"O crescimento explosivo da ciência cognitiva é importante para a filosofia. Mesmo que o estudo científico da mente e da cognição possa ser considerado como a invasão de um terreno tradicionalmente filosófico, o que se verifica é que a ciência cognitiva é em grande parte filosofia redescoberta."

Da Introdução

Sofia Miguens nasceu em 1969 no Porto, onde se licenciou em Filosofia.
Em 1995 concluiu, na Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa, uma tese de Mestrado intitulada *As Ciências Cognitivas e a Naturalização do Simbólico*, cujo tema foi a linguística cognitiva pós-chomskyana. Doutorou-se em 2002 com uma tese intitulada *Uma Teoria Fisicalista do Conteúdo e da Consciência – D. Dennett e os debates da filosofia da mente*.
É actualmente Professora Auxiliar do Departamento de Filosofia da Universidade do Porto. Tem leccionado nas áreas da Filosofia do Conhecimento, da Filosofia da Linguagem e da Filosofia da Mente. É membro do Gabinete de Filosofia Moderna e Contemporânea do Instituto de Filosofia da Faculdade de Letras da Universidade do Porto, em cujas actividades de investigação colabora já desde 1996.
Tem várias publicações em revistas da especialidade.
Este é o seu primeiro livro.

ISBN 972-610-653-2
9 789726 106531

Apoios:
FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA E DA TECNOLOGIA

Sofia Miguens

Uma Teoria Fisicalista do Conteúdo e da Consciência

CAMPO DAS LETRAS

Sofia Miguens

Uma Teoria Fisicalista do Conteúdo e da Consciência

D. Dennett e os debates da filosofia da mente

CAMPO DAS LETRAS

Campo da Filosofia/Noûs

Uma Teoria Fisicalista do
Conteúdo e da Consciência

UMA TEORIA FISICALISTA DO CONTEÚDO E DA CONSCIÊNCIA

Autor: Sofia Miguens
Capa: Hermínio Bastos

© CAMPO DAS LETRAS – Editores, S.A. – Porto, 2002
Rua D. Manuel II, 33 - 5.º, 4050-345 Porto
Tel. 226 080 870 Fax 226 080 880
E-mail: campo.letras@mail.telepac.pt
Site: www.campo-letras.pt

Impressão: Papelmunde - SMG, Lda.
Pré-impressão e acabamentos: Inforsete - V. N. de Famalicão
1.ª edição: Dezembro de 2002
Depósito legal: 193291/03
ISBN 972-610-653-2
Código de Barras: 9789726106531

Colecção: Campo da Filosofia/Noûs – 14

Sofia Miguens

Uma Teoria Fisicalista do Conteúdo e da Consciência

D. Dennett e os debates da filosofia da mente

CAMPO DAS LETRAS

Índice

Prólogo 17

Siglas utilizadas 19

Citações e traduções 21

INTRODUÇÃO
Apresentação geral do trabalho. A filosofia da mente. Ciência cognitiva como filosofia redescoberta. Será a psicologia (mais) importante para a filosofia (do que as outras ciências)? Estrutura específica do trabalho. 23

PRIMEIRA PARTE: As Origens 43

CAPÍTULO 1 - Dennett e a teoria da mente em 1965. O primeiro esboço da Teoria dos Sistemas Intencionais: princípios orientadores de Content and Consciousness. 45

1.1 A partir de Ryle e Wittgenstein, Quine e Putnam 51
1.1.1 Breve referência a E. Anscombe, C. Taylor e a alguns visionários na ciência e na engenharia. 70

1.2 O problema do conteúdo em *Content and Consciousness*: inícios de uma teoria teleológica do significado 74
1.2.1 A partir de fora e de cima. A referência segundo Quine. Existência e Identidade. 'Referencial' e 'não-referencial' segundo Dennett. O holismo e a fusão experimental das frases mentalistas nos seus contextos. 75
1.2.2 A partir de dentro e de baixo. Informação e Teleologia. Armazenamento inteligente de informação e comportamento. O

conteúdo e o funcionamento de estruturas apropriadas. Evolução no cérebro. 79
1.2.3 A estrutura do comportamento. O comportamento dirigido a fins (*goal directed behavior*) e a adscrição do conteúdo. Princípios da teleossemântica. Mente e linguagem. 86
1.2.4 Os níveis pessoal e sub-pessoal de descrição e explicação: como tratar o nível pessoal. A dor como exemplo. 93

1.3 O problema da consciência em *Content and Consciousness*: inícios de uma teoria deflacionária da consciência. 95
1.3.1 O funcionalismo segundo Putnam. A certeza introspectiva de um ponto de vista funcionalista. 96
1.3.2 Os sentidos do 'apercebimento' (*awareness*). Apercebimento e controlo (apercebimento-2), apercebimento e expressão verbal (apercebimento-1). 101
1.3.3 Imagens, *qualia*, preenchimentos e cores. Percepções de ausência ou ausência de percepção. 107

1.4 A Intenção: pensar e agir voluntariamente. 112
1.4.1 O nível pessoal e a teoria da acção: as razões na acção. A intenção segundo Anscombe e a sua adaptação à psicologia filosófica de Dennett. 118

1.5 A linguagem, o entendimento e o nível pessoal. O que fica estabelecido em *Content and Consciousness*. 128

SEGUNDA PARTE: O Modelo 133

CAPÍTULO 2 - A posteridade do funcionalismo de Putnam: diferendos acerca da natureza da psicologia. 135

2.1 Dos anos 70 aos anos 90: A teoria teleológica do conteúdo, as suas implicações e os seus opositores - *Brainstorms* (1978), *The Intentional Stance* (1987), *Brainchildren* (1998). 135
2.1.1 Linhas de análise da teoria do conteúdo: a formulação da Teoria dos Sistemas Intencionais (TSI), a oposição à Teoria dos Sistemas Intencionais (especialmente da Teoria Representacional da Mente), os princípios behavioristas da TSI e a sua ligação ao evolucionismo e a uma teoria geral do *design*. 135

2.1.1.1 Jerry Fodor, a Teoria Representacional da Mente e a Hipótese da Linguagem do Pensamento. 140
2.1.1.2 Instrumentalismo? - De *Intentional Systems a Real Patterns*. A TSI e as três estratégias (Estratégia Física, Estratégia do Design e Estratégia Intencional). 146
2.1.1.3 Princípios gerais. 146
2.1.1.4 *Intentional Systems*. 155
2.1.1.5 A suposição de racionalidade, o crente perfeitamente racional e o design da racionalidade. Das virtudes do behaviorismo skinneriano à TSI. 159
2.1.1.6 O valor de sobrevivência das crenças verdadeiras e a predominância do normal. A tensão entre racionalidade e incorrigibilidade. Racionalidade, holismo, indeterminação: as crenças nucleares e as outras crenças. Crenças e opiniões. 165
2.1.1.7 A oscilação entre instrumentalismo e realismo: o realismo enfraquecido, os verdadeiros crentes e os padrões reais. 171
2.1.1.7.1 *True Believers*. 171
2.1.1.7.2 *The Intentional Stance* e a exploração dos problemas delimitados em *True Believers*. 180
2.1.1.7.3 Aquilo em que uma rã acredita: *Evolution, Error and Intentionality*, a biologia evolucionista e a indeterminação quineana. 187
2.1.1.7.4 Dennett e o Quarto Chinês ou quando outros filósofos atacam a separação entre conteúdo e consciência. 194
2.1.1.8 *Real Patterns*. 197

2.2 A individuação do conteúdo e a explicação psicológica. Contra o interpretativismo da TSI: R. Millikan, F. Dretske, J. Fodor. 201
2.2.1 Funcionalismo e teleofuncionalismo: a reintrodução das funções biológicas na teoria da mente. Realismo acerca de representações. A TSI e o teleofuncionalismo. Problemas do externalismo. 201
2.2.2 *Beyond Belief*: mundos nocionais e proposições como medidas provisórias. 207
2.2.3 Explicação psicológica e individuação do conteúdo de acordo com outras posições externalistas 216
2.2.3.1 A biopsicologia de R. Millikan. 216
2.2.3.2 Informação e explicação do comportamento segundo F. Dretske. 225
2.2.3.3 J Fodor: a teoria nómico-informacional e o atomismo. 230
2.2.3.4 Sentencialismo, eliminativismo ou interpretação? 238

2.3 Dennett e o *design*: o foco das tensões internas da TSI 241

2.3.1 Oscilação entre *design* real e *design* como interpretação. 241
2.3.2 *Darwin's Dangerous Idea* (1995): a ideia de Darwin e o evolucionismo generalizado. Realidade e relatividade do *design*, de novo. 245

CAPÍTULO 3 - Pensamentos conhecendo outros pensamentos: Dennett e os debates em teoria da consciência dos anos 70 aos anos 90 (de Brainstorms *a* Brainchildren, *passando por* Consciousness Explained*).* 249

3.1 A consciência como problema dos anos 90. A posição heterodoxa e deflacionária de D. Dennett. Os *qualia* aniquilados, o Teatro Cartesiano desmontado, os *zombies* declarados inconcebíveis. Teorias empírico-especulativas da consciência. Um outro prisma: o problema metafísico da consciência fenomenal e da sua irredutibilidade. 249
3.1.1 Alguns marcos da investigação empírico-especulativa. 252
3.1.2 *Quining Qualia.* 261
3.1.3 Um outro prisma: metafísica e fundamentalidade da consciência. 267

3.2 Modelos de consciência e natureza das experiências: *Brainstorms* (1978). 278
3.2.1 Experiência, memória, apresentação, expressão: os sonhos. 286
3.2.2 Imagens mentais. 296
3.2.3 A dor: sentir-se ser e simulação. 308

3.3 Modelos de consciência e natureza das experiências: *Consciousness Explained* (1991) e o *close-up* do observador. 310
3.3.1 De novo a partir de fora e de cima: Shakey, SHRDLU e a heterofenomenologia. 317
3.3.2 O Modelo dos Esboços Múltiplos. 322
3.3.3 O tempo e a consciência. Temporalização. 324
3.3.4 De novo a partir de dentro e de baixo. A evolução da consciência do ponto de vista da terceira pessoa e evitando considerar a representação do que quer que seja. Fronteiras, razões, sensiência e futuro. Evolução no cérebro. Máquinas Virtuais instaladas. 337
3.3.5 A ilusão do utilizador da Máquina Virtual joyceana. 348
3.3.5.1 Ainda assim um interior: querer-dizer, pandemónio e actos de fala. 351
3.3.5.2 O Eu. 357
3.3.6 A última palavra (filosófica) quanto ao Teatro Cartesiano. 361
3.3.6.1 Mostrar ou dizer. 361
3.3.6.2 Querer-dizer, reportar e exprimir. 363

3.3.6.3 Quebrar a barreira da testemunha: uma interpretação da visão cega. 367
3.3.6.4 Ver é saber? 370
3.3.6.5 A cor, de novo (e uns certos gostos e desgostos ligados aos *qualia*). 373
3.3.6.6 Mary e os *zombies*. O Quarto Chinês. O morcego. 380
3.3.7 Ciência cognitiva ou teoria filosófica da consciência? O choque de intuições quanto à fundamentalidade da consciência: Dennett versus Searle, Nagel, Chalmers e Jackson. O concebível e o inconcebível. 384

CAPÍTULO 4 - As Pessoas e as suas Acções: a filosofia da mente e os fundamentos da filosofia moral. 391

4.1 Pessoa e acção como conceitos normativos: a filosofia moral e a teoria cognitiva. Naturalismo gradualista e compatibilismo. 391
4.2 A fragmentação do problema da vontade livre e as formas que o problema da vontade livre não tem: vontade numénica, indeterminismo físico, capacidades mentais não mecâncas. 397

4.3 A liberdade num mundo determinista: aleatoriedade, controlo, espaço de manobra (*elbow room*) e descrição intencional. 402
4.3.1 A rigidez, o espaço de manobra e o controlo. Condições da acção: determinismo físico, determinação do *design*, limitações cognitivas. A impossibilidade física da vontade pura e a sua substituição pela prudência. 403

4.4 Da teoria do controlo ao auto-controlo meta-reflexivo e à avaliação forte. 406
4.4.1 Deliberação, decisão, oportunidade. Previsibilidade e imprevisibilidade. A deliberação e a decisão: vantagens da insensibilidade e da arbitrariedade. Deliberação e possibilidade epistémica. 406
4.4.2 Razões e *self*. O eu e as suas ficções. Real, virtual. Indeterminação. Auto-exortação. 410
4.4.3 Do eu à pessoalidade. 414
4.4.4 'Quererei eu realmente ser aquilo que sou?' Os meta-problemas difíceis do controlo, a avaliação forte e a sorte moral. 422
4.4.5 Sorte moral ou responsabilidade. A prudência e o *design* do deliberador. 426

4.5 Manual de primeiros socorros morais e ética da virtude. 429

4.6 Problemas de fundo. 433

CAPÍTULO 5 - Questões aplicadas da TSI. Tipos de mentes: mentes animais, artificiais e humanas. 437

5.1 A horizontalidade de perspectiva da TSI sobre o mental: o natural e o artificial. Ciência cognitiva, engenharia invertida e/ou síntese. Cérebros, programas, robôs: da base para o topo e do topo para a base. A IA e as experiência de pensamento reais. Os tipos de mentes e a tipologia (a partir de dentro e de baixo) das criaturas cognitivas. 437

5.2 Questões aplicadas da TSI: as mentes animais e o problema da interpretação, o problema do enquadramento (*frame problem*) na IA e a incorporação das mentes. 445
5.2.1 Mentes animais: etologia cognitiva. 448
5.2.2 Mentes artificiais: O problema do enquadramento na IA. 462
5.2.3 A incorporação das mentes: perturbações do funcionalismo. 473

TERCEIRA PARTE: Os Problemas. 477

CAPÍTULO 6 - Fisicalismo, Conteúdo e Consciência: da filosofia da mente à ontologia. 479

6.1 Primeiro e fundamental ponto crítico para a avaliação da TSI: o fisicalismo e a irredutibilidade da intencionalidade. 479

6.2 A natureza e o seu interior I. Será a TSI necessariamente fisicalista? O realismo moderado, a teoria evolucionista da cognição e o estatuto dos interfaces. A (in)justificação do fisicalismo. 482
6.2.1 J. Haugeland e a teoria do entendimento (ou como reconciliar a Estratégia Intencional com a Intencionalidade Intrínseca de J. Searle). 490
6.2.2 Pós-antropologismo: B. Cantwell Smith e a origem dos objectos. 500

6.3 A natureza e o seu interior II. Racionalidade: a impossibilidade de irracionalidade e a racionalidade mínima. 509

6.4 A natureza e o seu interior III. Consciência fenomenal ou ilusão do utilizador de uma Máquina Virtual. Epifenomenismo. 520

CONCLUSÃO: Vale a pena fazer filosofia da mente? Um retorno às origens. 537

Contexto histórico-sociológico da filosofia da mente e da ciência cognitiva. A tradição filosófica. Vale ou não vale a pena? - A nossa natureza mental. 539

BIBLIOGRAFIA 561

ÍNDICE ONOMÁSTICO 589

ÍNDICE TEMÁTICO 593

«We must think of mind as a phenomenon to which the human case is not necessarily central, even though our minds are at the center of our world. The fundamental idea behind the objective impulse is that the world is not our world»
Thomas Nagel, *The View From Nowhere*

Prólogo

Este trabalho resulta de uma Dissertação de Doutoramento apresentada à Faculdade de Letras da Universidade do Porto em 2001. Espero que, na sua forma presente, ele sirva alguns propósitos práticos, nomeadamente que possa ser encarado como uma introdução ao campo da filosofia da mente e à história das relações desta com as ciências cognitivas. Espero que também possa servir como uma introdução ao pensamento do filósofo americano D. Dennett.

Um processo de reescrita é sempre complicado, sobretudo quando – e é esse o caso – se continua a trabalhar sobre os mesmos assuntos. Evitei portanto modificar demasiado o texto inicial.

Tenho muitos agradecimentos a fazer relativamente aos anos de preparação que conduziram à conclusão deste trabalho. Antes de mais, ao Professor Doutor Fernando Gil, o Orientador da minha tese de Doutoramento, pelo seu incentivo e pelo exemplo que para mim sempre representou. Em seguida, à Professora Doutora Maria José Cantista, Co-Orientadora da tese, por todo o apoio que sempre me deu em todos os aspectos da vida profissional e pela sua amizade. Agradeço-lhe também pelo facto de, enquanto Coordenadora do Gabinete de Filosofia Contemporânea do Instituto de Filosofia da Faculdade de Letras, ter apoiado e incentivado a presente publicação, como de resto tem vindo a fazer nos últimos anos com a investigação em filosofia contemporânea na FLUP. Ao Professor João Branquinho, da Faculdade de Letras da Universidade de Lisboa, que foi o arguente da tese, quero agradecer os comentários e sugestões, que não foram todos aqui incorporados ou considerados, mas que ficam registados para o meu futuro trabalho. Quero agradecer ainda a duas pessoas pela leitura e comentários que foram fazendo de várias versões do texto: o Dr. João Alberto Pinto, da Faculdade de Letras da Universidade do Porto, e o Professor Doutor António Machuco, da Universidade Lusófona de Humanidades e Tecnologias. Pelo contexto que propiciaram para um trabalho em curso, quero agradecer aos alunos da cadeira de Filosofia do Conhecimento da Faculdade de Letras da Universidade do Porto, no âmbito da qual tive a possibilidade, desde 1996 até 2000, de ensinar Filosofia da Mente.

Em termos institucionais, cabe-me agradecer vários apoios: do Gabinete de Filosofia Moderna e Contemporânea da Faculdade de Letras da Universidade do

Porto, coordenado pela Professora Doutora Maria José Cantista, no âmbito do qual foi possível uma primeira deslocação à New York University em Novembro/Dezembro de 1999, da Fundação Luso Americana para o Desenvolvimento e da Fundação Calouste Gulbenkian, que em conjunto financiaram a minha estadia como *visiting scholar* no Departamento de Filosofia da New York University no Semestre de Outono de 2000 e do Departamento de Filosofia da Faculdade de Letras da Universidade do Porto, que propiciou as condições para a elaboração deste trabalho.

Quero finalmente dedicar este livro a duas pessoas que para ele muito contribuiram: o João, e a minha Avó, Maria Alberta de Morais Miguens, que sempre me incentivou em tudo o que quis fazer.

Porto, Verão de 2002

Siglas Utilizadas

TSI – Teoria dos Sistemas Intencionais

MEM – Modelo dos Esboços Múltiplos

SI – Sistema Intencional

EI – Estratégia Intencional

ED – Estratégia do Design

EF – Estratégia Física

TRM – Teoria Representacional da Mente

RM – Representação Mental

C&C – *Content and Consciousness*

BS – *Brainstorms*

ER – *Elbow Room*

IS – *The Intentional Stance*

CE – *Consciousness Explained*

DDI – *Darwin's Dangerous Idea*

KM – *Kinds of Minds*

BC – *Brainchildren*

Nota: Optou-se por considerar que formas como SI e RM valem para o singular e para o plural.

Citações e traduções

Optou-se por traduzir todas as citações que aparecem no corpo do trabalho. As traduções são feitas directamente a partir da versão original dos textos. As citações que aparecem em nota são feitas na língua original. São por vezes mantidos termos e frases curtas na língua original no corpo do texto devido à sua particular expressividade.

Grande parte da terminologia da ciência cognitiva não está definitivamente fixada em português. Frequentemente recorreu-se ao uso habitual de termos em áreas como a Inteligência Artificial e a Psicologia de modo a estabelecer traduções. Assim, por exemplo, *embeddedness* é traduzido por embebimento, *frame problem* por problema do enquadramento, *top-down* e *bottom-up* como do topo para a base e da base para o topo. Os termos *self*, input e output são utilizados na sua forma inglesa. Decidiu-se não traduzir o termo *design*, que é corrente nessa forma por exemplo em áreas artísticas. No entanto, traduziu-se o termo *designed* por desenhado e por vezes por projectado. Traduziu-se o termo *awareness*, que é frequentemente considerado como sinónimo de consciência e que é um termo chave para estabelecer distinções entre conteúdo e consciência, por apercebimento.

Em geral traduziu-se a palavra *token* por espécime, a palavra *type* por tipo e a palavra *kind* por género. Os termos input e output, a priori e a posteriori são considerados como termos teóricos comuns e portanto não são escritos em itálico. O mesmo acontece com o termo *design* no âmbito da expressão 'Estratégia do Design'.

Introdução

Apresentação geral do trabalho. A filosofia da mente. Ciência cognitiva como filosofia redescoberta. Será a psicologia (mais) importante para a filosofia (do que as outras ciências)?

Termos como *design,* input e output, *top-down, bottom-up, embeddedness e awareness,* são amplamente utilizados em teoria da cognição – por exemplo na Psicologia e na Inteligência Artificial – e permitem-nos falar sobre sistemas cognitivos ou agentes racionais que manifestam inteligência ou consciência e que, como tal, presumivelmente utilizam representações, entre as quais auto-representações, para se comportarem no mundo físico. O estatuto de tais termos não é no entanto imediatamente claro. Este trabalho tem como objectivo avaliar as pretensões de esclarecimento desse estatuto avançadas por uma meta-teoria da cognição a que se chama Teoria dos Sistemas Intencionais (TSI).

De modo geral utiliza-se o termo 'sistema' no sentido de entidade isolada ou isolável. Utiliza-se o termo 'racional' para nomear sistemas que 'guiados por representações' se comportam de forma adaptada ao ambiente em função de 'estruturas de finalidades'. É relativamente a esses sistemas racionais, que se analisa a natureza e conteúdo dos estados mentais, a consciência e a auto-consciência, a existência de acções por oposição a meros movimentos, a identidade pessoal por oposição à mesmidade de um corpo que persiste ao longo do tempo e, finalmente, a racionalidade que de acordo com a TSI sustenta todas as atribuições de características mentais.

À primeira vista, este trabalho é sobre um autor, um filósofo americano contemporâneo, Daniel Dennett[1]. No entanto, aquilo que de facto se pretende é utilizar a obra de Dennett como fio con-

[1] N. 1942, MA Harvard 1963, DPhil Oxford 1965.

dutor para a exposição e avaliação de determinados problemas e teses da filosofia da mente. É por essa razão que a dissertação se intitula *uma* teoria: a TSI proposta por Dennett é uma de entre as várias teorias que na filosofia da mente competem no sentido de esclarecer a natureza do conteúdo e da consciência. Assim, embora se proponha que a TSI constitui um ponto de observação privilegiado dos principais debates que atravessaram a filosofia da mente e a ciência cognitiva nos últimos trinta e cinco anos, aproximadamente, admite-se que tais debates possam também evidenciar alguns dos seus limites. A particular teoria analisada neste trabalho, a TSI, é em última análise uma teoria normativa ou transcendental da mente. Entende-se por teoria normativa ou transcendental uma teoria segundo a qual objectos são inteligíveis em virtude de um compromisso prévio quanto àquilo que eles podem ser. É certo que o próprio Dennett não se refere nunca à TSI como uma teoria transcendental. No entanto, alguns dos intérpretes do seu pensamento não hesitam em fazê-lo[2] e este trabalho adoptará tal posição. É importante no entanto que fique claro que alguns dos tratamentos não meramente expositivos da TSI que aqui se leva a cabo não seriam facilmente aceites por Dennett. Por exemplo no que diz respeito à consideração da TSI como teoria transcendental da mente, tudo o que se encontra no próprio Dennett é a ideia de que uma suposição de racionalidade é constitutiva do mental enquanto mental, sendo unicamente mediante uma Estratégia Intencional (EI) que sistemas físicos são interpretados como mentais ou cognitivos.

A TSI é uma teoria quineana da interpretação de sistemas físicos supondo a racionalidade. Inicialmente definida como uma teoria fisicalista, funcionalista e instrumentalista do mental, ela abarca neste momento um conjunto de posições sofisticadas relativamente aos fenómenos do conteúdo e da consciência. A TSI é assim, nomeadamente, um realismo moderado (acerca da *natureza das representações* em sistemas cognitivos), um teleofuncionalismo interpretativista (quanto ao *conteúdo* dessas representações), um niilismo do significado (quanto ao *poder explicativo* das noções semânticas). Estas posições perante a natureza do significado são acompanhadas por uma teoria deflacionária da consciência, eliminativista relativamente aos *qualia*[3] e por

[2] Cf. por exemplo FODOR & LEPORE 1992 e HAUGELAND 1997.

[3] O termo *qualia* é utilizado na filosofia da mente para referir as qualidades da experiência sentida e o estatuto especial de tais qualidades no âmbito do mental.

uma teoria naturalista e normativa (e por isso gradualista) quanto à natureza de pessoas e de acções. Globalmente considerada, a TSI representa ainda uma posição ecumémica relativamente à distribuição da mentalidade por entidades naturais e artificiais.

Se o *fisicalismo* é, em traços gerais, a convicção de que a natureza do mundo é fundamentalmente física, o *funcionalismo* é a ideia segundo a qual a natureza da cognição é independente do substracto material desta. Esta exploração psicológica e filosófica da diferença entre *hardware* e *software* redunda no dualismo característico de muitas das teorias contemporâneas da cognição. Ser *instrumentalista* acerca de alguma coisa é utilizá-la por motivos pragmáticos sem lhe conceder realidade ou importância fundamental. O instrumentalismo aqui em causa é um instrumentalismo quanto à linguagem mental. Mas a TSI vem a mostrar não ser exactamente instrumentalista: o que ela faz é ligar o reconhecimento dos padrões da mentalidade ao ponto de vista de um intérprete, considerando que os padrões são reconhecíveis pelo intérprete devido à realidade do *design* que lhes subjaz.

O *niilismo* do *significado*[4] traduz-se, por sua vez, nos termos da TSI, na ideia segundo a qual não existem 'máquinas semânticas' (*semantic engines*): o facto de alguma coisa significar não pode ser responsável pela causação do que quer que seja. Sendo todo o sistema cognitivo um sistema *físico,* a linguagem mentalista (que descreve sistemas físicos em termos de razões, significados, pensamentos) não descreve nunca os processos causais que finalmente e realmente guiam os comportamentos. Num mundo de máquinas semânticas o mental seria causalmente eficaz e a linguagem mentalista seria explicativa. No nosso mundo, o 'mental' é apenas uma interpretação holista de sistemas físicos e um atalho heurístico. Deflacionar é, por outro lado, diminuir o valor de alguma coisa. Esta intenção traduz-se no âmbito da teoria dennetiana do mental na identificação da consciência com uma certa incorrigibilidade no auto-acesso de sistemas, característica da produção de relatos (linguísticos) acerca de si, por oposição a alguma aparição fundamental e a-conceptual ou a uma especificidade ontológica do sentimento de si ou da experiência de ser. O Modelo dos Esboços Múltiplos (MEM), a teoria da consciência que a TSI enquadra, é de natureza deflacionária.

A partir destas posições básicas, o funcionalismo, a deflação e um instrumentalismo que é afinal apenas interpretativismo, a TSI

[4] A expressão é utilizada por J. Fodor e E. Lepore. Cf. FODOR &LEPORE 1992.

pretende analisar a subjectividade a partir de vários ângulos. Um dos objectivos deste trabalho é verificar se a análise proposta é globalmente coerente, se ela constitui realmente uma teoria fisicalista do mental e se é legítimo fazer uso de uma suposição de racionalidade na interpretação mentalista de sistemas físicos sem oferecer uma teoria explícita da natureza da racionalidade.

Além de abordar problemas concretos de um domínio (a filosofia da mente) através das soluções para esses problemas exploradas na obra de um filósofo (Daniel Dennett), o presente trabalho tem como intenção situar Dennett no âmbito da história da filosofia analítica. Ora, a situação da filosofia da mente no âmbito da filosofia analítica é inseparável do desenvolvimento recente da ciência cognitiva[5]. Assim, um outro objectivo do trabalho é investigar a relação da filosofia da mente com a história da ciência cognitiva nas últimas décadas, nomeadamente focando debates com grande peso filosófico tais como os debates em torno da existência de uma linguagem do pensamento, da existência de imagens mentais, da natureza da consciência, da natureza da racionalidade, da natureza dos conceitos, etc. Não é sequer necessário introduzir a partir de fora tais discussões da ciência cognitiva uma vez que elas aparecem naturalmente no trabalho dos filósofos.

No entanto, não é apenas o teor da investigação empírica acerca de cognição que importa quando se procura investigar a relação da filosofia da mente com a ciência cognitiva: importam também as posições alternativas defendidas por filósofos relativamente a essa investigação. Embora nem todas as teorias do conteúdo e da consciência que divergem da de D. Dennett sejam aqui consideradas, procurar-se-á tanto quanto possível, dar voz a filósofos que, no mesmo período, no mesmo contexto e lidando com idênticos dados provenientes dos estudos científicos da cognição defenderam posições em tensão com as posições de Dennett[6]. Quanto à teoria da

[5] Traduzir-se-á *cognitive science* por ciência cognitiva e não por ciências cognitivas. Embora a 'ciência cognitiva' não seja uma disciplina mas uma constelação de várias disciplinas todas elas têm por objecto a 'cognição'.

[6] Um caso muito especial é o D. Davidson, que defende uma teoria da mente com inúmeros paralelismos com a de Dennett (cf. DAVIDSON 1980 e DAVIDSON 1984). O pensamento de D. Davidson não será muito explorado neste trabalho devido ao facto de Davidson não ser centralmente um filósofo da ciência cognitiva. No entanto, exactamente como Dennett, Davidson pensa que a teoria da mente parte da situação quineana de interpretação radical, é uma teoria (holista) da interpretação que (pre)supõe a racionalidade, nega a existência de leis psicofísicas e defende a irredutibilidade do aspecto normativo do pensamento.

representação são disso exemplo autores como Jerry Fodor, defendendo o realismo intencional[7], Fred Dretske investigando o papel da informação na explicação do comportamento[8], Ruth Millikan elaborando uma biossemântica teleofuncionalista[9] ou Paul Churchland propondo o materialismo eliminativo[10]. Relativamente ao problema da consciência David Chalmers[11], John Searle[12] e Frank Jackson[13] são alguns dos filósofos que ao contrário de Dennett consideram a questão da consciência uma questão metafísica fundamental.

A filosofia da mente, e não apenas a filosofia de Dennett, vive hoje da proximidade que estabelece com a investigação empírica multidisciplinar acerca da cognição. Os tópicos clássicos da filosofia da mente, tais como a consciência, a representação e a causação mental, são frequentemente tratados (tal como Quine prescreveu[14]) em continuidade com o trabalho científico sobre o mental. A enorme quantidade de investigação empírica em ciência cognitiva nas últimas décadas explica aliás em parte, de um ponto de vista sociológico, o protagonismo da filosofia da mente dentro da filosofia analítica no período em causa. Como afirma J. Kim, «nas últimas duas décadas, aproximadamente, a filosofia da mente tem sido uma área extraordinariamente activa e excitante. O campo cresceu enormemente, e (...) houve avanços significativos no nosso entendimento das questões respeitantes à mente (...) Este *boom* foi em parte, devido ao ímpeto fornecido pelo crescimento explosivo, desde meados do século, da ciência cognitiva»[15]. Ora, o crescimento explosivo da ciência cognitiva é importante para a filosofia. Mesmo que o estudo científico da mente e da cognição possa ser considerado como a invasão de um território tradicionalmente filosófico, o que se verifica é que «a ciência cognitiva é, em grande parte, filosofia redescoberta e (...) filosofia reabilitada»[16]. Um primeiro aspecto dessa reabilitação é o renascimento de teorias realis-

[7] FODOR 1975, FODOR 1987, FODOR 1998.
[8] DRETSKE 1981, DRETSKE 1988.
[9] MILLIKAN 1984, MILLIKAN 1993.
[10] CHURCHLAND 1981.
[11] CHALMERS 1996.
[12] SEARLE 1992.
[13] JACKSON [1986], JACKSON 1993.
[14] Faz-se alusão, evidentemente, à ideia de epistemologia naturalizada.
[15] KIM 1996: xi.
[16] FODOR 1981: 26.

tas da representação, que há muito pareciam abandonadas[17]. Embora o autor central neste trabalho pareça (pelo menos à primeira vista) defender uma posição anti-realista acerca da representação, mesmo para ele o realismo renascido é incontornável. É por essa razão aliás, que o realismo intencional de J. Fodor é o contraponto constante da apresentação da TSI. O facto de estar em causa uma investigação *científica* do mental é naturalmente uma das razões para o mencionado renascimento do realismo. No entanto, dado o retorno do realismo da representação, o problema é saber se ele sustenta uma posição metafísica tambem ela realista. Um segundo aspecto envolvido na reabilitação da filosofia é o surgimento de uma clara necessidade de fundamentação epistemológica. O trabalho empírico sobre o mental foi sendo naturalmente acompanhado por uma reflexão que só pode ser qualificada como filosófica, mesmo que muito dos seus praticantes não sejam filósofos profissionais. A reflexão epistemológica é uma vocação clássica da filosofia e uma vocação específica da filosofia moderna – pense-se em autores como Descartes, Leibniz, Hume, Kant – caracterizada por acompanhar a ciência natural em desenvolvimento. É essa mesma vocação epistemológica da filosofia que é solicitada contemporaneamente pela ciência cognitiva. Simplesmente não se trata agora de fundamentar a física clássica, a mecânica galilaica ou newtoniana, mas a investigação científica do mental como parte da natureza. Uma certa identidade de intenção da filosofia da mente contemporânea e da epistemologia clássica, dos séculos XVII e XVIII, explica o facto de os problemas de ambas se assemelharem tão estranhamente, tanto que não é raro encontrar tipologias de posições em filosofia da mente iniciadas com o dualismo de Descartes, o paralelismo de Leibniz ou o monismo de Espinosa ou posições acerca do conteúdo e da consciência nomeadas como harmonia pré-estabelecida ou epifenomenismo[18]. De certo modo, uma nova inocência parece ter sido ganha após o divórcio entre a filosofia e as ciências naturais notório em muito do trabalho filosófico dos séculos XIX e XX.

[17] Essa morte nunca foi definitiva, mas em vários quarteirões da filosofia o realismo era simplesmente inadmissível. Basta pensar, relativamente à teoria analítica do mental, na grande influência de G. Ryle e de L.Wittgenstein.

[18] Cf. DENNETT 1998b. Dennett louva Dretske por, precisamente, não se coibir de utilizar tais termos (epifenomenismo, harmonia pré-estabelecida) nas discussões sobre conteúdo mental, uma vez que é precisamente isso que está em causa.

A questão epistemológica do fundamento aparecerá, neste trabalho, em primeiro lugar. Segundo Dennett, cabe à filosofia da mente *fundamentar* (e não competir com) teorias neurofisiológicas, psicológicas, computacionais da cognição, procurando clarificar as suposições metafísicas que estas inevitavelmente fazem. Uma teoria filosófica da mente deverá, assim, estar preparada para responder a questões tais como: Existem mentes? Em que sentido *existem*? Serão as mentes físicas? A teoria da cognição porá em causa que a maneira básica de existir seja física? Como sabemos seja o que fôr acerca de mentes? Em que sentido são as mentes responsáveis pela constituição de mundos dentro do mundo? Podem esses mundos ser incompatíveis? Estas questões mais directamente teóricas (metafísicas, ontológicas, epistemológicas) não podem deixar de trazer consigo questões éticas ou proto-éticas. Exemplos de tais questão são: Porque é que para algumas mentes, ao contrário do que acontece com as pedras ou as estrelas, as coisas importam, sendo não apenas apercebidas mas também sentidas? Como se relacionam o mental e o sentimento de si? Como se relacionam o eu e a consciência com o controlo do comportamento? Em que é que uma acção voluntária de uma entidade consciente difere de um mero movimento? O que é que isso tem a ver com a existência de 'pessoas'?

Mesmo se todas estas questão estão tacitamente presentes na investigação empírica da cognição, as respostas não aparecem simplesmente a partir dos resultados desta. De facto, a TSI não é uma hipótese empírica acerca do mental mas sim, em grande parte, uma hipótese apriorista. Alguns autores chamam-lhe mesmo, como foi dito, uma hipótese transcendental, entendendo por transcendental, repita-se, a ideia segundo a qual objectos teóricos apenas são inteligíveis em virtude de um compromisso prévio que marca os limites daquilo que eles podem ser. Esse seria o caso do mental como objecto de pensamento. Admitindo que uma teoria filosófica da mente consiste num trabalho conceptual de fundamentação, não duplicando o trabalho empírico, ainda assim a resposta às questões acima enumeradas não pode ser puramente conceptual e apriorista. Ela só poderá surgir, segundo Dennett, de um inquérito impuro. Noutras palavras, a teoria filosófica da mente não pode ignorar o trabalho empírico sobre cognição. A 'impureza' da teoria da mente defendida por Dennett reflectir-se-á neste trabalho numa troca constante (e numa deambulação, sem necessidade de delimitação estrita) entre filosofia e psicologia.

É verdade que se atribui aqui um sentido muito lato ao termo 'psicologia', de acordo com o qual a psicologia diz respeito a qualquer tipo de mente e não apenas às mentes humanas ou biologicamente realizadas. Considera-se que este sentido de 'psicologia' foi justificado por H. Putnam com o desenvolvimento das ideias funcionalistas nos anos 60[19]. Segundo o funcionalismo, como já se afirmou, a natureza dos estados mentais é de uma maneira importante independente do substracto material que os implementa. Os predicados psicológicos são assim, de certa maneira, não físicos. Esta afirmação só obtem, obviamente, todo o seu peso se fôr entendida no âmbito de um estudo científico e materialista da cognição.

J. Fodor, em *The Language of Thought* (1975), uma obra central no período de que se ocupará este trabalho, afirma que durante muito tempo ficou mal aos filósofos psicologizarem e aos psicólogos filosofarem[20]. Afirma-o antes de declarar que é exactamente isso que vai fazer. Psicologia especulativa, conceptualmente disciplinada, e filosofia empiricamente informada: é esse o teor do trabalho de filósofos como D. Dennett e de J. Fodor, e é isso que os torna representantes centrais da inovação metodológica em filosofia que a filosofia da mente das últimas décadas constitui. Uma tal opção metodológica, que não é obviamente partilhada por todos os filósofos analíticos contemporâneos, é uma consequência possível do imperativo quineano de naturalização da epistemologia, e de algum modo conduz a considerar a filosofia da mente como um ramo da filosofia da ciência. Mais especificamente a filosofia da mente é filosofia da psicologia num sentido generalizado, um sentido em que o termo 'psicologia' nomeia todas as disciplinas científicas que se ocupam do mental, desde aquelas que estão mais próximas da engenharia e da biologia, como é o caso, respectivamente, da Inteligência Artificial e das neurociências, até aquelas que se ocupam com níveis mais afastados da implementação material, como a etologia cognitiva e a linguística.

O trabalho de filósofos da mente como Dennett e Fodor pode parecer, à primeira vista, pertencer àquele a que Putnam chama o *shallower level*[21], um nível mais superficial da filosofia, em que se trata

[19] O lugar clássico destas ideias funcionalistas são os artigos *Minds and Machines* (PUTNAM 1960), *The Nature of Mental States* (PUTNAM 1967a), *Robots: Machines or Artificially Created Life* (PUTNAM 1964), *The Mental Life of Some Machines* (PUTNAM 1967b) reunidos no 2º volume *(Mind, Language and Reality)* dos *Philosophical Papers* de H. Putnam (PUTNAM 1975).

[20] FODOR 1975: vii-viii «Since (...) psychologizing and philosophizing are mutually incompatible activities, these accusations were received with grave concern».

[21] Cf. PUTNAM 1988.

de questões 'pequenas e práticas' como 'Pode-se atribuir crenças a paramécias, ostras ou macacos?', por oposição às grandes questões como 'Será que a realidade é independente do pensamento?' ou 'Será que o pensamento é independente da linguagem?'. No entanto, as explorações desse nível mais superficial, ou menos directamente metafísico, contribuem muito para a própria qualidade do tratamento teórico das grandes questões metafísicas. Mas se filósofos como Dennett e Fodor se dedicam à investigação da cognição lado a lado com cientistas, não é conveniente ocultar que, para muitos filósofos analíticos, o puro *armchair knowledge*, o esclarecimento conceptual em grande medida apriorista, continua a ser o núcleo metodológico da filosofia. Deste ponto de vista, que se poderia considerar representado por exemplo por Thomas Nagel ou Michael Dummett, a filosofia da mente tal como Fodor ou Dennett a praticam representa uma concessão excessiva do filósofo às suas próprias inclinações científicas.

Face a essa crítica possível e admitindo que o impacto da ciência cognitiva foi decisivo no desenvolvimento da filosofia americana pós-quiniana, um outro aspecto da intenção de alguma maneira histórica deste trabalho, é a caracterização desse desenvolvimento. Escolhendo Dennett como referência num trajecto pela filosofia da mente procura-se explicitar algumas raízes da recente maneira americana de fazer filosofia, de modo a enfrentar directamente a crítica de certos autores (como por exemplo M. Dummett) segundo os quais essa maneira representaria uma regressão relativamente à sofisticação conseguida pela filosofia analítica anterior. Esta, na tradição de Frege e de Wittgenstein, ao conceber a metodologia filosófica como uma análise do pensamento feita através da análise da linguagem e ao conceber portanto a filosofia como primeiramente filosofia da linguagem, conseguiria uma 'extrusão' (na expressão de M. Dummett) do pensamento em relação à mente individual[22] que evitaria o erro de conceber sem mais que os pensamentos, a vida mental, fazem parte do fluxo de consciência do indivíduo empírico. Esta orientação teria sido perdida em grande parte da filosofia americana contemporânea, com a desastrosa opção pela filosofia da mente em detrimento da filosofia da linguagem, trazendo consigo os perigos do cientismo, do psicologismo e da esterilidade.

[22] Cf. DUMMETT 1993: 22. M. Dummett considera aliás que foi também este o imperativo que presidiu à génese da fenomenologia.

No pensamento de Dennett permanecem muitas margens indefinidas e muitas respostas incompletas, possivelmente por razões como as apontadas por Dummett. Há nomeadamente um lugar vazio para a especificação de uma teoria da objectividade que sustente a teoria da representação proposta, bem como a necessidade de uma ontologia mais elaborada, que explore a relação entre o ponto de vista evolucionista acerca da ontologia e da metafísica na TSI e o fisicalismo. A filosofia de Dennett é, no entanto, uma filosofia prática, no sentido em que está comprometida com o trabalho empírico. Para além disso, pode-se sempre pensar que se a teoria da mente aparece como a teoria de um híbrido deselegante, e se ela captura a natureza nas suas articulações, então é porque o mental é um híbrido deselegante. É preciso considerar essa hipótese, por mais que ela contradiga séculos de especulação filosófica em torno de uma razão 'pura', teórica e prática[23], que permitiria enfrentar mais directamente as acima mencionadas questões da verdade e da ontologia. O trabalho de Dennett não é aliás uma filosofia prática apenas na medida em que acompanha a ciência empírica mais relevante e aposta numa ontologia e epistemologia científicas *standard* e conservadoras. O próprio autor assume praticar «a escrita filosófica com um espírito de engenharia», considerando que «desenha e constrói dispositivos – argumentos, bombas de intuição (*intuition pumps*), máquinas de metáforas – que supostamente atingirão determinados efeitos»[24]. Por outro lado, e ainda num terceiro sentido, a filosofia de Dennett é uma filosofia prática na medida em que considera a teoria das pessoas e das acções como o ponto culminante da teoria da mente. Os conceitos de pessoa e de acção esclarecem-se mutuamente e um tal esclarecimento apenas tem lugar, evidentemente, após o esclarecimento das noções de conteúdo e de consciência. A teoria das pessoas e das acções é um ponto delicado da filosofia da psicologia, no qual está em causa a relação entre o conhecimento empírico e a normatividade. Apesar dessa dificuldade, a exploração dos fundamentos da ética é uma constante no trabalho de Dennett, que considera que a teoria da mente envolve inevitavelmente no caso humano uma teoria da pessoalidade (*personhood*) e que o conceito de acção é central para a própria possibilidade de pessoas (por oposição a, ou em acrescento a, organismos humanos vivos ou outro tipo de entidades).

[23] Evidentemente, os acima citados séculos de divórcio entre filosofia e ciências naturais (os séculos XIX e XX), muito claro pelo menos na filosofia a que os filósofos analíticos chamariam continental, corresponderam a filosofias da razão impura (pense-se na filosofia de Nietzsche, ou no existencialismo). Simplesmente essa teoria da razão impura foi sobretudo uma teorização existencial e não tanto epistemológica, como será o caso aqui.

[24] DENNETT 1993: 203.

De acordo com o que se tem vindo a afirmar, a ciência cognitiva não poderia deixar de constituir o pano de fundo do trajecto filosófico de Dennett. Ela é hoje um campo disciplinar plenamente constituído (sobretudo nos Estados Unidos), presente em inúmeros programas universitários de ensino e investigação. A expressão *cognitive science* só começou a ser utilizada nos anos 70, mas a gestação da chamada revolução cognitiva[25] pode ser reportada à cibernética dos anos 40[26]. O nascimento de facto da ciência cognitiva é normalmente fixado em 1956, e não é uma coincidência indiferente que essa seja também a data oficial de nascimento da Inteligência Artificial. O nascimento da ciência cognitiva é fixado em 1956 por ser essa a data de um encontro acerca de teoria da informação no MIT, na qual estiveram presentes Allen Newell e Herbert Simon, Noam Chomsky e George Miller, apresentando respectivamente os artigos *The Logic Theory Machine, Three Models for the Description of Language, The Magical Number Seven, Plus or Minus Two*[27]. Foi aí que, na conhecida descrição do psicólogo George Miller, a convicção de que várias disciplinas eram peças de um todo mais vasto se impôs[28]. Esse todo mais vasto visado por várias disciplinas pode ser chamado a cognição, o mental, a agência (*agency*) inteligente, e as maneiras de o estudar são muito variadas[29].

[25] Cf. GARDNER 1984 para um estudo histórico da revolução cognitiva, em várias disciplinas (psicologia, filosofia, Inteligência Artificial, neurociência, antropologia).

[26] Cf. DUPUY 1994. J.P. Dupuy parte da hipótese de que a ciência cognitiva tem a sua origem no movimento cibernético, no qual teria sido pela primeira vez explorada a ideia segundo a qual o que é importante para a existência do mental não é a organização concreta ou os materiais de um determinado sistema mas uma organização causal abstracta que se mantém invariante de sistema para sistema. Na obra citada J.P.Dupuy analisa a cibernética nascida nos anos 40, nos Estados Unidos, numa pequena comunidade de neurobiólogos, matemáticos, engenheiros e economistas, entre os quais, numa primeira fase se encontram por exemplo John Von Neumann e Norbert Wiener e numa segunda fase Heinz von Foerster, William Ross Ashby, Warren McCulloch e Walter Pitts. A tese geral de Dupuy é que a cibernética não esteve à altura das suas próprias intenções e que em parte por isso as ciências cognitivas (que hoje frequentemente desconhecem as suas origens cibernéticas ou mesmo as renegam) teriam falhado o encontro com uma tradição filosófica que lhes teria sido bem mais útil do que a filosofia analítica, centrada na linguagem, nomeadamente a tradição fenomenológica.

[27] GARDNER 1985: 28-29.

[28] BECHTEL, ABRAHAMSEN & GRAHAM 1998: 37.

[29] Apenas a ideia de peças de um todo mais vasto explica o facto de ao lado de figuras mais óbvias tais como Noam Chomsky, Allen Newell, Herbert Simon, Marvin Minsky, John McCarthy e George

Assume-se neste trabalho que a TSI só pode ser compreendida, em primeiro lugar, como uma resposta à necessidade de esclarecimento conceptual que a convergência de áreas de estudo tão díspares gera. Em segundo lugar, propõe-se que a TSI deve ser comprendida como o desenvolvimento de uma das posições possíveis na resposta à seguinte questão: 'A psicologia é ou não importante para a filosofia?' Para Dennett a psicologia, no sentido geral de ciência cognitiva, é importante para a filosofia, mas considerar que não também constitui uma posição defensável. O facto de considerar que a psicologia é importante para a filosofia explica por um lado o interesse de Dennett pela psicologia como engenharia invertida (i.e. pela psicologia como compreensão dos mecanismos da cognição enquanto 'condições de possibilidade' da própria cognição) e, por outro lado, a sua concepção da filosofia da mente como consistindo em grande parte numa teoria da linguagem psicológica (i.e. numa teoria da linguagem mentalista e finalista usada na descrição de sistemas físicos com *design* complexo). Ainda assim, e mesmo se, devido à importância concedida à psicologia, a linguagem introspectiva e teleológica está por toda a parte nos trabalhos de Dennett, trata-se de uma situação ambígua. Essa linguagem é tomada pelo seu valor facial, mas não é, em útima análise, considerada fundamental. No que respeita ao estatuto do vocabulário mentalista, Dennett é um quineano, i.e. um fisicalista e um niilista do significado, alguém que considera a linguagem intencional como sendo apenas uma 'estratégia' de descrição da realidade física, sendo esta aquilo que realmente e fundamentalmente existe.

Miller, aparecerem usualmente, como fontes de grandes contributos para a ciência cognitiva (cf. por exemplo a lista de curtas biografias dos autores dos mais relevantes contributos para a ciência cognitiva que aparece em BECHTEL &GRAHAM 1998: 750-776), filósofos como H. Putnam, J. Searle, D. Dennett, J. Fodor, Paul Churchland e Patricia Churchland e teóricos como G. Boole, C. Babbage, A. Church e A. Turing, estes últimos devido ao seu 'trabalho preparatório' em lógica e teoria da computação. Uma noção da convergência (e divergência) temática que se pode hoje esperar encontrar em torno do estudo da cognição pode ser dada pela *MIT Encyclopedia of the Cognitive Sciences* (WILSON &KEIL 1999) que, sob os seis títulos gerais de 'Inteligência Computacional', 'Cultura, Cognição e Evolução', 'Linguística', 'Neurociências', 'Filosofia' e 'Psicologia' congrega entradas que vão desde Algoritmos, Autómatos, Redes, Propriedades de Sistemas Formais, Aquisição de Linguagem, Processamento de Linguagem, Memória, *Binding Problem*, Visão cega, Fisiologia da Cor, Fisiologia da Dor, Emoções, Afasias, Sonhos, Sono, Imagens Mentais, Sistema Límbico, Etologia, Psicologia Evolucionista ou Altruísmo, até entradas relativas a temas gerais da Filosofia como Lógica, Ética, Fisicalismo, Causação mental, Eu, Significação, Introspecção, Intencionalidade, Superveniência, *What-it's-like*, *Qualia*, Semântica dos mundos possíveis, Descartes, Hume e Kant.

Estrutura específica do trabalho.

Na Primeira Parte deste trabalho, intitulada *As Origens*, procurar-se-á esclarecer a genealogia do pensamento de Dennett dentro da filosofia analítica e ao mesmo tempo expor a teoria da mente avançada em *Content and Consciousness,* o seu primeiro livro. O Capítulo 1 (*Dennett e a teoria da mente em 1965*) é um capítulo histórico. Na origens da TSI encontram-se W. V. Quine e a indeterminação da tradução, H. Putnam e o funcionalismo, E. Anscombe e a análise das acções intencionais, C. Taylor e a crítica à explicação behaviorista do comportamento, bem como a Teoria da Identidade psicofísica, um fisicalismo reducionista que era o principal adversário contra o qual foram feitas, na época, as propostas da TSI. Aliás, a teoria da mente apresentada em *Content and Consciousness* é o resultado da conjugação de influências ainda mais alargadas, entre as quais estão Gilbert Ryle, o mestre de Dennett em Oxford, e Wittgenstein, o seu 'herói filosófico'. Na influência de ambos os autores em Dennett está em causa a definição de uma posição acerca da relação entre *descrições conceptuais* e *explicações causais* na teoria da mente. É no entanto da filosofia americana, e nomeadamente de Quine e Putnam, que Dennett recebe o núcleo da sua posição quanto ao mental: a intenção naturalista em epistemologia, a sofisticação do materialismo requerida pelo desenvolvimento do estudo da cognição e uma posição geral quanto ao estatuto do significado. De todas estas influências resulta um primeiro esboço da TSI, uma teoria da mente fisicalista e anti-reducionista.

Na Segunda Parte do trabalho, intitulada *O Modelo*, procurar-se-á analisar as propostas teóricas específicas de Dennett quanto à Intencionalidade (Capítulo 2 – *A posteridade do funcionalismo de Putnam: diferendos acerca da natureza da psicologia*), quanto à Consciência (Capítulo 3 – *Pensamentos conhecendo outros pensamentos*) e quanto a Pessoas e Acções (Capítulo 4 – *As Pessoas e as suas Acções: a filosofia da mente e os fundamentos da filosofia moral*). Embora se considere frequentemente que o funcionalismo cognitivista elimina a subjectividade do domínio mental – ou pelo menos aspectos essenciais da subjectividade como os *qualia* – assistir-se-á nestes três capítulos a uma reinstauração da subjectividade. Esta reinstauração far-se-á no entanto a partir de uma distinção de dimensões e nomeadamente de uma distinção entre três grupos de questões: questões relativas à interpretação, à intencionalidade e à racionalidade, questões relativas à

unificação e à incorrigibilidade, e questões relativas ao *self*, ao controlo, à identidade pessoal e à voluntariedade.

O Capítulo 2 é acerca de intencionalidade e tem duas partes distintas. Numa primeira parte trata-se da natureza representacional de alguma coisa física, numa segunda parte trata-se do conteúdo das representações. Na primeira parte, o principal contraponto da apresentação das posições da TSI é a Teoria Representacional da Mente (TRM) de J. Fodor. O capítulo inicia-se com a apresentação das ideias de J. Fodor, o funcionalista simbólico típico, que definiu a agenda da discussão em filosofia da psicologia a partir do funcionalismo putnamiano. Começar assim é um dispositivo argumentativo, já que em grande medida a TSI se define contra o funcionalismo forte da TRM e contra a teoria da explicação psicológica que esta envolve, mas é também um reflexo dos próprios princípios da TSI, nomeadamente do seu anti-representacionismo. De acordo com a TSI aquilo que existe para haver mentalidade em sistemas físicos são padrões (reais) e reconhecimento (desses padrões) por intérpretes e não um nível autónomo de realidade que seria intrinsecamente representacional e que teria forma de linguagem.

A segunda parte do segundo capítulo é especialmente importante para compreender o niilismo do significado, entendido como a ideia segundo a qual não existem máquinas semânticas, na sua relação com a explicação psicológica. A ideia central é a ideia de mundo nocional ou heterofenomenológico, a qual não é uma categoria causal. Repare-se que são, assim, necessárias pelo menos duas ideias distintas para compreender a posição sobre o significado desenvolvida na TSI, os '*padrões reais*' (uma noção relativa à natureza representacional de alguma coisa) e os '*mundos nocionais*' (uma noção relativa ao conteúdo das representações, i.e. ao representado na representação). A ideia de mundo nocional combina-se com uma posição teleofuncionalista acerca do conteúdo das representações. O teleofuncionalismo pretende reintroduzir a noção biológica de função na teoria da mente e é uma teoria basicamente externalista acerca do conteúdo das representações. O externalismo é a ideia segundo a qual é aquilo que está fora da mente que determina os significados e culmina na afirmação segundo a qual significar – ou representar – não é um funcionamento epistemicamente controlado pelos sistemas cognitivos nos quais veículos de significado acontecem, nem um funcionamento necessariamente relacionado com a consciência ou com a consciência de si. Há um certo consenso

entre os autores aqui seleccionados para a análise comparativa das soluções para o problema do conteúdo das representações. Todos admitem que o mundo (exterior) é o maior responsável pela fixação do conteúdo das representações internas aos sistemas de uma forma que passa ao lado do que o sistema sabe. Aquilo que se discute são interpretações específicas dessa solução e sobretudo as razões pelas quais a posição de Dennett fazendo apelo, do mesmo modo que as outras posições teleofuncionalistas, ao *design* dos sistemas cognitivos é afinal menos realista do que as restantes.

O Capítulo 3 é acerca de consciência. A forma mais elaborada da teoria dennettiana da consciência é o Modelo dos Esboços Múltiplos (MEM). Este pretende impossibilitar a concepção da mente como Teatro Cartesiano[30], aniquilar os *qualia* (qualidades da experiência sentida) e mostrar que os *zombies* (seres fenomenalmente apagados[31] embora comportamentalmente indistinguíveis de criaturas conscientes) são inconcebíveis. Embora teoria do conteúdo seja por princípio independente da teoria da consciência, esta teoria da consciência pressupõe a teoria do conteúdo analisada no capítulo anterior. O esquema orientador do capítulo é o seguinte. Por duas vezes Dennett propôs modelos funcionalistas de consciência, nos anos 70 e nos anos 90, em *Brainstorms* (BS) e em *Consciousness Explained* (CE). À proposta de modelos funcionalistas chama-se a parte construtiva da teoria da consciência. É com tais modelos que se pretende preencher o abismo nageliano entre fisiologia e fenomenologia[32]. A análise destes modelos, que são a parte construtiva da teoria da consciência, é o primeiro fio condutor do capítulo. Este primeiro fio condutor tem uma dupla face: se por um lado se trata de expor modelos, por outro trata-se de construir um caso favorável ao tipo de abordagem proposto. Aparece assim uma parte desconstrutiva da teoria da consciência que tem como núcleo uma determinada posição quanto aos *qualia,* alcançada em função da aplicação de princípios intelectualistas e verificacionistas ao estatuto epistémico dos relatos que sistemas cognitivos fazem dos seus próprios estados.

[30] O 'Teatro Cartesiano' é a ideia metafórica acerca do mental de acordo com a qual este é como um palco no qual tudo o que se está a passar numa dada mente num dado instante parece estar presente conjuntamente para um único Sujeito. Este supervisionaria assim a sua própria vida mental.

[31] A expressão 'fenomenalmente apagado' é retomada de PINTO 1999.

[32] Cf. NAGEL [1974] para uma influente descrição de um tal abismo entre o sentimento de ser e de experienciar (cuja pedra de toque é o *what-it's-like*) e aquilo que se sabe acerca dos mecanismos físicos subjacentes.

Os modelos funcionalistas permitem um posicionamento face a questões empíricas da ciência cognitiva (como as imagens mentais, os sonhos, a produção de linguagem e a visão cega) e a exposição deste posicionamento é o segundo fio condutor do capítulo. Para além disso, os modelos permitem também a estruturação de uma posição face às típicas abordagens da consciência pelos filósofos (usualmente apresentadas em torno de casos como o morcego de T. Nagel[33], Mary-a-neurocientista de F. Jackson[34], o *zombie* de D. Chalmers[35] ou as situações de inversão dos *qualia*[36]). A exploração das implicações dos modelos funcionalistas relativamente a *qualia* e *zombies*, ao Argumento do Conhecimento, ao *what-it's-like* distinguido por Nagel como marca da experiência consciente ou mesmo à experiência mental do Quarto Chinês de J. Searle[37], constitui o terceiro fio condutor do capítulo. Todo o Capítulo 3 assenta na suposição de que o MEM só pode ser compreendido quando contrastado ao mesmo tempo com o estatuto e os objectivos de outras teorias a que se chama teorias empírico-especulativas da consciência (como as de F. Crick e C. Koch[38], L. Weiskranz[39], G. Edelman[40], B. Baars[41] ou R. Penrose[42]) e teorias metafísicas da consciência (como as de D. Chalmers, F. Jackson, J. Searle ou T. Nagel). A posição de Dennett é de resto ela própria uma metafísica da consciência, embora o autor recue perante tal terminologia.

O Capítulo 4 é acerca de pessoas e de acções, dois temas que pressupõem o esclarecimento das questões do conteúdo e da consciência. O capítulo situa-se na intersecção da teoria da mente com a filosofia moral e a intenção que lhe preside é a de identificar e examinar as consequências da TSI e do MEM na filosofia moral. Propõe-se que as noções de pessoa e acção devem ser examinadas à luz da TSI e que as noções de Eu, Controlo e Consciência devem ser analisadas à luz do MEM. A análise da noção de pessoa faz-se em dois momentos: o primeiro visa as Condições de Pessoalidade,

[33] NAGEL 1974.
[34] JACKSON [1986].
[35] CHALMERS 1996.
[36] Cf. LOCKE [1690], CHALMERS 1996 Cap.7, BLOCK 1997a.
[37] SEARLE 1980.
[38] CRICK & KOCH 1990.
[39] WEISKRANZ 1986.
[40] EDELMAN 1989.
[41] BAARS 1988, BAARS 1996.
[42] PENROSE 1989, PENROSE 1994.

o segundo a Identidade Pessoal. Localizam-se três pontos focais: uma teoria da agência e da liberdade que conduz à ideia de auto-controlo meta-reflexivo, uma teoria da unidade e da identidade pessoal que conduz às ideias de unificação virtual e de auto-avaliação constitutiva da pessoalidade e da racionalidade próprias e uma teoria da deliberação que conduz à ideia de prudência naturalizada.

O Capítulo 5, intitulado *Questões aplicadas da TSI.Tipos de Mentes: mentes animais, artificiais e humanas*, conclui a Segunda Parte do trabalho. Nele são expostos mais claramente os princípios da eliminação da distinção apriorista entre o natural e o artificial, eliminação essa que conduz à consideração horizontal de todos os tipos de mentes – humanas, animais e artificiais. São ainda tratados dois problemas aplicados da TSI, envolvendo respectivamente a mentalidade animal e a mentalidade artificial: o problema da interpretação do comportamento na etologia cognitiva e o problema do enquadramento (*frame problem*) na Inteligência Artificial. Conclui-se com a localização de perturbações do funcionalismo que não conduzem no entanto a um afastamento deste, mas sim a uma objecção ao funcionalismo des-cerebralizado clássico, o funcionalismo simbólico, que fundamentou algumas das primeiras apologias da indistinção entre o natural e o artificial.

A Terceira Parte, intitulada *Os Problemas,* é constituída pelo Capítulo 6 (*Fisicalismo, Conteúdo e Consciência: da filosofia da mente à ontologia*) e tem como objectivo uma avaliação global do pensamento de Dennett. São consideradas eventuais incongruências de uma teoria da mente que se declara simultaneamente fisicalista e interpretativista e que tem a pretensão de abarcar o leque total das dimensões da subjectividade. O capítulo parte da localização um primeiro ponto crítico da TSI: o fisicalismo. Sugere-se que uma interpretação ontológica coerente das propostas de Dennett acerca da intencionalidade conduz a uma subversão do fisicalismo, nomeadamente porque o ponto de vista da física não sustenta a subsistência discreta das coisas e a sua individuação, mas apenas a particularidade e a localidade. Isto perturba o tipo de fisicalismo ligado à concepção hierárquica de natureza – que se insinua na TSI como prioridade absoluta concedida à Estratégia Física, i.e. à descrição do mundo feita pela física. No entanto, o facto de o ponto de vista físico não sustentar a subsistência discreta das coisas e a sua individuação confirma as razões pelas quais a TSI se opõe à Teoria Representacional da Mente, nomeadamente confirma a pertinência do anti-sentencialismo da teoria dennettiana. No contexto deste

esclarecimento sugere-se ainda que a TSI não pode abdicar totalmente da distinção entre intencionalidade intrínseca e intencionalidade atribuída se pretende explicar a sua noção fundamental, a noção de intérprete. É o intérprete que assegura, através da sua unidade, cada uma das estratégias de abordagem da realidade previstas na filosofia de Dennett (a Estratégia Física, a Estratégia do Design e a Estratégia Intencional). A unidade do intérprete é essencial para compreender a natureza da representação, não se identificando directamente com a instituição de uma representação de unidade (o Eu) num sistema cognitivo de agentes múltiplos. Corrige-se assim um defeito de perspectiva da TSI, a ideia segundo a qual o intérprete visa sempre outros sistemas cognitivos, sempre a partir de fora. A única maneira de não deixar um intérprete como resto na TSI é conceber o estatuto da interpretação de si próprio como mental realizada por um sistema cognitivo físico. Ora, é isso que constitui o entendimento genuíno. Feitos estes esclarecimentos e porque a questão da racionalidade está sempre subjacente à teoria da intencionalidade, retoma-se a questão da racionalidade. Concede-se que a TSI, embora sendo uma teoria quineana da interpretação supondo a racionalidade, não é obrigada a definir 'racionalidade'. São retomadas e justificadas as ideias de Dennett segundo a quais (1) a racionalidade é um conceito sistematicamente pré-teórico e (2) a racionalidade não pode ser comprovada ou infirmada empiricamente. Sublinha-se que a TSI pode ser considerada como apresentando um argumento duplo, um argumento a priori e um argumento empírico, a favor da impossibilidade de irracionalidade dos Sistemas Intencionais. De acordo com os próprios princípios da TSI conclui-se que não é legítimo supor que existe algo que seria uma verdadeira natureza ou um valor intrínseco da racionalidade e que a estipulação de condições necessárias e suficientes para a racionalidade, ou a abordagem do problema da racionalidade através de análise conceptual, fechariam a porta ao inquérito empírico de uma forma incompatível com os princípios metodológicos assumidos. Não é possível no entanto deixar de notar insuficiências na explicitação da 'suposição de racionalidade'. Procura-se nomear essas insuficiências e encontrar aquilo que a TSI deveria afirmar explicitamente acerca da racionalidade. O capítulo é concluído com um retomar do tratamento da consciência. A teoria dennetiana da consciência é 'intencionalista', i.e. é uma teoria do conteúdo. Para extrair as vantagens da teoria da consciência a partir dos seus defeitos (ou mesmo contradições) distingue-se antes de mais a teoria da subjectividade latente na Estratégia Intencional do 'absolutismo da terceira pessoa' que rege a teoria da consciência. Defende-se que a noção de subjectividade presente na TSI como teoria do conteúdo constitui base suficiente para discernir e corrigir os defeitos cientistas do tratamento da consciência no MEM.

Na Conclusão do trabalho, intitulada *Vale a pena fazer filosofia da mente?*, procede-se a um retorno às origens. Esse retorno tem vários sentidos. Num primeiro sentido trata-se de um retorno ao início do trabalho e às questões históricas e sociológicas que se colocam a propósito da filosofia da mente e da ciência cognitiva como domínios teóricos. Num segundo sentido trata-se de um retorno às origens da TSI na tradição filosófica amplamente considerada, embora partindo de Quine e de Putnam, os dois filósofos que mais marcaram a TSI em termos ontológicos e epistemológicos. Num terceiro sentido trata-se de um retorno à vida mental individual.

O presente trabalho guiou-se pela seguinte convicção: a pertinência de uma teoria filosófica é avaliável pelo esclarecimento que ela é capaz de produzir em relação ao conhecimento e compreensão previamente operantes, nomeadamente em áreas científicas. Quando se trata da ciência cognitiva o desafio é grande: são muitas as disciplinas e muitos os dados aos quais se reporta uma meta-teoria ou filosofia da cognição. No entanto, aquilo que ao longo deste trabalho se joga é um esclarecimento desse género, possibilitado pela Teoria dos Sistemas Intencionais proposta por D. Dennett.

Primeira Parte: As Origens

Capítulo 1

Dennett e a teoria da mente em 1965. O primeiro esboço da Teoria dos Sistemas Intencionais: princípios orientadores de Content and Consciousness.

O objectivo da teoria da mente é resumido da seguinte maneira por Dennett em *Content and Consciousness*[43]: a teoria deve ir da matéria e movimento ao conteúdo e propósito e depois 'voltar'. Este objectivo seria alcançado em C&C através da defesa de um *fisicalismo não reducionista*, reportado à *interpretação holista* de sistemas. De acordo com Dennett, estas duas posições básicas explicam a natureza do conteúdo e da consciência – os dois traços básicos do mental – evitando a situação clássica da teoria da mente, caracterizada como 'uma oscilação pendular do dualismo cartesiano para o materialismo de Hobbes, para o idealismo de Berkeley'. A teoria da mente apresentada em C&C não considera a consciência, o conteúdo e a finalidade como características não físicas. Não os identifica no entanto também – apesar de partir do pressuposto de que aquilo que basicamente existe é físico – com entidades ou propriedades espaço-temporais. As características mentais dependem segundo Dennett de uma determinada postura perante sistemas físicos a que Dennett então chama abordagem centralista e no futuro chamará Estratégia Intencional.

A dependência da mentalidade em relação a uma abordagem revela, em C&C, a característica postura interpretativista e deflacionária de Dennett já a postos, com tudo aquilo que ela envolve: uma oposição à reificação de significados e estados de consciência

[43] *Content and Consciousness* foi inicialmente uma tese de PhD apresentada em Oxford em 1965. Essa é a data tomada como ponto de referência neste capítulo, embora o livro resultante da tese seja referido daqui em diante como DENNETT 1969. Utilizar-se-á também a sigla C&C no corpo do texto.

e uma inimizade em relação aos *qualia*, à inefabilidade da experiência subjectiva e à suposta unidade do problema da consciência. Apesar de Dennett pretender apresentar uma teoria não reducionista do mental, Thomas Nagel não hesitará em classificar C&C como mais um exemplo da "recente onda de euforia reducionista", na qual inclui por exemplo as teorias defendidas por Jack J.C.Smart, David Lewis, David Armstrong e Hilary Putnam[44]. Em consequência da clareza da posição defendida, os adversários deste tipo de abordagem (por exemplo o próprio Nagel) estão já definidos. Na linguagem que Dennett virá a usar futuramente em *Consciousness Explained*, está iniciado o "desmantelamento do programa de protecção de testemunhas". Mais exactamente, está iniciado o desmantelamento da particular testemunha que seria em cada mente o 'observador daquilo que acontece', uma figura que, mais ou menos subrepticiamente mas com enorme persistência, se introduz na teoria científica ou filosófica do mental, assim como nas descrições introspectivas comuns. Eliminar o observador ou a testemunha equivale, na prática, a mostrar a ilegitimidade de postular na teoria da mente em geral – na teoria da linguagem, da percepção, da acção, do raciocínio, da imaginação, etc – elementos explicativos não analisáveis que possuam capacidades idênticas às dos sistemas globais a explicar. Este movimento redunda evidentemente na ausência de explicação. O primeira constação do teórico da mente deve então ser que uma vez deixada para trás a pessoa, o agente cognitivo global, não há mais ninguém.

Mas se a testemunha fôr eliminada ou dispensada, como Dennett pretende, o teórico da mente ficará a braços com "o problema de Hume": se não existe um sujeito-agente que faz com que o pensamento pense, aparentemente o pensamento acontece por si, sem supervisão, agenciamento ou intenção. Resta saber se na

[44] NAGEL 1974: 165. Cf. J.J. C. SMART 1963, *Philosophy and Scientific Realism*, David LEWIS 1966, *An Argument for the Identity Theory*, H.PUTNAM 1967, *Psychological Predicates*, D.M. ARMSTRONG 1968, *A Materialist Theory of the Mind*. Note-se que as posições que T. Nagel classifica como reducionistas incluem não apenas as versões da teoria da identidade defendidas por J. Smart, D. Armstrong e D. Lewis como também o funcionalismo proposto por H.Putnam como alternativa à teoria da identidade. J.J.C. Smart foi o primeiro filósofo a defender uma teoria da identidade. Esta é uma teoria materialista segundo a qual todo o estado mental é *idêntico* a algum estado físico, mesmo se os estados mentais e os estados físicos são *identificados* de maneiras diferentes (por exemplo, respectivamente, como dor e como disparo das fibras-C). D. Armstrong e D. Lewis mantêm a ideia de J.J.C. Smart segundo a qual apesar de aparentemente serem ou mentais ou físicas as propriedades dos estados são 'neutras' (*topic-neutral*) e especificam-na defendendo que para um estado mental ser um estado mental ele deve ter uma relação causal característica com outras ocorrências.

ausência de algo como um "motor primeiro" o pensamento se "moveria". O problema de Hume consiste precisamente em saber como é possível que existam movimentos de pensamento (transições cognitivas, inferências, associações) se os movimentos de pensamento não são acções de um autor. Hume, obrigado pela sua teoria da mente a explicar a maneira como as ideias pensam sozinhas, escolheu como solução o associacionismo, que é, para – Dennett, uma não-solução. Mas embora o associacionismo não seja a boa solução, Hume viu seguramente o bom problema, tendo além do mais, escolhido o rumo metodológico acertado para o seu tratamento, o rumo naturalista, justificado pelo princípio segundo o qual não temos mais acesso à essência dos fenómenos mentais do que à essência dos fenómenos físicos[45]. Felizmente, entre Hume e Dennett, algo de novo aconteceu no estudo da cognição. A vantagem do teórico contemporâneo em relação ao empirista clássico é o facto de o primeiro dispôr de uma teoria científica da cognição, a teoria computacional, que permite conceber a forma como as ideias 'pensam sozinhas', sem sentirem a falta da testemunha cartesiana e ao mesmo tempo evitando as armadilhas do associacionismo. Anteriormente à teoria computacional da cognição, «ninguém fazia a mínima ideia de como podiam processos meramente materiais implementar as leis (...) que governam uma mente semânticamente coerente (...) e ali estava o problema até Alan Turing ter tido o (...) melhor pensamento acerca de como a mente funciona que alguém teve até agora[46]». De facto, a exploração concreta, no estudo científico da cognição e da agência inteligente, da ideia de *sistema simbólico físico* – uma ideia surgida na sequência do pensamento de Alan Turing sobre autómatos abstractos[47] – constitui em grande medida para Dennett, quando escreve C&C[48], o pano de fundo dos problemas relevantes da

[45] «For me it seems evident that the essence of mind being equally unknown to us with that of external bodies, it must be equally impossible to form any notion of its powers and qualities otherwise than from careful and exact experiments (...) But if this impossibility of explaining ultimate principles should be esteemed a defect in the science of man, I will venture to affirm that 'tis a defect common to it with all the sciences (...) None of them can go beyond experience, or establish any principles which are not founded in that authority» (HUME [1739]: 44-45). As alusões a David Hume e ao problema de Hume para caracterizar a situação de partida da teoria da mente são frequentes na obra de Dennett.

[46] FODOR 1994b: 296.

[47] Alude-se aqui às Máquinas de Turing.

[48] Cf. BECHTEL, ABRAHAMSEN & GRAHAM 1998:1, *The Life of Cognitive Science*, para uma exposição da cena contemporânea do estudo científico da cognição e dos agentes inteligentes.

mente e do conhecimento. O tratamento científico da cognição como informação e computação começa a indiciar que o discurso mentalista acerca de sistemas físicos não está necessariamente ligado a descrições introspectivas autoritárias de substâncias imateriais. Por exemplo os computadores mostram a possibilidade de sistemas físicos inteligentes (ou pelo menos susceptíveis de descrições mentalistas) sem o envolvimento de uma regressão infinita de observadores ou testemunhas. É essa a lição que deve ser aplicada ao caso humano.

É claro que a ilegítima testemunha do que se passa na mente (o "fantasma na máquina" na expressão de Gilbert Ryle) não tem por hábito aparecer claramente na teoria cognitiva, que não prevê usualmente uma alma cartesiana placidamente introduzida. A testemunha aparece subrepticiamente em expressões como "interpretações de estímulos", "reconhecimentos", "comandos", "analisadores" de vários tipos. Basta percorrer os textos da psicologia cognitiva ou da neurofisiologia para encontrar o fantasma na máquina nesta sua variedade de disfarces. O problema é que alguém que procura acabar com a "protecção das testemunhas," como Dennett se propõe fazer, acaba certamente com uma dúvida em relação a si próprio. Dennett virá a enfrentar claramente esse problema: «Sou um behaviorista? Searle e Nagel sempre insistiram que sim, agora (Patricia) Churchland vem dizer o mesmo»[49]. Caracterizando Dennett, B. Dahlbom afirma: «Dennett é um behaviorista e no seu behaviorismo conjuga ideias de Wittgenstein, Ryle e Quine, bem como da psicologia experimental. O behaviorismo é o funcionalismo aplicado a organismos, a ideia segundo a qual um organismo é uma máquina, produzindo comportamento com uma mente-cérebro como sistema de controlo. É uma versão do naturalismo, tratando os seres humanos como organismos biológicos e indo buscar as contribuições do evolucionismo para a nossa auto-compreensão (...) O behaviorismo e o funcionalismo de Dennett estão ligados ao verificacionismo, a ideia segundo a qual onde não existe evidência que permita decidir uma questão, não existe questão (...) Mas o behaviorismo também é a mais particular aplicação do verificacionismo à linguagem psicológica, aquilo a que os filósofos em Ox-

Embora a hipótese dos sistemas simbólicos físicos só tenha sido formulada com este nome tardiamente (cf. NEWELL 1980, *Physical Symbolic Systems)*, o processo que lhe dá origem começa várias décadas mais cedo.

[49] DENNETT 1993a: 210.

ford costumavam chamar "behaviorismo lógico"»[50]. A eliminação de testemunhas conduz portanto ao behaviorismo. Mas a teoria proposta em C&C define-se exactamente por oposição ao behaviorismo psicológico. Saber em que sentido é Dennett um behaviorista é, assim, complicado. Ele sempre se opôs por exemplo ao behaviorismo periferista skinneriano, que não leva em conta os resultados das teorias computacionais da cognição, mas de facto a sua primeira inspiração são as ideias de behavioristas lógicos como L. Wittgenstein e G. Ryle, e a ideia chave do behaviorismo lógico, segundo a qual as entidades mentais são construções lógicas a partir de eventos comportamentais, nunca anda longe.

Existe portanto algo no behaviorismo que não deve ser abandonado. O problema é que a pedra de toque do behaviorismo psicológico é o afastamento de toda a contaminação de teleologia e intencionalidade na descrição de comportamentos e a causa de Dennett como filósofo da psicologia é precisamente a *defesa das descrições intencionais e teleológicas* (é nisso se traduz o centralismo de C&C e posteriormente a Estratégia Intencional). Mas a causa intencional e teleológica apenas obscurece provisoriamente o compromisso behaviorista: a última palavra de Dennett é que a lei do efeito – de acordo com a qual as associações estímulo-resposta recompensadas serão repetidas – não desaparecerá[51] na teoria da cognição. A lei não desaparecerá mas terá que ser re-instalada: o âmbito do behaviorismo defendido por Dennett é o *interior dos sistemas cognitivos físicos*, no qual se encontram por exemplo cérebros, e não os comportamentos globais de sistemas como animais ou humanos.

É conveniente sublinhar desde já um limite da teoria dennettiana do mental: ela não permite, segundo a maioria dos críticos, diferenciar um *zombie,* um ser desprovido de vida mental em primeira pessoa que se comporta de modo previsível intencional e teleologicamente (o que inclui obviamente a possibilidade de afirmar "eu sou um ser consciente" e de descrever os conteúdos da consciência), de um ser consciente. Esta sempre foi a principal crítica dirigida a Dennett. Segundo Nagel, Dennett elabora uma teoria da mente em terceira pessoa e depois pergunta: «Poderia um sistema físico que satisfizesse esta descrição realmente não ser consciente?- Alguma coisa necessária para a consciência estaria a faltar? E a res-

[50] DAHLBOM 1993a: 4-5.

[51] A expressão é o título de um artigo de Dennett (DENNETT 1978c, *Why the Law of Effect will not Go Away).* A Lei do Efeito foi assim chamada por E. Thorndike, psicólogo behaviorista.

posta dele é: Não – isto é tudo o que a consciência é»[52]. Como se vê, a situação faz temer o pior: e se Dennett é ele próprio um *zombie*? Afinal aquilo que "sabemos" sobre a consciência experienciamo-lo em primeira pessoa e pode dar-se o caso de tal experiência não existir em Dennett. Se assim fosse, nunca o saberíamos. Como nota D.Chalmers, isso não ajuda nada[53].

Além deste possível "limite", todos os temas futuros de Dennett estão presentes em C&C, nomeadamente a dependência da teoria da consciência relativamente à teoria do "conteúdo", a relação entre significado, função e evolução, a desconstrucção dos equívocos no tratamento da consciência, a não unidade desta e a teoria das pessoas e suas acções intencionais. Alguns erros embaraçosos estão ainda presentes, nomeadamente a suposição de uma linha divisória entre consciência e não consciência (uma *awareness line*), além de admitidos exageros na negação da existência de imagens mentais. C&C contem em gérmen a futura obra de Dennett e por isso partirei da análise das teses que aí se encontram, procurando pôr em relevo a orientação particular que elas representam em filosofia da mente, com ocasionais referências a textos posteriores. É o próprio Dennett que autoriza este tratamento, na medida em que considera que os principais livros que posteriormente escreveu constituiram sucessivas retomas dos dois tópicos centrais da filosofia da mente, o conteúdo e a consciência[54]. Assim, a primeira parte de *Brainstorms* (1978) retoma a questão do conteúdo, a segunda parte a questão da consciência. *Intentional Stance* (1987) tem por objecto a questão do conteúdo e *Consciousness Explained* (1991) a questão da consciência. Os livros posteriores, *Darwin's Dangerous Idea – Evolution and the meanings of life* (1995), *Kinds of Minds – Towards an Understanding of Consciousness* (1996) e *Brainchildren – Essays on Designing Minds* (1998) são concretizações, aplicações e esclarecimentos, nomeadamente em termos de biologia e IA, das grandes teses acerca do conteúdo e da consciência delineadas nos primeiros livros. *Elbow Room* (1983) é totalmente dedicado às questões da acção e da pessoalidade, que pressupõem as teorias do conteúdo e da consciência.

O trajecto teórico de Dennett está bastante próximo do trabalho empírico sobre cognição. No entanto, uma tal proximidade não

[52] NAGEL 1995b: 88.

[53] Cf. CHALMERS 1996:190 e PINTO 1999: 123, para comentários às consequências da posição de Dennett segundo a qual todos somos *zombies*.

[54] DENNETT 1998y, *Self Portrait*.

é necessariamente acompanhada por um materialismo como o defendido por Dennett[55], apenas torna mais premente o esclarecimento das implicações metafísicas do trabalho científico sobre cognição. Aliás, o próprio Dennett começará a tornar-se no filósofo-cientista cognitivo que é hoje apenas após C&C. Apesar disso, como afirma numa narração do seu itinerário teórico, o fascínio por mecanismos e pelo funcionamento destes, conjuntamente com a convicção de que o dualismo representa um beco sem saída na teoria da mente, fizeram com que Dennett desde sempre se decidisse a trabalhar a partir do ponto de vista de terceira pessoa, i.e. a partir das ciências naturais. Um tal gosto e uma tal estratégia, a que se juntou a influência de Quine e de Ryle–que eram, nas suas palavras e «embora (se) sentisse movido pelo (...) desacordo com eles», a «fonte da segurança intelectual que (...) tinha em virtude da (...) profunda concordância quanto ao que eles pensavam ser a filosofia[56]» – definem a figura de Dennett como filósofo.

Passar-se-á em seguida à descrição da teoria da mente apresentada em C&C, seguindo aquele que será um posicionamento geral deste trabalho: tentar fazer aparecer a partir da obra de D. Dennett os problemas da filosofia da mente, sem que as suas soluções sejam consideradas centrais ou prioritárias, embora se procure defendê-las.

1.1 *A partir de Wittgenstein e Ryle, Quine e Putnam.*

A dissertação de doutoramento que virá a tornar-se o primeiro livro de Dennett foi escrita em Oxford, em 1965, com a supervisão de Gilbert Ryle. A agenda teórica dos filósofos em Oxford era então ainda em grande medida marcada pelo programa de psicologia filo-

[55] Um "Materialismo de tipo A" ou defesa da superveniência lógica da consciência ao mundo físico, por razões funcionalistas ou eliminativas, de acordo com a classificação que David Chalmers faz das posições possíveis quanto à experiência consciente (CHALMERS 1996: 165). D. Chalmers classifica como materialistas de tipo A, além de D. Dennett, D. Armstrong, D. Lewis, G. Ryle, F. Dretske, D. Rosenthal, J.J.C.Smart e outros. Aliás, Chalmers considera que a alternativa que se coloca hoje a qualquer teórico da consciência é fundamentalmente uma escolha entre o materialismo de tipo A, como o defendido por Dennett, e a posição a que chama tipo C e que inclui vários tipos de dualismo de propriedades. De acordo com as posições de tipo C, o materialismo é suposto ser falso e as propriedades fenomenais ou proto-fenomenais são tomadas como irredutíveis.

[56] DENNETT 1998y: 356-357. Resta saber se é de facto possível tomar conjuntamente a maneira como Ryle e Quine entendem a filosofia, pois trata-se de entendimentos deveras diferentes.

sófica estabelecido por Ryle[57] em *The Concept of Mind*[58], bem como pela filosofia linguística[59]. Em *The Concept of Mind* Ryle propunha como método filosófico a determinação da geografia lógica de conceitos. A obra é aliás uma aplicação do método aos conceitos mentais, argumentando Ryle que certas utilizações de conceitos mentais violam regras lógicas. É a esta violação que corresponde a ideia de erro categorial que Ryle ilustra com a célebre história do estrangeiro a quem são mostrados os *colleges*, bibliotecas e museus de Oxford ou de Cambridge e que no fim pergunta "E onde é a Universidade?"[60]. Erros categoriais resultam da transgressão do âmbito legítimo da aplicação de conceitos. Na teoria da mente eles são muito frequentes e têm consequências teóricas perversas.

A psicologia filosófica de Ryle visava delimitar a natureza do mental sem presumir qualquer separação entre eventos mentais, por um lado, e eventos físicos, por outro. Ryle afastava assim não apenas o dualismo mas também a necessidade de escolha entre monismo materialista e monismo idealista. Idealismo e materialismo eram, para Ryle, respostas a uma má questão sobre a natureza da realidade. A questão era má pois pressupunha a disjunção "ou existem mentes, ou existem corpos", restando depois saber em que sentido iria a redução. A necessidade de escolha entre mente e matéria aparecia na sequência de um mito comum de acordo com o qual haveria dois tipos de existência rigorosamente separáveis, uma existência física (no espaço e no tempo, que consiste em matéria ou é função de matéria) e uma existência mental (no tempo e não no espaço, que consiste em consciência ou é função da cons-

[57] Cf. RYLE 1949, especialmente o Capítulo x (*Psychology*). Para uma avaliação da importância da psicologia filosófica rylena, cf. GUTTENPLAN 1994b. Cf. também BYRNE 1994: 135, acerca das estranhas intenções de G. Ryle, que estaria, ao escrever *The Concept of Mind*, muito mais interessado em teoria do método filosófico do que em filosofia da mente.

[58] «Ryle's *The Concept of Mind*, when it came out, made *the* basis for discussion for a long time: it was quite difficult to get away from it» (DUMMETT 1993: 168). M. Dummett (n. 1925) refere-se nesta passagem, é certo, a um tempo anterior à chegada de Dennett a Oxford.

[59] Para uma caracterização "da geração da linguagem comum" ou da filosofia linguística de Oxford (título sob o qual podem ser considerados autores como G. Ryle, J. Austin, A. J. Ayer) , cf. DUMMETT 1978: 341. Dummett vê, aliás, muito pouco de comum no pensamento dos autores mencionados, à parte o facto de terem estado a certo a certo momento (antes da 2ª Guerra Mundial) unidos contra o *status quo* da filosofia oxfordiana representada por H. Joseph e H. Pritchard e terem sido mais ou menos influenciados ou por G. Moore, ou por L. Wittgenstein ou pelos positivistas lógicos. Dummett aponta ainda como razão para a devoção ao culto da linguagem comum, considerada como o caminho para a análise de conceitos, um desgosto ou falta de vocação dos filósofos em relação à lógica matemática.

[60] RYLE 1949:17.

ciência). Este mito subjaz à "doutrina oficial"[61] segundo a qual todo o humano tem uma mente e um corpo, os corpos estão situados no espaço, as mentes não, os corpos estão sujeitos a leis mecânicas, as mentes não, o funcionamento do corpo é ou pode ser público enquanto o funcionamento da mente é privado e só conhecido directamente pelo seu possuidor. Por tudo isto, qualquer pessoa teria duas histórias, uma história pública e uma história privada, que só ela pode contar. Para Ryle, a "doutrina oficial" é um gigantesco erro categorial.

Defendendo que mente e matéria estão em diferentes categorias mas não são tipos ou espécies de existência separáveis, Ryle pretendia tornar ociosa a identificação ou não identificação do mental e do físico. Apenas quando dois termos pertencem à mesma categoria é apropriado construir conjunções que os incorporam. Assim, «"*she came home in a flood of tears and a sedan chair*" é uma piada conhecida, que vive do absurdo da conjugação de termos de tipos diferentes»[62], cuja disjunção seria igualmente absurda. O mesmo se passa com a disjunção "ou existem mentes ou existem corpos". Ryle considera perfeitamente possível afirmar num tom de voz lógico que existem mentes e noutro tom de voz lógico que existem corpos, sem qualquer compromisso com diferentes *espécies* de existência, apontando apenas dois *sentidos* diferentes de "existir". Os erros categoriais envolvidos no mito do fantasma na máquina resultam da suposição de espécies de existência diferentes e da transposição de modelos de uma espécie para a outra, gerando por exemplo as teoria para-mecânicas do mental. Estas consideram a psicologia como a contraparte mental da mecânica e lidam com questões psicológicas por meio de categorias idênticas às categorias do "outro mundo" (como coisa, atributo, estado, processo, mudança, causa). Exemplos dessa teoria para-mecânica em *The Concept of Mind* são a ideia de volições como "empurrões internos ocultos" (*occult inner thrusts*) ou questões de fios e roldanas (*wires and pulleys questions)* acerca da vida mental como "Quando uma pessoa imagina uma montanha ou ouve uma melodia mentalmente, *onde está* a montanha que está a ser imaginada e a música que está a ser ouvida?".

[61] À qual Ryle chama também o Dogma do Fantasma na Máquina (RYLE 1949:17). Mentes seriam, de acordo com essa perspectiva, "*spectral machines*", "*bits of not-clockwork*", "*ghosts harnessed to machines*", "*another engine inside the human body*".

[62] RYLE 1949:23.

Dada a centralidade das crenças na vida mental, um objectivo importante de Ryle é mostrar que quando pessoas são descritas por meio de termos mentais como "acreditar" ou "saber", não estão a ser descritas ocorrências mentais interiores que seriam as causas do comportamento observável, mas sim habilidades, tendências, disposições comportamentais. As crenças são tendências para dizer ou fazer coisas, coisas que incluem declarar essas crenças, sendo portanto caracterizadas por referência ao comportamento. Esta ligação entre comportamento e crenças é conceptual e não causal. Ryle defende que não existem processos mentais ocultos por trás das acções públicas das pessoas, que explicariam a sua "inteligência". Características como inteligência e carácter estão patentes, não actuam por trás do pano. Esse é o princípio que guia as análises da vontade, das emoções, da sensação, da imaginação, do pensamento e do auto-conhecimento em *The Concept of Mind*. A suposição oposta constitui o Mito do Intelectualismo, alvo do ataque de Ryle. O intelectualismo é a inversão que consiste em tomar fenómenos como pensamentos privados e consulta de regras como modelo do comportamento inteligente. Ora Ryle considera que existem competências anteriores a esses fenómenos, mais fundamentais do que eles, e que pensar que tais competências são elas próprias processos intelectuais conduziria a uma regressão infinita. De acordo com Ryle, o que está primeiro é *know-how* tácito, não expressões intelectuais explícitas. Uma das consequências deste anti-intelectualismo é a ideia segundo a qual quando dizemos que uma pessoa está consciente de alguma coisa parte do que queremos dizer é que ela está "pronta a relatar ou reportar algo" sem investigação especial. Esta ideia passará directamente para Dennett.

Embora Dennett tenha acabado por se instalar em Oxford, onde Ryle, com o acima esboçado programa de psicologia filosófica, era, ou tinha sido, "rei"[63], ele partira para Inglaterra tendo um outro "herói filosófico", Wittgenstein[64]. No entanto, a prática da filosofia tal como era entendida pelos (muitos) auto-declarados wittgensteinianos que vem a encontrar em Oxford não o satisfaz. Por isso, Dennett desiste de tentar ser um wittgensteiniano, limi-

[63] Cf. DUMMETT 1997: 34.

[64] «My debt to Wittgenstein is large and longstanding. When I was an undergraduate, he was my hero, so I went to Oxford, where he seemed to be everybody's hero. When I saw how most of my fellow graduate students were (by my lights) missing the point, I gave up trying to be a wittgensteinian» (DENNETT 1991: 463).

tando-se a procurar pôr em prática aquilo que pensava ter aprendido nas *Investigações Filosóficas*. É uma questão delicada saber o que encontra Dennett nessa leitura. Nomeadamente, as *Investigações* estabelecem e praticam a precedência do problema semântico em relação ao problema epistemológico[65] enquanto Dennett pratica directamente epistemologia das ciências da mente. Existe no entanto, sem dúvida, um denominador comum a Dennett e ao Wittgenstein das *Investigações:* o ataque à reificação de estados de consciência. Saber em que medida é ainda wittgensteiniano o modo que o próprio Dennett encontrará de prosseguir esse ataque, constitui um problema em aberto ao longo deste trabalho, e se a terapia conceptual wittgensteiniana se compromete a deixar os factos tal como estão, o mesmo não se pode dizer da filosofia de Dennett.

A discordância de Dennett relativamente aos filósofos que se consideram wittgensteinianos é, como é óbvio, principalmente metodológica. A pura análise conceptual não parece a Dennett um método frutífero em teoria da mente. A tarefa wittgensteiniana do filósofo, a tarefa terapêutica de "mostrar à mosca a saída da garrafa" não passa certamente, para os wittgensteinianos, por investigações empíricas acerca da "garrafa" ou da "mosca". Ora, ao contrário dos wittgensteinianos, Dennett decide que deve começar por conceber como é que, em geral, sistemas cognitivos físicos podem existir e funcionar antes de confiar em intuições conceptuais acerca da mente[66]. Tal decisão conduz Dennett a admitir "hipóteses

[65] Fazem-no «baseando-se justamente no facto de que é a pré-existência de uma linguagem que permite a formulação de proposições de carácter cognitivo, as quais para terem um lugar no conjunto das proposições da referida linguagem necessitam de utilizar com sentido os termos dessa mesma linguagem» (ZILHÃO 1993: 96).

[66] Cf. DENNETT 1998: 356. Cf. também DENNETT 1995j: 59: «I was completely frustrated by the work that was being done by philosophers because they did not know anything about the brain and they did not seem to be interested». Para uma experiência semelhante de Inglaterra de uma americana menos sensível ao interesse de Wittgenstein para a filosofia da mente, cf. Patricia CHURCHLAND 1995: 21: «that book (P. Churchland refere-se a *Word and Object* de Quine) was one of the few things in philosophy that made any sense to me at that time. Several other graduate students (...) were utterly contemptuous of the later Wittgenstein in particular and of the so called ordinary language philosophy in general. (...) Apart from being willfully obscure, Wittgenstein was at best insensitive to the possible role of scientific data in philosophical questions, and at worst, hostile to science». O confronto entre a inspiração quiniana e a inspiração wittgensteiniana é comum aos filósofos da mente da geração de Dennett e Churchland. Note-se que Quine aparece como a fonte de uma inclinação naturalista muito mais disseminada na filosofia americana do que na filosofia inglesa. Segundo Dummett (DUMMETT 1993: 168), pouca gente em Inglaterra se apercebia da importância do trabalho de Quine. O próprio Ryle promoveu o desprestígio de Carnap, que foi o mestre dos filósofos americanos da geração de Quine (DUMMETT 1997: 34).

mecânicas" sobre o funcionamento de sistemas cognitivos na própria filosofia, o que não se enquadra nem na ideia wittgensteiniana nem na ideia ryleana de filosofia. A influência de Ryle, o suposto behaviorista, em Dennett deve-se assim antes ao facto de Ryle ser, «num importante sentido, um fenomenólogo introspeccionista»[67]. E o mesmo poderia ser afirmado do Wittgenstein das *Investigações*, o outro suposto behaviorista.

Em C&C é já visível o afastamento não apenas da análise conceptual como também da ideia segundo a qual a linguagem vulgar delimita o âmbito do pensamento filosófico. Para Dennett, «a ordem só poderá ser trazida para este caos (*a teoria da mente*) se abandonarmos a convicção de que o uso vulgar da linguagem é conceptualmente são (...) logo que abandonarmos esta convicção, não poderemos mais apoiarmo-nos na totalidade do uso para sabermos o que está por trás das noções de apercebimento (*awareness*) e consciência (*consciousness*), pois nada de consistente está por trás da totalidade do uso. Quando procuramos distinções que servem para marcar sentidos consistentes e separáveis destas palavras, tudo o que podemos encontrar são distinções de função e para capturar estas distinções de função precisamos de um quadro mecanicista plausível»[68]. O problema de Wittgenstein e de Ryle enquanto filósofos é que eles foram de certa forma hostis à ideia de investigação científica da mente, e de acordo com a sabedoria standard saída da revolução cognitiva ter-se-ia ultrapassado as suas análises "grosseiramente a-científicas" do mental. Mas para Dennett, isto não é verdadeiro: «Devemos tolerar a sua frequentemente frustrante incompreensão daquilo que são boas questões científicas, a sua quase total ignorância de biologia e ciência do cérebro pois eles conseguiram fazer observações profundas e importantes que só agora começamos a ser capazes de apreciar»[69]. As observações profundas de Wittgenstein e Ryle, não são decerto observações de filósofos naturalistas, ou fisicalistas. Para além da hostilidade em relação à investigação científica do mental, Ryle considerava a própria questão "Será o mundo em última análise físico?" conceptualmente confusa e Wittgenstein nunca se mostrou particularmente interessado na exploração filosófica do fisicalismo. Wittgenstein e Ryle

[67] DENNETT 1998: 191, n.10.
[68] DENNETT 1969: 130.
[69] DENNETT 1996: 169.

são dois anti-interioristas muito especiais, muito diferentes dos filósofos da mente materialistas que se lhes seguiram, muitos dos quais vieram a apontá-los como precursores. O que há de especialmente importante (e comum) no pensamento de Wittgenstein e de Ryle é o facto de ambos procurarem estabelecer o estatuto daquilo que se pode dizer ao "nível pessoal" – o nível em que "nós" descrevemos a nossa vida mental – relativamente ao nível sub-pessoal, do discurso científico-natural, legitimamente mecânico, sobre cognição. É o facto de o nível pessoal ser por princípio insusceptível de descrições *mecânicas* que justifica o desinteresse de Wittgenstein e de Ryle pelas investigações empíricas sobre o mental.

Uma vez que a teoria naturalista da cognição não é o território de Wittgenstein ou de Ryle, as preciosas observações que os dois autores facultam ao teórico naturalista da mente só podem dizer respeito às descrições de nível pessoal. Wittgenstein, nas *Investigações Filosóficas*, traça imagens esclarecedores da descrição de nível pessoal: as *Investigações* consistem afinal em verificações e descrições atentas dos estados de consciência nas mais variadas circunstâncias da vida mental, sem presunção de se possuir assim alguma autoridade epistémica ou acesso directo à estrutura do mundo. Pelo contrário, aquilo que Wittgenstein persegue continuamente é o ponto em que aquilo que se pode dizer numa descrição mentalista pára e não há mais nada a dizer, acentuando o quanto as condições de aplicação dos conceitos para experiências interiores se afastam de uma hipotética situação de auto-observação com intenções epistémicas, cujos resultados seriam expressos em linguagem referencial. Por exemplo, se eu digo "Eu tenho uma dor", essa expressão, na célebre formulação de Wittgenstein, substitui o grito, não o descreve nem explica (*Investigações Filosóficas* §244). As asserções de experiência interior são exteriorizações (*Äusserungen*) e não expressões de um conhecimento directo do interior. Traduzido para a terminologia de Dennett, nada do que está a passar-se a nível sub-pessoal é cognitivamente acessível ao sujeito que é capaz de descrever os acontecimentos mentais a nível pessoal.

Ryle, por seu lado, faz, ao longo de *The Concept of Mind*, uma caricatura do uso de metáforas para-mecânicas em teoria da mente. Ironiza acerca da maneira de falar dos teóricos da mente que parecem envolvidos em anatomia mental especulativa, e dá como exemplos das absurdas "*wires and pulleys questions*" assim geradas, questões como "Onde estão armazenadas as experiências passadas?", "Como

é que uma mente captura (*grasps*) a realidade física exterior?", "Como é que uma pessoa pode ouvir música dentro da sua cabeça, só para si, se não há música lá dentro para ouvir?", etc. Ryle compara o estatuto destas questões, sem dúvida vulgares em contextos psicológicos e epistemológicos[70], com questões mais evidentemente absurdas como: «Quantos actos cognitivos cumpriu John Doe antes do pequeno almoço e como foi cumpri-los? Foram cansativos? Ele apreciou a passagem das premissas à conclusão?»[71]. Perguntas como estas são esquisitas e infrutíferas. John Doe não sabe responder às perguntas dos epistemólogos acerca dos seus próprios juízos e raciocínios. Mas, com toda a certeza, só ele mesmo poderia dizer como foi passar por essas experiências. Infelizmente, como nota Ryle, ele não costuma contar essas coisas. É importante compreender porque é que estas questões estão mal formuladas e propiciar tal compreensão é o objectivo de Ryle, que basicamente considera as questões mal formuladas devido ao compromisso tácito com o "fantasma na máquina" e à transposição de modelos gerada pelo (pressuposto e ilegítimo) dualismo.

O que pode então fazer o teórico da mente que não assuma tais compromissos? É neste ponto que entra o behaviorismo de Ryle, que afirma: «a tendência geral deste livro será sem dúvida e sem que isso seja um mal, estigmatizada como behaviorista. Por isso é pertinente dizer alguma coisa acerca do behaviorismo. O behaviorismo foi inicialmente uma teoria dos métodos correctos para a psicologia científica (que sustentava que) (...) as teorias dos psicólogos deveriam ser baseadas em observações publicamente comprováveis e repetíveis e em experiências. Mas as supostas contribuições da consciência e da introspecção não são publicamente comprováveis. [*no entanto*] Romancistas, dramaturgos e biógrafos sempre se satisfizeram com a exibição dos motivos, pensamentos, perturbações e hábitos das pessoas através da descrição das suas palavras, actos, imaginações, das suas expressões, gestos e tons de voz. Concentrando-se naquilo em que Jane Austen se concentrou, os psicólogos começaram a ver que esta era a própria matéria dos seus assuntos, e não uma mera prisão exterior»[72]. Em suma, Ryle pensa que se o mito cartesiano não gerou o melhor tratamento da mente, também os psi-

[70] E que são, pelo menos a primeira e a terceira, mais legítimas do que Ryle pensava.
[71] RYLE 1949: 275.
[72] RYLE 1949: 308.

cólogos behavioristas sucumbiram às motivações erradas. O erro comum a cartesianos e behavioristas é, segundo Ryle, a crença na especificidade isolável dos fenómenos mentais (depois considerados como um mito pelos behavioristas, e como insusceptíveis de investigação científica pelos cartesianos). O problema da psicologia científica behaviorista é, assim, o facto de na própria palavra "psicologia" ter ficado incrustrado o mito dos dois mundos. Uma vez afastado o mito dos dois mundos fica a conclusão de que os comportamentos humanos não são meros sintomas ou expressões de algo de muito importante que se passa "lá dentro" e que seria a verdadeira matéria do estudiosos da mente: «Dizer alguma coisa com significado não é (...) fazer duas coisas, nomeadamente dizer alguma coisa em voz alta ou na nossa cabeça e ao mesmo tempo, ou pouco tempo antes, passar por algum outro processo sombrio»[73]. De forma muito semelhante, Wittgenstein defenderá nas *Investigações* que verbos mentalistas como "significar" e "compreender" não devem ser pensados como referindo processos e estados interiores. Quando se decidiu falar assim já se decidiu tudo e decidiu-se mal [74].

Em suma, na medida em que atacam o interiorismo imaterialista Wittgenstein e Ryle "apoiam" o naturalismo. Não é no entanto de todo esclarecedor classificá-los simplesmente como behavioristas. Por um lado, porque o "behaviorismo" foi originalmente (ou pelo menos foi melhor delimitado como) um termo para uma recomendação metodológica em psicologia empírica e nem Witgenstein nem Ryle se interessaram particularmente por inquéritos empíricos acerca do mental[75]. Por outro lado, o behaviorismo psicológico está ele próprio muito mais comprometido com o mito dos dois mundos do que se poderia supor: os behavioristas consideram, como os cartesianos, que há dois mundos, apenas escolhem de forma diferente qual deles eliminar. Ora, ao contrário do que se passa com o

[73] RYLE 1949: 279.
[74] Cf WITTGENSTEIN 1953, §308.
[75] Cf. por exemplo WITTGENSTEIN 1953 §109: «Correcto foi não deixar a nossa investigação ser uma investigação científica. (...) E não devemos produzir nenhuma espécie de teoria. Na nossa investigação não deve haver nada de hipotético. Toda a explicação tem que acabar e ser substituída apenas pela descrição. E esta descrição recebe a sua luz, isto é, a sua finalidade, dos problemas filosóficos. É claro que estes não são problemas empíricos, a sua solução estará antes no conhecimento do modo como a linguagem funciona, de maneira a que de facto esse modo seja reconhecido (...) Estes problemas serão resolvidos não pela adução de novas experiências mas pela compilação do que é há muito conhecido. A Filosofia é um combate contra o embruxamento do intelecto pelos meios da nossa linguagem».

behaviorismo, as teorias do mental de Ryle e de Wittgenstein não redundam na exclusão ou eliminação do mental nem em desatenção à linguagem psicológica. Assim, aquilo que Ryle e a Wittgenstein exemplificam aos olhos de Dennett é um modo de descrever o nível pessoal que não é, como diria Ryle, nem cartesiano nem hobbesiano, nem imaterialista nem materialista. Essas maneiras de tratar a mente são igualmente – e erradamente – mecânicas; ao nível pessoal apenas é possível constatar a vida mental tal como esta se dá, não se pode pretender ter a iluminação na posse que corresponderia à posição do "autor" da vida mental. É esse estilo de constatação do mental como acontecimento que Dennett aprecia em Wittgenstein e Ryle, é essencialmente por essa razão que Dennett os considera como "fenomenólogos". No entanto, a vida mental modesta e atentamente descrita deverá, segundo Dennett e ao contrário do que Wittgenstein e Ryle defenderam, ser posteriormente relacionada com o *nível sub-pessoal* e portanto com *modelos mecânicos da cognição*. É com o nível sub-pessoal que lidam os inquéritos empíricos e a atenção aos inquéritos empíricos é a diferença resultante da intenção naturalista em filosofia. Dennett pretende portanto seguir as intuições de Wittgenstein e Ryle quanto ao nível pessoal mas sem hostilidade em relação às investigações empíricas sobre a mente, antes começando precisamente por elas.

Para o auxiliar a contornar o défice de naturalismo que era, em ultima análise, aquilo que Dennett criticava em Ryle e em Wittgenstein, o jovem americano trouxera dos Estados Unidos a marcante influência da leitura de Quine. No seu auto-retrato teórico, Dennett coloca a questão desta maneira: «O que se obtem quando se cruza um Ryle com um Quine? Um Dennett, aparentementemente»[76]. Dennett partia portanto para a teoria da mente com uma atenção ao que as ciências têm para dizer acerca da cognição (além de uma mais idiossincrática inclinação tecnófila[77]) que embora se tenha difundido bastante com a ascenção das ciências cognitivas de então para cá[78],

[76] DENNETT 1998y: 365.

[77] «In fact, had I not been raised in a dyed-in-the-wool "arts and humanities" academic family, I probably would have become an engineer» (DENNETT, 1998: 356). Essa inclinação fez de Dennett o filósofo de eleição de muitos cientistas da comunidade da IA e mesmo o filósofo "residente" no projecto COG – o projecto de construcção de um robô humanóide – do laboratório de IA do MIT (cf. http://www.ai.mit.edu, projecto COG).

[78] Cf. DENNETT 1986 a: ix «There are now more than a few philosophers of mind who are vastly more knowledgeable about the brain then I was then (or am now). A fairly professional knowledge of the other cognitive sciences is now considered a virtual qualification for professional status in the discipline» (...) «we are almost all naturalists today».

não era muito comum entre os filósofos nos anos 60, sobretudo em Inglaterra. Além do naturalismo como posicionamento metodológico alternativo à terapia conceptual ou à geografia lógica, Dennett vai buscar a Quine a sua primeira conclusão definitiva acerca de significado. Essa conclusão é uma posição quanto à chamada tese de Brentano, a tese acerca da "in-existência" intencional como marca do mental. A expressão "tese de Brentano" generalizou-se sobretudo a partir de uma passagem de *Word and Object* de Quine à frente analisada. Aquilo que Brentano afirma na *Psychologie vom Empyrischen Standpunkt* (1874) é que «Todos os fenómenos mentais são caracterizados por aquilo a que os Escolásticos na Idade Média chamavam a inexistência (*Inexistenz*) Intencional (e também mental) de um objecto *(Gegenstand)* e aquilo a que chamariamos embora em termos não totalmente desprovidos de ambiguidade, a referência a um conteúdo, o direccionar-se a um objecto»[79].

Em C&C Dennett toma a ideia de Brentano como uma ideia acerca do estatuto do significado e considera que a interpretação quineana dessa ideia é a correcta. A referida interpretação aparece numa passagem de *Word and Object* (1960), mais exactamente o §45 do Capítulo VI (Flight from Intension[80]), intitulado Double Standard. Nessa passagem Quine acerta as suas contas com a tese de Brentano, afirmando que se «pode aceitar a tese de Brentano como mostrando a indispensabilidade dos idiomas intencionais e a importância de uma ciência autónoma da intenção, ou como mostrando a ausência de bases dos idiomas intencionais e o vazio de uma ciência da intenção. A minha atitude, ao contrário da de Brentano, é a segunda. Aceitar o uso intencional pelo seu valor facial é, como vimos, postular relações de tradução como sendo de alguma maneira válidas, embora indeterminadas em princípio, relativamente à totalidade das disposições de fala. Tal postulado promete muito pouco *insight* científico se não tem melhor fundamento do que o facto de as supostas relações de tradução serem pressupostas pelo vernáculo da semântica e da intenção[81]». Este parágrafo de Quine, citado em C&C, é apontado em *The Intentional Stance*[82] como o ponto originário dos debates da filosofia da mente nas últimas décadas.

[79] BRENTANO [1874]: 8.

[80] O Capítulo VI aparece após a apresentação da teoria da linguagem e da verdade (I), da tradução e da significação (III), da referência e da regimentação.

[81] QUINE 1960: 221.

[82] DENNETT 1987: 340.

Não é exagero afirmar que, para Dennett, toda a filosofia da mente posterior a 1960 consiste de certo modo em notas de pé de página ao parágrafo de Quine.

As afirmações de Quine na passagem citada relacionam-se com uma conhecida posição sua: Quine não aceita que os "significados" sejam entidades. Esta decisão ontológica tem raízes profundas e repercussões em toda a filosofia quineana. Uma das razões apresentadas por Quine é a economia ontológica, o princípio segundo qual não há necessidade de admitir mais entidades do que as essenciais para o fim visado (pode-se considerar que o fim visado por Quine em filosofia é a explicação da possibilidade de todo o conhecimento sério possuído pelo indivíduo e pela espécie, desde a matemática, a física e a lógica até ao conhecimento envolvido no uso de línguas naturais[83]). Outra razão é o facto de a sinonímia, que seria um critério de identificação de entidades intensionais, ter sido posta em causa por Quine[84], ficando assim as supostas entidades intensionais "sem identidade". A ideia de sinonímia envolveria segundo Quine a ficção de um conteúdo que permanece fixado enquanto palavras e sintaxe variam. Sendo a fixação de conteúdo uma ficção, não existem critérios para a identificação de significados. Ora, segundo Quine, sem identidade não poderá haver entidades. Não existem portantos entidades que seriam "significados". Mas a posição de Quine deve ser compreendida sobretudo no âmbito da indeterminação da tradução. Esta é a principal razão pela qual Quine toma a decisão ontológica de não aceitar significados como entidades. A indeterminação da tradução é introduzida e analisada em *Word and Object* nos capítulos anteriores ao capítulo onde se inclui o citado §45, nomeadamente no Capítulo 2 (*Meaning and Translation*). No entanto, os temas envolvidos na questão vêm de longe no pensamento de Quine, e relacionam-se com a defesa da separação dos destinos, defendida desde por exemplo *From a Logical Point of View* (1953), de dois tipos de problemas semânticos: os problemas da referência (envolvendo a designação, a denotação e a verdade – problemas de Semântica I) e os problemas do significado (envolvendo sinonímia, analiticidade e tradução – problemas de Semântica II). Para Quine, se a Semântica I tem um estatuto sólido (i.e. existe uma boa teoria que permite lidar com os problemas, nomeadamente uma teoria da verdade) o mesmo não acontece com a Semântica II. O problema começa desde logo com a situação em que se pretenderia responder à questão "como é que se sabe que duas expressões têm o mesmo sentido?" (esta seria uma questão básica da semântica II). A partir daqui, todos os caminhos conduzem à indeterminação.

A maneira de abordar a questão da mesmidade do significado e da indeterminação da tradução em *Word and Object* – a obra maior de Quine, que Dennett teria em consideração no período em escreve – é a análise de frases ocasionais de observação. Estas são traduzíveis por terem o mesmo estímulo-sentido (*stimulus-meaning*)[85]. No quadro da teoria apresentada em *Word and Object*, estímulo-sentido de uma frase é o conjunto de estimulações que incitam ao assentimento ou ao dissentimento a essa frase. É aqui que se insere o célebre exemplo quineano da elocução "Gavagai", pronunciada pelo nativo, cuja tradução "radical" é tentada pelo linguista. O objectivo de Quine é chegar a mostrar que até mesmo frases de observação como "gavagai" – ligadas à ostensão que está, segundo Quine, na base da possibilidade de linguagem – apenas podem ser traduzidas holofrasticamente, i.e. como todos, uma vez que são condicionadas directamente pela situação estimulatória relevante. Logo, mesmo na tradução de frases deste tipo, e ao contrário do que inicialmente se pensaria, poderá haver ambiguidade. É o que se analisa em torno de "gavagai" e "rabbit" no §12 (Synonymy of terms) de *Word and Object*. As frases de observação devem ser consideradas como unidades fónicas, sem partes significativas salientes, condicionadas por situações estimulatórias acessíveis a todos os membros da comunidade linguística, e portanto por situações públicas. Mas a ideia segundo a qual poderia existir sinonímia baseada em estímulos (*stimulus-synonymy*) de frases ocasionais de observação (elas seriam mesmo o único ponto da linguagem em que isso pareceria possível) revela não ser plausível. Mesmo ao nível das fra-

[83] Em *From Stimulus to Science* (QUINE 1995: 16), Quine define a epistemologia naturalizada como a reconstrucção racional da aquisição efectiva por parte de um indivíduo ou da espécie de uma teoria séria do mundo exterior. Na formulação de *Pursuit of Truth* (QUINE 1990) o fim visado é distinguir que asserções são verdadeiras. Para atingir tal fim, Quine sempre foi favorável a «treating cognition from within our own evolving theory of a cognized world, not fancying that firmer ground exists somehow outside all that» (QUINE 1960: 235).

[84] Cf.QUINE 1953 a. Uma definição operatória de sentidos/intensões que permite a leitura da formulação de Brentano constrataria os sentidos ou intensões com as referências ou extensões. Sentidos são os traços característicos que todos (e apenas os) membros da referência/extensão de um termo possuem/partilham e que seriam, na terminologia brentaniana, os conteúdos referidos nos actos mentais ou os objectos para que estes se direccionam.

[85] Adopta-se a tradução de *stimulus-meaning* usada por P. Melo em MELO 1994.

ses de observação a tradução radical não é feita por estrita identidade de estimulo-sentidos. Segundo Quine, a segmentação ou tentativa de segmentação destas frases para a sua tradução envolve já "hipóteses analíticas", que não são para Quine verdadeiras hipóteses, i.e. hipóteses no sentido científico, hipóteses testáveis, já que não existe matéria de facto que as confirme ou refute. As hipóteses analíticas constituirão manuais de tradução e o ponto da indeterminação da tradução é que os manuais de tradução podem ser ao mesmo tempo incompatíveis entre si e cada um deles adequado a toda a evidência relevante: «Não pode haver dúvida de que sistemas rivais de hipóteses analíticas podem adequar-se perfeitamente à totalidade do comportamento verbal assim como à totalidade das disposições ao comportamento verbal e ainda assim especificar traduções mutuamente incompatíveis de inúmeras frases não susceptíveis de controlo independente»[86].

Se esta é a situação mesmo das frases de observação, para além destas tudo o resto na tradução de linguagens – nomeadamente as "frases permanentes"[87] – é, para Quine, indeterminado. A situação de tradução radical é exemplar do uso da linguagem em geral e particularmente do uso da linguagem com propósitos epistemológicos: « No mesmo grau em que a tradução radical de frases é subdeterminada pela totalidade das disposições para o comportamento verbal, as nossas próprias teorias e crenças são em geral subdeterminadas pela totalidade da evidência sensorial possível para todo o sempre»[88]. A posição eliminativista acerca de significados[89] está, como é sabido, ligada ao naturalismo: o pressuposto da análise quiniana da linguagem é que esta deve ser tratada como um fenómeno natural (Quine afasta-se aqui de Wittgenstein e Ryle). Transpondo as ideias de Quine acerca de significados para a teoria da mente, não é difícil

[86] QUINE 1960: 72.

[87] Frases permanentes (por oposição às frases ocasionais, como as frases de observação que têm vindo a ser mencionadas) são por exemplo as frases científicas.

[88] QUINE 1960: 78.

[89] Sublinhe-se desde já porque o facto é importante para o propósito do presente trabalho, que Quine defende uma posição eliminativista ou niilista acerca do significado e não uma posição eliminativista acerca da consciência. A existência desta última não é considerada problemática. Esta diferença é importante para a avaliação das teses de Dennett – que se afirma um seguidor de Quine – acerca dos *qualia*. A expressão "niilismo do significado" é utilizada para analisar a teoria quineana do significado em FODOR 1987 e em FODOR &LEPORE 1992. Fodor faz muita questão em distinguir o holismo do significado – a posição acerca de significados supostamente defendida por Quine mas que é segundo Fodor antes um niilismo – do holismo da confirmação, que é uma tese epistemológica.

constatar que as frases da linguagem mentalista são intensionais[90]. E Quine ensina a fugir das intensões, uma vez que as regras da lógica cuja base é verofuncional[91] não se lhes aplicam. A lógica necessária e suficiente para o discurso científico, a partir de cujo estado se responderá à questão ontológica (a questão acerca do que há), é a lógica extensional. Esta supõe a regimentação na notação lógica da quantificação. Antes da regimentação nenhuma noção séria e cientificamente pertinente de *referência* se aplica. Quine pretende portanto tratar as questões da referência e da existência em conjunto e de uma forma que deixa de fora todo o âmbito do mental. É claro que Quine admite que em termos de linguagem e de ligação do uso desta à reificação, se começa *in media res*, sendo as primeiras coisas faladas e referidas coisas físicas, médias e públicas. Essa é no entanto apenas uma limitação quanto ao ponto de partida e não quanto ao ponto de chegada. O ponto de chegada (as questões lógico-ontológicas acerca daquilo que há) será decidido na ciência.

A interpretação quiniana da tese de Brentano é um dos fundamentos epistemológicos e ontológicos da teoria da mente apresentada em C&C. É importante por isso não esquecer que um filósofo da mente pode encontrar em Quine a ideia segundo a qual a linguagem mentalista é pura e simplesmente excêntrica ao ponto de chegada da ontologia. De acordo com a interpretação quineana da tese de Brentano falar de intencionalidade, usar a "linguagem da mente", é apenas uma *maneira de falar*. Nada do que é dito em linguagem da mente é determinado ou refere. A tese de Brentano combina-se assim, como «Quine famosamente insistiu»[92], com a indeterminação da tradução[93]. A teoria da mente só pode ser, a partir daqui, uma teoria da interpretação. Recapitulando, para um quineano os significados não são entidades mas apenas maneiras (interpretativas) de descrever comportamentos. É com base na declaração de Quine segundo a qual a semântica (do sentido) não é "séria", que Dennett considera que a linguagem da mente é puramente instrumental. Toda a teoria se devotará então a mostrar que os cérebros não podem ser máquinas semânticas (*semantic engines*), por mais que pareçam muito sê-lo: nada pode ser movido ou causado por entidades inexistentes, os significados. É

[90] DENNETT 1969: 29.

[91] Com a expressão "lógica cuja base é verofuncional" pretende-se englobar a lógica interproposicional e o que na lógica intraproposicional lhe corresponde. A lógica verofuncional é a lógica proposicional e nela não fazem muito sentido estas questões.

[92] DENNETT 1998y: 357.

também por razões quineanas que Dennett considerará toda a pretendida precisão em caracterizações intencionais como um objectivo ilusório (é esta nomeadamente a raíz da sua rejeição do realismo intencional e da Hipótese da Linguagem do Pensamento de Jerry Fodor).

O naturalismo e o interpretativismo situam Denett numa linhagem filosófica ligada a Quine, possivelmente a mais influente figura na filosofia americana (e não só) dos últimos cinquenta anos. No entanto, para além de Quine, uma outra forte influência pessoal marca a filosofia americana neste período. É durante a década de 60 e no contexto do debate do problema mente-corpo, que Hilary Putnam escreve célebres textos nos quais propõe e defende uma nova teoria materialista da mente, o funcionalismo, que se opõe à teoria da identidade. Textos como *Minds and Machines* (1960), *Robots: Machines or Artificially Created Life?* (1964), *The Nature of Mental States* (1967) *The Mental Life of Some Machines* (1967), alguns dos quais vieram a tornar-se manifestos do funcionalismo, são exemplares da teoria da mente então desenvolvida por Putnam. Nesses textos, Putnam declara que o problema mente-corpo é um problema lógico e linguístico e não um problema empírico. O problema não se relaciona sequer directamente com o cáracter único da subjectividade humana, podendo surgir em relação a qualquer sistema capaz de responder a questões sobre a sua estrutura: «os aspectos lógicos do problema mente-corpo são aspectos de um problema que deve surgir para qualquer sistema computacional que satisfaça as condições (1) usa linguagem e constrói teorias (2) não conhece inicialmente a sua própria constituição física, excepto superficialmente (3) está equipado com orgãos sensoriais e é capaz de levar a cabo experiências e (4) vem a conhecer a sua constituição através de investigação empírica e construção de teorias»[94]. Aliás, e embora o propósito de Putnam seja compreender como devemos falar da vida mental de humanos e não como devemos falar de máquinas, várias apresentações da teoria envolvem sistemas não humanos que satisfazem essas condições, por exemplo uma Máquina de Turing em *Minds and Machines* e robôs (além de robôs feitos por robôs) em *Robots: Machines or Artificially*

[93] A indeterminação quineana não é uma tese pontual de filosofia da linguagem, isolável do resto da filosofia de Quine. Como o próprio Quine faz notar, há três indeterminações fundamentais no seu pensamento e que aliás relacionam a filosofia da ciência, a filosofia da linguagem e a filosofia da mente: a indeterminação da tradução, a inescrutabilidade da referência e a subdeterminação da teoria pelos dados. Cf. QUINE 1990 a, *Three Indeterminacies.*

[94] PUTNAM [1964]: 388.

Created Life?.O funcionalismo é concebido por Putnam como uma forma de esclarecimento das perplexidades lógicas e linguísticas do problema mente-corpo, e é uma forma sofisticada de materialismo que se opõe ao materialismo "mais simples" da teoria da identidade. Esta, segundo Dennett, fizera valer uma certa plausibilidade mais pela falta de atractivos da sua alternativa, o dualismo, do que por qualidades próprias. Para não pôr em perigo a unidade da ciência é preciso identificar ou reduzir as "ofensivas coisas mentais" a coisas físicas e esse é o propósito geral dos teóricos da identidade. Ora, para que muitos filósofos chegassem a ver, como Dennett via quando escreveu C&C, a teoria da identidade como um monismo «metafisicamente extravagante e implausível»[95] provavelmente ninguém foi tão importante como H. Putnam.

A enorme virtude de H. Putnam[96] foi ter feito reconhecer a importância da distinção entre estados físicos e estados funcionais ou lógicos de sistemas. Um sistema com estados lógicos ou funcionais relevantes tem propriedades que podem ser consideradas em certo sentido não físicas. O funcionalismo é inicialmente formulado nos termos das Máquinas de Turing[97], posteriormente em termos de autómatos probabilistas[98]. Trata-se contudo sempre de uma tese geral sobre sobre a "natureza dos estados mentais" ou dos "predicados psicológicos". A caracterização funcionalista dos estados mentais por Putnam clarifica o estatuto do acesso de um sistema cognitivo físico aos seus próprios estados internos, permitindo considerar a uma nova luz as virtualidades epistémicas dos relatos resultantes de auto-monitorização, nomeadamente as descrições introspectivas da vida mental feitas por humanos. Putnam recoloca o problema do acesso directo e privilegiado à mente numa nova arena, ao defender que a questão "Como é para um sistema estar num estado funcional?" difere de forma importante da questão "Como é para um sistema estar num estado físico?".

O artigo *Minds and Machines* tem como centro uma analogia entre um homem e uma Máquina de Turing e pretende dissolver o problema-mente corpo, mostrando que este é análogo ao problema da relação entre estados lógicos e estados estruturais. Mostrar que os dois problemas são *um* único problema é decisivo: ninguém leva-

[95] Cf. DENNETT 1969:5-6.

[96] De acordo com o próprio Putnam. Cf. PUTNAM 1994.

[97] Cf. PUTNAM 1960.

[98] Cf. PUTNAM 1967. Cf. também KIM 1996 para as precisões.

ria a sério o problema da identidade ou não-identidade entre estados estruturais e estados lógicos, uma vez que a resposta não tem qualquer importância. Após explicar o conceito de Máquina de Turing, Putnam faz notar o seguinte em relação a uma particular operação a ser executada por uma particular Máquina de Turing (no caso, o cálculo da 3000ª casa da expansão decimal de π). Faz todo o sentido perguntar como é que a máquina calcula a 3000ª casa da expansão decimal de π (a resposta terá que incluir a sequência de estados por que a máquina passa no processo de cálculo, a descrição das regras da tabela da máquina). No entanto, não faz nenhum sentido objectar que a máquina, para realizar a tarefa deve passar pelos estados A, B e C e que é um problema saber como é que ela pode assegurar que está nesses estados. Putnam mostra o absurdo da objecção notando que a descrição lógica (a tabela da máquina) descreve os estados apenas em termos das suas relações uns com os outros. Desde que existam estados distintos A,B,C e que eles se sucedam uns aos outros tal como é especificado na tabela da máquina a realização física é irrelevante. Se para estar no estado A a máquina tivesse que assegurar-se de que estava no estado A haveria um círculo vicioso. A máquina passa pelos estados sem ter que se assegurar de que está neles.

Putnam considera ainda o caso em que a máquina imprime a frase "Eu estou no estado A". Esse facto não supõe que ela tenha tido que assegurar-se de que estava no estado A? Putnam considera esta situação análoga à situação wittgensteiniana em que alguém diz: "Sinto dor". A pessoa, para dizer "sinto dor", não tem que reflectir para se assegurar. Ela não está a descrever coisa alguma. Quer o indivíduo wittgensteiniano quer a máquina putnamiana não passaram por qualquer processo de recolha de evidência e de realização de inferências para chegarem às respectivas "afirmações". O indivíduo "sabe" que sente dor e a máquina "sabe" que está no estado A respectivamente *sentindo dor* e *estando no estado A*. O relato verbal sai directamente do estado que é relatado, sem qualquer recolha de dados, sem qualquer cálculo ou raciocínio.

Putnam explora em seguida a analogia entre estados mentais e estados lógicos da máquina e entre estados físicos de humanos e estados estruturais da máquina. Uma Máquina de Turing abstracta, uma vez realizada, passa a "ter" estados físicos e problemas com eles. A máquina tem algum acesso a esses estados físicos ou estruturais: ela é capaz de alguma monitorização que lhe permite por

exemplo detectar um mau funcionamento. Nisso ela é semelhante a um humano, que de uma forma mais ou menos confiável pode detectar um mau funcionamento do seu corpo. A máquina pode até imprimir "o tubo de vácuo 312 falhou" e é razoável perguntar como é que ela se assegurou disso. Se o imprime erradamente, pode ter havido um mau cálculo ou um sinal dos mecanismos de auto-inspecção interpretado erradamente. Mas nenhum desses casos se põe quando ela imprime "Estou no estado A". Em suma, Putnam defende que pode haver uma descrição de humanos análoga não ao ponto de vista do engenheiro que descreve uma máquina física mas ao ponto de vista do matemático que especifica a tabela da Máquina de Turing. Esta descrição é feita em termos de estados lógicos e não de estados físicos e, para Putnam, é esse o estatuto dos estados mentais, cuja descrição pode ser efectuada sem referência à realização física.

A analogia entre um homem e uma Máquina de Turing resulta assim na aproximação entre estados mentais e estados funcionais e é a esta luz que Putnam vai considerar o estatuto das identificações teóricas em teoria da mente. Este era o objecto central de discussão dos teóricos da identidade, na medida em que as identificações teóricas permitiriam compreender a natureza e a possibilidade de leis psicofísicas, de previsão e de explicação causal. As *identificações teóricas* em causa são alcançadas através de *investigação empírica*, e portanto embora possam ser aparentemente definições (por exemplo "A dor é estimulação das fibras-c") elas não são *analíticas*. Putnam afirmava nestes textos – como é sabido Putnam veio a renegar o seu passado funcionalista – que se daria no futuro uma identificação teórica entre estados psicológicos humanos e os estados cerebrais correspondentes (análoga às identificações teóricas entre luz e radiação electromagnética ou entre água e H_2O, nos exemplos de *Minds and Machines*). No entanto os estados psicológicos em causa nas identificações deveriam ser concebidos não como análogos aos estados físicos, estruturais, definidos ao nível do *hardware* (como pretendiam os teóricos da identidade) e sim como análogos aos estados lógicos de máquinas, definidos ao nível da programação. Esta tese, segundo a qual os estados psicológicos viriam a ser empiricamente idênticos a estados funcionais, é considerada por Putnam como sendo ela própria uma hipótese empírica, menos vaga do que a hipótese da identidade de estados mentais com estados físico-químicos de sistemas e mais susceptível de investigação matemática e empírica. Para Putnam, na medida em que a questão

da natureza dos estados psicológicos era uma questão factual, a investigação consistiria em produzir modelos mecânicos. Para o Putnam "cientista" de então, isso e apenas isso era a psicologia.

No centro da ideia funcionalista está assim o facto de sistemas "psicologicamente isomorfos" não terem necessariamente de estar no mesmo estado físico para estar no mesmo estado lógico, e portanto não terem que estar no mesmo estado físico para serem descritos por um mesmo predicado psicológico. A teoria da identidade revelava-se a esta nova luz uma concepção paroquial (demasiado humana) do mental. Aos olhos de um jovem filósofo como Dennett, Putnam argumentava persuasivamente que não há razão para afirmar que todos os seres, humanos ou outros, que podem pensar o mesmo pensamento devam necessariamente instanciar uma mesma descrição física (por exemplo, não há necessidade alguma de considerar que, quando dois homens pensam por exemplo em "Espanha", eles partilham algum estado fisicamente descritível específico). Mostrar que é assim e porquê foi a grande contribuição de Putnam para a filosofia da mente nos anos 60.

1.1.1 *Breve referência a E. Anscombe, C.Taylor e a "alguns visionários da ciência e da engenharia"*[99].

Até aqui foi visto o modo como, em termos epistemológicos e ontológicos, Dennett é guiado na teoria da mente pelas obras de G. Ryle, W. V. Quine, L. Wittgenstein e H. Putnam. Ryle mostra que a alternativa entre monismo materialista e monismo idealista não é uma alternativa a não ser que o problema da mente seja mal pensado. Wittgenstein e Ryle em conjunto constituem bons exemplos metodológicos do que deve ser a descrição da mente ao nível pessoal. A Quine, Dennett vai buscar por um lado o naturalismo, que não se encontra em Wittgenstein nem em Ryle, e por outro lado uma tese sobre o estatuto do significado, que virá a justificar a sua posição instrumentalista, interpretativista e holista. Em Putnam Dennett encontra um tipo sofisticado de materialismo, o funcionalismo, que lhe permite fundamentar a oposição à teoria da identidade psicofísica. No entanto, no que respeita aos problemas da mente como problemas de acção as grandes influências são Charles

[99] DENNETT 1998y.

Taylor, com a sua crítica filosófica ao behaviorismo e Elizabeth Anscombe, com a sua análise da intenção.

A importância da crítica filosófica ao behaviorismo em psicologia reside no facto de ela apoiar a convicção de Dennett de que os fenómenos psicológicos não podem deixar de ser caracterizados *intencionalmente sem que a psicologia se veja impossibilitada de existir*. O verdadeiro problema (quer para C. Taylor, quer para Dennett) consiste em esclarecer o estatuto dessa caracterização. O behaviorista caricaturado por C. Taylor em *The Explanation of Behavior* considera que os idiomas intencionais e as explicações pela finalidade devem ser abandonados pois não se enquadram no esquema explicativo científico. Esta decisão conduz à procura de uma pura linguagem de dados. Ora, numa tal situação auto-imposta, o behaviorista nunca poderia sequer falar, por exemplo, do que é aprendido na aprendizagem, pois «o que é aprendido, mesmo no condicionamento rigidamente constrangido não são movimentos mas acções»[100]. Taylor fala de "acções" pois quando se trata de processos classificados pelo seu produto ou resultado (por exemplo "descobrir onde está a comida") este pode tipicamente ser alcançado de muitas maneiras (através de "diferentes movimentos" cuja "comunidade" reside apenas, precisamente, na caracterização intencional). A conclusão aplica-se mesmo aos contextos experimentais de condicionamento, embora nestes exista aparentemente um afunilamento dos movimentos para o que parece ser uma solução única.

É aliás a C.Taylor que Dennett vai buscar a oposição, que percorre C&C, entre *teorias periferistas do comportamento* (especificamente o behaviorismo) e *teorias centralistas do comportamento* (que explicam o comportamento através de finalidades). O termo "centralismo" que Dennett utiliza para classificar a sua própria posição em C&C é o termo utilizado por Taylor para uma teoria do comportamento que assume a indispensabilidade das descrições intencionais e finalistas desprezadas pelos behavioristas. Apoiado na análise das tentativas infrutíferas do behaviorismo psicológico de passar sem "desejos", "acções" e "crenças" na explicação do comportamento de organismos, Taylor defende a ineliminabilidade das explicações teleológicas (*explanations by purpose*) do comportamento humano e animal. O valor do livro de Taylor reside aliás no facto de ele analisar de forma

[100] TAYLOR 1964: 206.

muito específica o insucesso das tentativas de utilização de uma pura linguagem de dados pelos behavioristas. Em *The Explanation of Behavior* Taylor mostra que tais tentativas redundam em «futilidade (...) deselegantes adereços *ad hoc* e cláusulas de salvaguarda (*inelegant ad hoc props and provisos*), que tornam as teorias estímulo-resposta espantosamente complexas»[101]. Taylor conclui que «A Teoria E-R é rica em hipóteses que são já petições de princípio, soluções meramente verbais que deixam o problema intocado – pistas condicionais, estímulos relativos, integração sensorial, *drives* adquiridos de todos os tipos – que (...) são um sinal confiável da má saúde de uma teoria»[102].

A *Intention*, de Elizabeth Anscombe, Dennett vai por seu lado buscar uma caracterização sofisticada da intenção e da voluntariedade na acção humana, admirando nomeadamente a ideia segundo a qual a descrição intencional de comportamentos é uma espécie de abstracção e não um relato de eventos actuais.

Além destas influências determinantes de filósofos no pensamento de Dennett, é necessário considerar outras influências, alheias ao campo disciplinar da filosofia. Segundo Dennett «nesse tempo – em meados dos anos 60 – mais ninguém na filosofia estava a tentar construir essa estrutura [*Dennett refere-se a uma estrutura física que pudesse ser pensada como cumprindo o trabalho da mente*] por isso tratava-se de uma tarefa bastante solitária, e a maior parte da iluminação e do encorajamento que eu podia encontrar vinham do trabalho de alguns visionários na ciência e na engenharia: Warren McCulloch, Donald McKay, Donald Hebb, Ross Ashby, Allen Newell, Herbert Simon and J. Z. Young são alguns dos nomes que me vêm à mente»[103]. Noções como

[101] Cf DENNETT 1969: 34, remetendo para TAYLOR 1964: 272. É interessante notar que as referências filosóficas de C.Taylor incluem nomeadamente M. Merleau Ponty (*La Structure du Comportement* e *La Phénoménologie de la Perception*), o qual aliás nos seus últimos textos (escritos em 1956-1960 e reunidos em *La Nature*) abordava problemas semelhantes aos de C&C, citando aliás em alguns casos os mesmos exemplos (do contexto da etologia, da cibernética, do neo-darwinismo).

[102] TAYLOR 1964: 270.

[103] Cf DENNETT 1998: 356. W. McCulloch e W. Pitts publicaram em 1943 o famoso artigo *A Logical Calculus of the Ideas Imanent in the Nervous Activity* (cf. McCULLOCH & PITTS 1943 e comentário em GARDNER 1985:18). D. McKay utilizava modelos de teoria da informação em psicologia. D. Hebb, o neuropsicólogo canadiano cujo nome está ligado à famosa lei de Hebb, publicou em 1949 *The Organization of Behavior*, onde defendia que os padrões comportamentais se constituem ao longo do tempo através da conexão de conjuntos de células (*cell assemblies*) Cf. comentário em GARDNER 1985:271. W. Ross Ashby, um médico e matemático inglês interessado na elaboração de uma teoria mecanicista da mente, procurou construir uma máquina capaz de comportamento adaptativo, utilizando métodos lógicos axiomáticos, sem falar de consciência ou de propósitos, tentando

armazenamento de informação, transmissão de mensagens e execução de comandos tinham-se tornado incontornáveis em todo um conjunto de áreas disciplinares que iam desde a teoria da informação e a Inteligência Artificial até à biologia. Saber o que se está a fazer quando se fala de significados e finalidades era, assim, já um problema teórico para muitos dos "visionários" das ciências cognitivas nos anos 60. Era já também omnipresente e popular a crítica segundo a qual as máquinas eventualmente criadas como resultado de explorações científicas da cognição não poderiam por princípio ter inteligência genuína mas apenas aquela com que os seus criadores as dotassem. Esta ideia, que virá a ser o núcleo da experiência mental do Quarto Chinês de John Searle, é evidentemente decisiva para a abordagem filosófica do problema da intencionalidade. Em oposição a esse argumento popular e num movimento teórico que será constante ao longo da sua obra, D. Dennett vê a questão da seguinte maneira: afimar que máquinas só têm intencionalidade em virtude dos seus criadores é uma aplicação do princípio geral segundo o qual uma entidade inteligente só pode ser inteligente em virtude do seu criador. Então, o mesmo princípio deverá aplicar-se aos próprios humanos enquanto entidades inteligentes: eles só podem ser, como são, entidades inteligentes em virtude do seu criador. Para Dennett, no caso humano, esse criador é a evolução por seleccção natural. Ora, na selecção natural não há intenção inteligente nem poder de previsão. Portanto, não é verdade que entidades inteligentes só possam ser criação de entidades inteligentes: nós, os humanos, somos a prova viva desse facto, ao sermos resultado da selecção natural, que não é uma entidade inteligente. Em suma, Dennett recusa-se já nesta altura a admitir argumentos que concluam pelo carácter não genuíno da cognição artificial. O que lhe parece, pelo contrário, é que a cognição artificial poderá ser esclarecedora acerca da cognição humana. Dennett nunca teve aliás problemas em se declarar um defensor da IA forte[104].

É a partir desta conjugação de influências em epistemologia, ontologia, teoria da acção e do comportamento e ciência empírica

imitar o cérebro vivo. Cf GARDNER 1985: 19 e DUPUY 1994. A. Newell e H. Simon (economista, teórico da organização) são os fundadores da IA e formularam a hipótese dos Sistemas Simbólicos Físicos. Cf. GARDNER 1985:145 e MACHUCO ROSA 2002. J.Z.Young, neuroanatomista, publicou *A Model of the Brain* em 1965.

[104] DENNETT 1996: 62. Na definição de J. Searle em SEARLE 1992, a *IA forte* é a ideia segundo a qual *dotar um sistema do programa adequado é condição suficiente para ele ter uma mente exactamente no mesmo sentido em que os humanos têm mentes.*

da cognição que Dennett define aquela que tem sido de então para cá a sua posição básica em teoria da mente, à qual chama na altura "centralista", por oposição à teoria periferista por excelência, o behaviorismo. Ao centralismo Dennett chama hoje Estratégia Intencional (*intentional stance*) e à sua teoria da mente chamar-se-á em geral, daqui em diante Teoria dos Sistemas Intencionais (TSI). Note-se que a Estratégia Intencional, polémica e muito atacada sobretudo devido aos pressupostos instrumentalistas e implicações antirealistas, é avançada por Dennett como um passo além do veredicto ryleano de erro categorial. Dennett pensa que é preciso escolher, i.e. que é preciso dar prioridade ontológica, quando se trata de falar da mente, ou à linguagem das ciências naturais ou à linguagem da mente e é precisamente isso que faz. A escolha de Dennett sempre foi clara: só a linguagem das ciências naturais *refere*. Entre a linguagem das ciências naturais e a linguagem mentalista situa-se uma "barreira da fusão", o que significa que as descrições de sistemas em linguagem mentalista devem ser consideradas de forma holista, como descrições aproximativas. Todos os problemas relativos ao carácter provisório e ao estatuto interpretativo das atribuições intencionais e teleológicas decorrerão desta opção.

1.2 *O problema do conteúdo em* Content and Consciousness*: início de uma teoria teleológica do significado.*

Logo no início de C&C, Dennett declara que a teoria da identidade é implausível e que apenas a ausência de alternativas conduz a aceitá-la. T. Nagel fará notar[105] que Dennett assume, com esta rejeição sumária, que todos os teóricos da identidade defendem identidades gerais entre tipos mentais e tipos neuronais, o que não corresponde à verdade. Nagel admite no entanto que Dennett está certo quando vê uma motivação negativa para a teoria da identidade. Para Dennett, o problema com a teoria da identidade é supor que identificações teóricas entre o mental e o físico são sequer possíveis. Ora, identificações são possíveis apenas quando existem entidades distintas para identificar, e de acordo com Dennett, não existem entidades mentais. Os fenómenos mentais não podem ser (ou não ser) identificados com eventos

[105] NAGEL 1995a: 82.

neuronais, já que a linguagem mental na qual esses fenómenos são descritos não é referencial.

A única forma de evitar supor que termos mentalistas nomeiam entidades é "fundi-los" nos seus contextos. O passo seguinte consiste em constatar que é incoerente pensar que todas as coisas ditas acerca de um sistema físico em linguagem mentalista são falsas: devem existir verdades em linguagem da mente, que é preciso relacionar com a linguagem por meio da qual as ciências físicas tratarão o mesmo sistema. Na recensão que faz do livro de Dennett, Nagel nota que «como este novo método de análise é a espinha dorsal do livro, uma maior discussão do seu carácter lógico teria sido apropriada»[106]. Qual é afinal o teor do método de análise proposto? Para Dennett o problema ontológico da mente é um problema de relacionamento de linguagens, nomeadamente da linguagem da mente, cujos termos não referem, com a linguagem das ciências físicas. Embora os termos da linguagem da mente não refiram, a linguagem da mente cumpre alguma função: *ela adscreve conteúdo a sistemas físicos globalmente considerados*. Ora se a linguagem da mente na qual se adscreve conteúdo não é referencial, põe-se a questão de saber de onde provém o conteúdo com o qual ela lida. Segundo Dennett, o conteúdo resulta da interpretação de sistemas físicos como Sistemas Intencionais. Em C&C, Dennett não utiliza ainda a noção de Sistema Intencional (SI) no sentido especifico que lhe atribuirá posteriormente, a partir de *Intentional Systems* (1971), o artigo no qual especificará a Estratégia Intencional (*intentional stance)* através da relação com a Estratégia Física e a Estratégia do Design, aqui aglomeradas sob o título de linguagem das ciências físicas. No entanto, a noção de conteúdo é já tratada como relativa a sistemas intencionais abordados globalmente, a partir de fora.

1.2.1 *A partir de fora e de cima. Da referência segundo Quine à distinção dennetiana entre "referencial" e "não-referencial". Existência e Identidade. O holismo e a fusão experimental das frases mentalistas nos seus contextos.*

Em C&C, um Sistema Intencional é definido como um sistema capaz de discriminar traços complexos do seu ambiente e de reagir a esses traços, algo que nenhum sistema pode fazer sem interpretação

[106] NAGEL 1995a: 83.

da estimulação periférica. O sistema tem portanto não apenas que «produzir em si estados que co-ocorram normalmente com condições generalizadas do objecto no campo perceptual do sistema»[107] como também que ligar esses estados a estruturas eferentes. Um SI é um sistema físico teleológico, envolvido em afazeres no mundo: o teor intencional da descrição que dele é feita estará ligado ao carácter intencional (aqui entendido como "com propósito") dos sistemas, e este será ligado à evolução por selecção natural. Estabelece-se assim a continuidade entre a noção biológica de "função" e a noção de "conteúdo".

Como foi dito, Dennett coloca o problema do mental de uma forma peculiar, identificando-o como uma questão acerca do estatuto ontológico da linguagem da mente. Na origem deste tratamento está o pensamento de Quine acerca de ontologia e de linguagem, nomeadamente as teses acerca do sentido e da referência. No 1º Capítulo de C&C (*The Ontological Problem of Mind*), o exemplo central é, neste ponto, o exemplo das vozes. Dennett pretende mostrar as semelhanças entre a existência de mentes e de vozes. Vozes não são entidades identificáveis e no entanto a sua existência não é problemática. O exemplo das vozes é introduzido por analogia com os exemplos de Quine, no 6º Capítulo de *Word and Object*, exemplos como "sake", "miles" e "Fahrenheit degrees". Neste último capítulo de *Word and Object*, Quine procurava responder não apenas à questão ontológica geral "O que é que há?" como também à meta-questão "Em que é que consiste o compromisso de uma teoria com objectos?". Entre os objectos abstractos e concretos com cuja existência as teorias podem estar comprometidas – como *sense data*, objectos físicos, números, classes, proposições – Quine considera pseudo-objectos como "sakes" e "miles", para exemplificar o ponto segundo o qual o facto de um termo ser um termo na linguagem natural não é prova alguma de que ele tenha "carácter de termo" (nas palavras de Quine, «uma ocorrência superficial sob a forma de termo não é prova do carácter de termo» (*a superficial termlike occurence is no proof of termhood*)[108]).

O projecto explícito do 6º Capítulo de *Word and Object* é "restaurar a lei e a ordem na ontologia", nomeadamente por oposição à linguagem ambígua de filósofos que distinguem um sentido de

[107] DENNETT 1969: 73.
[108] QUINE 1960: 236.

existir para objectos concretos e outro sentido de existir para objectos abstractos (se Quine não pensava em Ryle poderia ter pensado). Para Quine, o âmbito da resposta à questão ontológica deve ser limitado à referência e aos objectos que serão admitidos como valores de variáveis. Por isso «Parafrasear uma frase na notação canónica é antes de mais tornar o seu conteúdo ôntico explícito»[109]. É a esta teoria que Dennett vai buscar o modelo para o tratamento de "vozes", que por sua vez exemplificará o bom tratamento do problema ontológico da mente. Os exemplos acima referidos (*sakes, miles, Fahrenheit degrees*) são considerados por Quine como nomes defeituosos (*defective*), de modo que questões sobre a sua identidade são absurdas. Com as vozes passar-se-á algo de análogo.

A situação é a seguinte: nós ouvimos vozes, temos vozes boas ou más, vozes de tenor, de soprano, podemos ficar sem voz, etc. Impõe-se saber se uma voz é uma entidade, e no caso de o ser, que tipo de entidade é. Será uma parte física do corpo de uma pessoa? Mas nesse caso como seria possível que, uma vez gravada, ela fosse por exemplo apreciada por outros, após a morte da pessoa? Em alternativa, uma voz poderá ser identificada com movimentos vibratórios de partículas. Mas nesse caso como pode alguém dizer que perdeu ou esforçou a voz? Olhando para os exemplos de Quine, que mostram que palavras que são aparentemente termos podem não ter "carácter de termo" – nomeadamente as palavras "sake" e "behalf" – o que se verifica é que a significação desses termos não é separável dos contextos. No caso particular em causa, "sake" e "behalf" não são separáveis de expressões como "for the sake of" e "on someone's behalf". As expressões marcam o âmbito da significatividade. Termos assim são "nomes degenerados", "encerrados" dentro de "idiomas", e apenas os idiomas na sua totalidade significam. O facto de termos como "sake" serem morfologicamente nomes não nos deve, portanto, iludir. "Sake" não se comporta como "head" ou "wife": não podemos "escolher" um "sake" determinado como podemos escolher um exemplar de "head" ou "wife". Não nos comprometemos com a existência de objectos que são *sakes* pelo facto de usarmos de forma significativa o nome (degenerado) "sake". Quine faz ainda notar que termos para unidades de medida como "miles" e "Fahrenheit degrees" estão numa situação semelhante à situação de "sake" e "behalf": são nomes defeituosos e não faz sentido fazer cer-

[109] QUINE 1960: 242.

tas perguntas sobre a sua identidade. O que é importante notar é que esse facto não é suficiente para perturbar as descrições de distâncias em milhas e de temperaturas em graus Fahrenheit.

Dennett sugere em C&C que o funcionamento dos "nomes degenerados" de Quine é comparável com o funcionamento dos termos em linguagem da mente, como "pensamentos", "dores" e "desejos". Fixando-se no exemplo da "voz", Dennett constata que a "voz" resiste à identificação (com uma parte do corpo, com vibrações do ar, etc) e se o mundo não é ele próprio ambíguo, como os sentidos de "voz" aparentemente são, deve haver algo de vão na tentativa de identificar a porção do mundo físico que constitui uma "voz". É exactamente esse o caso segundo Dennett. É vão tentar apontar uma porção do mundo físico que constitui uma voz, mas isso não acontece pelo facto de a voz ter uma forma insatisfatória de existência (pois todas as coisas que podemos dizer acerca de vozes têm explicações físicas) ou por "voz" ser uma palavra insatisfatória (pois sabemos perfeitamente utilizar a palavra "voz" de forma significativa). O que o exemplo mostra é que é possível falar significativamente de "vozes" sem necessidade de as identificar com uma parte específica do mundo físico. Podemos (e devemos) ser fisicalistas acerca de vozes sem presumir identificações. Assim como afirmar que fazermos alguma coisa "*for someone's sake*" não nos obriga a dizer que existem "sakes", também não estamos comprometidos com a existência de vozes, pelo facto de reconhecermos vozes, disfrutarmos de vozes, etc. O interesse de Dennett não são evidentemente as vozes mas o mental, e é ao mental que deve ser aplicada esta mesma conclusão.

Enfrentar uma questão ontológica só tem interesse quando «conceder existência a alguma coisa nos permite perguntar (e esperar respostas para) coisas muito gerais acerca dela como, por exemplo, que tipo de coisa é?, existe no tempo? e especialmente, essa coisa é idêntica a x?»[110]. Se existem de facto duas coisas, não poderia ser evitada a questão da sua identidade ou não identidade, mas se se negar a existência de uma delas, já não seremos obrigados a identificá-la. É esse o caso da descrição de vozes e também de entidades mentais. *Apenas se entidades mentais existissem específica e distintamente como mentais é que a questão da identificação com eventos físicos se colocaria.* E se assim fosse, voltar-se-ia a todo o quadro cartesiano e pós-cartesiano, aos debates em torno de interaccionismo, paralelismo, identificação ou não identificação, em suma, a uma re-encenação da filosofia moderna. A filosofia ryleana aparecera como um primeiro modo de evitar este destino. No entanto para Dennett a teoria da mente não pode contentar-se com o veredicto de erro categorial quanto a certos usos da linguagem. Se não for dada prioridade ontológica a uma categoria nunca será evitado algum resquício de dualismo. A virtude do naturalismo quineano é tornar clara a obrigação de escolher o que é mais fundamental na realidade. É então, para estabelecer as consequências da prioridade ontológica, que Dennett introduz o contraste entre linguagem *referencial* (caso de nomes ou nominalizações que denotam ou nomeiam ou se referem a coisas existentes no sentido forte definido acima, objectos com os quais a teoria se compromete e que serão valores das variáveis na notação canónica) e *não referencial* (caso de "sake", "mile", "voz", e de todos os termos mentais). Esta distinção entre referencial e não refencial é reportada a *Word and Object* (embora o sentido de "referencial" não seja idêntico ao proposto por Quine). Dennett considera como não referenciais todas as expressões dependentes do contexto, que devem ser consideradas como fundidas nesse contexto. A ocorrência de tais expressões, mesmo que aparentemente nominais, não refere nem compromete o enunciador com a existência das entidades nomeadas. De facto, não é sequer possível analisar essas expressões para além do "fechamento holista" no contexto, sem incorrer na situação que Dennett exemplifica pela análise de "potable" em "po"-"table". Ora, é este o estatuto da linguagem da mente, é por isso que frases em linguagem da mente devem ser tratadas como todos. Posteriormente, tratar-se-á de saber se os "todos mentais" podem ser de alguma maneira correlacionados de modo explicativo com o domínio referencial das ciências físicas. Mas assim como a explicação de vozes pode não fazer referência a vozes, a teoria do mental poderá não fazer referência a entidades mentais.

[110] DENNETT 1969: 11.

1.2.2 *A partir de dentro e de baixo. Informação e Teleologia. Armazenamento inteligente de informação e comportamento. O conteúdo e o funcionamento de estruturas apropriadas. Evolução no cérebro.*

A linguagem mentalista envolve crenças e desejos. Estes interpenetram-se na descrição do mental e não podem ser caracterizados independentemente. Um comportamento só é interpretável

como indiciando uma determinada crença se fôr atribuído um desejo ou propósito à entidade que supostamente possui essa crença: o facto de a pessoa A se pôr debaixo de um telhado só constitui evidência da sua crença de que "está a chover" se ela tiver a intenção de não se molhar. É esta interdependência que justifica a abrupta caracterização, usual na filosofia da mente, do mental em termos de *beliefs and desires,* como se não houvesse outros estados mentais. A interpenetração de crenças e desejos (i.e. de significação e teleologia), é muito importante para Dennett e sugere, antes de mais, uma particular teoria da natureza da crença. Dada a interpenetração de significado e função, não seria muito surpreendente se a capacidade que a teoria da selecção natural tem de dar conta das funções de orgãos e comportamentos de sistemas vivos fosse acompanhada por uma capacidade de dar conta dos significados de estados desses sistemas. Dito isto, compreende-se que Dennett inicie em C&C o desenvolvimento da sua teoria do conteúdo com um capítulo sobre a evolução no cérebro[111].

O pressuposto geral daquela que vai ser uma teoria evolucionista do significado é que *o significado não é um traço a mais que se descubra num sistema mas sim uma interpretação.* Não sendo o significado um traço a mais, presume-se que deve ser possível uma descrição extensional do sistema físico que não lhe faça alusão, mesmo que esta seja difícil ou impossível de alcançar. Ora, o importante para Dennett é fazer notar que mesmo que tal descrição fosse feita, ela não incluiria a descrição do que o sistema está a fazer (estando incluído em "fazer" o "pensar", ter estados mentais). É aqui que se instala o centralismo proposto. A vantagem deste é antes de mais heurística. Uma descrição centralista permite descrever rapidamente o que o sistema está a fazer, sem necessidade de esperar pela descrição extensional. Evidentemente, o valor heurístico do centralismo é directamente proporcional à complexidade do sistema: quanto mais complexo fôr um sistema, maior vantagem e utilidade terão as descrições centralistas. Assumindo que as descrições centralistas intencionais são úteis para fazer sentido dos afazeres de sistemas físicos, o problema é saber o que é que as legitima.

[111] Poder-se-ia utilizar a expressão "darwinismo neuronal" (de G. Edelman, cf. EDELMAN 1987) para caracterizar a evolução no cérebro. DENNETT 1986: xi, Preface to Second Edition of DENNETT 1969 «Chapter III, on *Evolution in the Brain* stands up well, in spite of its technical naiveté. The very recent upsurge in enthusiasm among neuroscientists for theories of learning as intracerebral evolution is particularly gratifying. While Edelman at Rockfeller, Changeux in Paris, the New Connectionists in artificial intelligence and others are now developing evolutionary models at a level of empirical detail and sophistication I could not imagine in the 1960's I am pleased to see that their accounts appeal heavily to the concerns I outlined in this chapter».

A primeira coisa a constatar é que comportamento intencional não pode existir sem armazenamento (inteligente) de informação pela entidade que exibe esse comportamento. Para o armazenamento de informação poder ser considerado como *inteligente* os fins da entidade devem estar de alguma maneira dentro dela e a informação deve poder ser utilizada para prosseguir esses fins. Essa situção opõe-se ao caso em que uma entidade é meramente suporte de informação (ser suporte de informação corresponderia por exemplo ao caso em que se diz que numa montanha existe "informação" geológica). Por outro lado, a base para a capacidade de armazenamento inteligente de informação só pode ser o armazenamento não inteligente de informação. Esse armazenamento não inteligente de informação não é mais do que a "confiável plasticidade" do que quer que esteja entre input e output.O que faz a diferença entre o armazenamento não inteligente de informação e a informação inteligente é assim o facto de esta última ser informação para o sistema, que deve ter usos para ela. Ora, para haver usos para a informação, o sistema, a entidade, deve ter fins próprios, mesmo porque «O critério do armazenamento inteligente de informação (é precisamente) o carácter apropriado (...) do comportamento resultante»[112]. Ora, o carácter apropriado do comportamento não é uma caracteristica física intrínseca ao sistema, que possa ser encontrada nos seus inputs e outputs. Para o perceber é preciso olhar para "fora", i.e. para os afazeres do sistema no seu ambiente.

O passo seguinte é perguntar como pode surgir um sistema capaz de comportamento adaptado ao ambiente através do uso inteligente de informação, já que a capacidade de armazenamento inteligente de informação não surge "automaticamente" (por exemplo a partir de uma determinada *quantidade* de informação). Para Dennett o problema da especificação das características de um tal sistema capaz de um uso inteligente de informação é o problema da "evolução das estruturas apropriadas". Um cérebro, por exemplo, só é "útil" na medida em que controla comportamento apropriado ao ambiente do sistema guiado por esse cérebro. O problema assume então a forma de compreender a génese das estruturas no cérebro que controlam o

[112] DENNETT 1969: 6.

comportamento apropriado do sistema global. O sistema é uma entidade separada, delimitada no ambiente e a apropriação ao ambiente de uma entidade assim individuada não acontece casualmente num ambiente complexo. Para que exista adequação do comportamento do sistema, o cérebro (ou outra estrutura de controlo) deve ter a capacidade de discriminar o que, no ambiente, é significativo para o sistema. Ora, o problema é saber como é que a discriminação daquilo que é significativo pode ser cumprida por um dispositivo isolado do ambiente. Não é razoável supor que a discriminação do que é significativo seja feita por leitura ou decifração do significado dos eventos físico-químicos internos causados pelos estímulos ambientes (embora evidentemente essas metáfora sejam continuamente utilizadas): eventos físicos não têm significado intrínseco e não existe um "leitor interno". Logo, a capacidade de discriminar pela significação não pode ser simplesmente uma capacidade de analisar o input. O cérebro é em si "cego" às condições externas (dito de outra maneira, H. Putnam tem, de certa maneira, razão acerca de todos os seres com cérebro com a ideia de "cérebro numa redoma").

O exemplo que Dennett dá da "cegueira interior" do sistema físico que produz comportamento adequado é o facto de «os critérios pelos quais o grupo do MIT (H. Maturana, W. McCulloch, J. Lettvin e W.Pitts[113]) determina que certos sinais aferentes da retina da rã significam contraste ou bordos que se movem ou convexidade não pode(re)m ser usados pelo cérebro da rã para discriminar estes sinais, pois o cérebro da rã não pode observar a retina da rã, não pode dizer de onde vêm esses sinais»[114]. O significado de coisas no ambiente não pode ser discriminado por testes físicos. A única possibilidade de explicar a significação na relação sistema/ambiente é através da conexão entre os eventos aferentes e os eventos eferentes que controlam o comportamento apropriado. Um sistema capaz de gerar e preservar esse tipo de conexões será capaz de comportamento apropriado em relação ao seu ambiente. Para explorar esta via, Dennett introduz a noção de "Estrutura funcional". Uma estrutura funcional é aquela da qual se pode esperar que opere de uma certa maneira, tendo um modo de funcionamento normal e outro desviante. Então, «dado um cérebro com plasticidade inicial ou capacidade de produzir diferentes estruturas funcionais como resultado de input, a chave

[113] Cf. BECHTEL GRAHAM 1998: 30 (*Single Neuron Electrophysiology*).
[114] DENNETT 1969: 48.

para a utilidade no cérebro deve estar na capacidade de seleccionar por entre estas estruturas funcionais»[115], i.e. na capacidade de manter aquelas que são úteis e de eliminar as que são prejudiciais. O problema é evidentemente caracterizar o princípio de selecção das estruturas funcionais no cérebro. Dennett considera que esse princípio deve ser análogo à selecção natural. A selecção das estruturas não acontece em virtude da "aptidão física" destas, pois não é isso que as torna úteis ou prejudiciais: o processo de selecção tem que ser exteriormente fundamentado.

Considerando a história evolutiva dos sistemas nervosos, constata-se que todas as respostas desencadeadas num sistema nervoso e que controlam, por exemplo, comportamentos primitivos alternativos, são respostas cegas. No entanto, respostas cegas que controlam comportamentos que por acaso resultam apropriados serão mantidas, pela própria sobrevivência do sistema que as tem instaladas (*built-in*), enquanto que as respostas inapropriadas se extinguirão com os organismos nos quais existem: «As espécies que sobrevivem são as espécies que por acaso têm impulsos eferentes (...) ligados a impulsos aferentes (...) de maneiras que as ajudam a sobreviver»[116]. Uma vez estabelecidas geneticamente as conexões entre eventos aferentes e eventos eferentes que controlam um dado comportamento, poder-se-á afirmar dizer que os estímulos que despoletam os eventos aferentes "têm significado": eles têm significado em virtude dos efeitos que lhes estão "ligados".

O controlo de comportamentos apropriados começa por ser controlo de comportamentos rígidos e instintivos, e portanto controlo fixado geneticamente em sistemas. Todo o comportamento de que um sistema é capaz poderia (e pode, em muitos sistemas) ser assim rigidamente fixado. Nesses casos, poder-se-ia afirmar que o organismo "sabe" desde logo tudo o que tem que "saber", desde o nascimento até à morte, pelo facto de ser como é. Padrões rígidos de comportamento semelhantes existem em muitos organismos vivos e constituem todo o leque do seu comportamento. Em alternativa, pode existir plasticidade no sistema de controlo do organismo, nomeadamente no sistema nervoso. Essa situação alterará as condições do controlo do comportamento do sistema. Para dar um exemplo da utilidade da circunstância "nova" de plasticidade, basta

[115] DENNETT 1969: 48.
[116] DENNETT 1969: 49.

notar que tropismos, i.e. sequências comportamentais rígidas, fixadas, mostram ser incapazes de lidar com a transposição do organismo para longe do ambiente ao qual são apropriados.

Alguma instalação (*pre-wiring*) de ligações entre eventos aferentes e eventos eferentes existe sempre que exista um organismo que se "comporta", mesmo que exista plasticidade: não são possíveis organismos "tábua rasa". O problema interessante é imaginar como é que uma pré-instalação poderá vir a possibilitar, em sistemas capazes de aprendizagem, um comportamento apropriado em relação àquilo que não é geneticamente dotado de significação. Noutras palavras, é preciso conceber como poderá acontecer um *re-wiring* (uma re-instalação de controlos) a partir do *pre-wiring*. Uma má concepção partiria da ideia de um sistema de controlo (um cérebro) com dois lados naturalmente separados de forma absoluta, que seriam o lado aferente e o lado eferente, em alguns casos existindo estruturas ligadas de forma pré-estabelecida e noutros estruturas à espera de serem correctamente ligadas. Dennett sugere, em alternativa, que estruturas aferentes e eferentes estão desde logo interconectadas, em parte aleatoriamente. Existe portanto uma abundância inicial, da qual serão seleccionados os bons candidatos, por um processo de evolução selectiva. O problema é saber como é que as conexões apropriadas, de entre as muitas que há, são seleccionadas e portanto "sobrevivem", e não saber como é que são estabelecidas as conexões. Assim, aquilo que importa conceber é um processo de extinção do inapropriado e de propagação do apropriado. Dennett sugere que o factor decisivo é o fenómeno neurofisiológico do limiar (*threshold*) variável para o disparo dos neurónios: este limiar baixa com a maior frequência dos disparos do neurónio e sobe com a inactividade do neurónio. Segundo Dennett, para além dos detalhes neurofisiológicos, o que interessa em termos de modelo abstracto de uma entidade capaz de cognição, é apenas isso: «tudo o que precisamos é de uma multidão de interruptores (...) e de uma regra geral segundo a qual o disparar de um elemento interruptor aumenta a probabilidade de que esse disparo se repita»[117]. É esta condição que fornece o princípio da propagação das espécies (i.e. dos padrões neuronais que se estabelecem e propagam) no cérebro. Dennett considera ainda exemplar outra característica da neurofisiologia humana, necessária aliás em qualquer sistema confiável de

[117] DENNETT 1969: 54.

processamento de informação: o cérebro humano "utiliza" a redundância num contexto de "sinais" neuronais "ambíguos". Se os impulsos neuronais constituem sinais com "conteúdo ou significado" e se um particular neurónio dispara se e só se um padrão de estimulação determinado está presente no ambiente, esse output não é de todo ambíguo. Mas se um padrão ambiente A (mas também um padrão B e também um padrão C, etc) têm o mesmo resultado, existe ambiguidade. Acontece que, excepção feita aos níveis mais periféricos, essa parece ser a situação mais geral no sistema nervoso humano. E a ambiguidade é, sugere Dennett, essencial para o funcionamento confiável de um orgão como o cérebro. Se ela não existisse, se houvesse neurónios com funções exclusivas, não duplicadas por outros neurónios, a lesão ou morte de neurónios específicos seria o fim absoluto da função. Mas, se assim fosse, as lesões e degenerescências não apresentariam as características que apresentam. O que acontece é que o funcionamento globalmente considerado do cérebro é muito mais confiável do que o funcionamento das partes e que uma dada percentagem de "mau funcionamento" das partes é simplesmente normal. A fiabilidade da transmissão por elementos não fiáveis em isolamento é então conseguida através de uma duplicação de sinais. A redundância, num sistema de neurónios em "dedicação exclusiva" a uma função, resultaria em multiplicação ineficiente. No entanto, dada a ambiguidade, a redundância pode ser conseguida com menos elementos: «São portanto as mais ou menos simultâneas concatenações de vários outputs neuronais e não os outputs de neurónios isolados que são não ambíguas»[118]. De tudo isto, o essencial para a teoria do conteúdo é a ideia de composição dos "veículos de conteúdo" (as estruturas funcionais de conexões eventos aferentes/ eventos eferentes têm partes, e podem ser decompostas e reconstituídas) e a ideia de desambiguação global.

Os elementos referidos (limiares variáveis, sinais "compostos", abundância de conexões aferentes/eferentes) permitem conceber um processo de evolução no cérebro que explica a capacidade de discriminação pelo significado. Não se espera que as conexões abundantes iniciais sejam desde logo apropriadas, dado o carácter composto dos sinais. Existiria antes, como na evolução das espécies, algum conflito entre candidatos e características do ambiente (neste caso, as condições de partida são as conexões pré-instaladas). As

[118] DENNETT 1969: 56.

conexões geneticamente estabelecidas devem ter a capacidade de inibir a concorrência, i.e. a programação genética deve poder predominar sobre as propostas possíveis devido à plasticidade (ou poderia ser desfeito no indivíduo o trabalho da evolução na espécie, o qual tornou o organismo capaz de responder adequadamente ao ambiente de várias maneiras). Como acontece com as espécies, não é a morte que conta definitivamente como "inibição" mas a incapacidade de se reproduzir. Pela acção de um mecanismo como este, reproduz-se não apenas tudo o que for compatível com as conexões já fixadas como também o que não fôr inapropriado, mesmo traços que excedam o imediatamente funcional. O efeito gradual serão novas estruturas funcionais, capazes elas próprias de se sobreporem à concorrência. O processo resulta, segundo Dennett, numa repetida autopurificação da função, ganhando em eficácia à medida que mais estruturas não inapropriadas se estabelecem. Quanto mais inteligente estiver o sistema destinado a ser, mais *soft programming* deverá ser possível, i.e. mais programação não inicialmente instalada deverá poder ter lugar, a qual é naturalmente mais apropriada para lidar com novos estímulos. Biologicamente, essa possibilidade corresponde aos longos períodos de infância dos animais mais inteligentes. A velocidade da evolução no cérebro é incomparável com a velocidade da evolução das espécies: nela, as gerações podem durar milésimos de segundos, em vez de meses ou décadas. No entanto, o processo é idêntico. Assim, é relativamente pouco importante discutir se uma particular instalação, uma implantação material do uso inteligente de informação, é geneticamente programada ou resulta de *soft-programming* e aprendizagem (nomeadamente porque um comportamento geneticamente controlado e um comportamento dependente de um controlo que se estabelece dadas regularidades presentes no ambiente de aprendizagem da espécie têm o mesmo valor de sobrevivência). Em termos de cognição é aqui que se situa a verdade do behaviorismo: um processo de ensaio e erro é necessário para explicar a instalação destes "controlos de comportamento".

1.2.3 *A estrutura do comportamento. O comportamento dirigido a fins* (goal directed behavior) *e a adscrição de conteúdo. Princípios da teleossemântica. Mente e linguagem.*

As estruturas apropriadas produzidas pela evolução têm não apenas causas mas também *razões*. Convém no entanto fazer uma

distinção entre razões instaladas *no sistema* e razões *para o sistema*. As razões das estruturas apropriadas não têm que ser apercebidas para funcionarem como tal (este é o motivo dos *free floating rationales* ou razões não apercebidas pelas entidades que as têm "instaladas"). No entanto, razões-apercebidas não são por natureza distintas de meras razões no sistema.

Tendo como objectivo defender que o comportamento estímulo-reposta não é a princípio distinto do comportamento "com razões" ou dirigido a fins, Dennett começa por analisar estruturas aferentes-eferentes simples, que controlam comportamento elementar de discriminação, por estímulo-resposta. Para estabelecer a natureza do comportamento dirigido por fins, Dennett analisa o modo como a ideia de fim *(goal)* aparece em programação. O exemplo é o programa General Problem Solver[119] (GPS) de A. Newell e H. Simon. Este é descrito pelos seus autores como operando de forma dirigida por fins (por oposição a simplesmente seguir os passos programados, i.e. dados pelo programador, que teria, nesse caso, apenas ele próprio um fim para o programa): o fim (*goal*) está incorporado na especificação do estado final. Para o estado final ser atingido são recrutados meios, i.e. sub-rotinas. O problema é compreender em que sentido se pode dizer que o GPS tem fins e selecciona meios para os atingir. A alternativa a esta análise comparativa seria, obviamente, distinguir em absoluto e por princípio o genuíno comportamento dirigido por fins (por exemplo o humano, em que fins são intencionados por agentes conscientes) do comportamento que apenas termina em fins (o que seria o caso do GPS).

O comportamento dirigido por fins do GPS envolve um estado final especificado, que vai sendo procurado. Curiosamente, sublinha Dennett, Newell e Simon afirmam que o programa é dirigido por fins apesar de ele não ser capaz de mudar a ordem das sub-rotinas pelas quais procura nem de excluir sub-rotinas inaplicáveis. O programa não tem *insight* para reconhecer progressos na tarefa: simplesmente tenta todas as possibilidades. Poder-se-ia objectar que, exactamente por isso, não se trata de um *genuíno* comportamento dirigido por fins. No caso humano, por exemplo, um fim dirige-nos durante a tarefa, no caso do GPS o fim é a terminação da actividade. Dennett pensa no entanto que não é possível estabelecer semelhante distinção. Se fosse possível estabelecê-la, apenas

[119] Cf. GARDNER 1985: 148-151 e BODEN 1977: 354-357 e 359.

os humanos seriam dirigidos por fins. No entanto o comportamento animal parece ser também dirigido por fins. Aliás, pensamos que o comportamento animal é dirigido por fins precisamente porque os animais ensaiam padrões motores alternativos até conseguirem atingir um determinado estado final. Ser dirigido por fins não pode então excluir por princípio a repetição de tentativas, mesmo que estas sejam aleatórias ou vãs. Assim, segundo Dennett, mais do que a inteligência óbvia da execução (incluindo o reconhecimento do progresso e a exclusão automática das tentativas que seriam vãs), as marcas distintivas do comportamento dirigido por fins são as próprias tentativas e o reconhecimento do estado final. A capacidade de raciocinar não acompanha necessariamente o comportamento dirigido por fins. A posse cognitiva prévia do fim não poderia ser o critério do genuíno comportamento dirigido por fins, até porque não dispomos de critérios para assegurar essa posse: não seria certamente critério a declaração do fim, pois foi admitido que animais podem ser dirigidos por fins, e estes certamente não os declaram. O único critério seguro do comportamento dirigido por fins é a verificação das tentativas para atingir o fim e a verificação do reconhecimento do estado final, uma vez atingido. Ora, esse critério admite o comportamento do GPS como comportamento dirigido por fins. O programa seria então um bom modelo abstracto pelo menos do comportamento animal. Resta saber se, aceitando o GPS como um modelo simples e adequado do comportamento dirigido por fins, a evolução por selecção natural no cérebro pode produzir sistemas semelhantes. Ora de acordo com Dennett, o modelo da evolução atrás apresentado pode de facto dar conta do surgimento de tais sistemas dirgidos por fins.

Um problema imediato neste ponto é que os estados finais especificados de comportamentos animais complexos (por exemplo: "voltar a casa" ou "encontrar um parceiro sexual") não são caracterizáveis em termos de estados neuronais periféricos. A actividade do sistema de controlo deve portanto poder ser determinada por descrições bastante gerais das condições externas e o estado final do comportamento dirigido por fins deve ser actividade neuronal de nível muito mais alto do que o nível periférico. De forma muito geral, o propósito do comportamento de humanos e outros animais é divisível em hierarquias de fins, a partir das necessidades básicas (como sobrevivência, nutrição, procriação). O exemplo que Dennett dá de uma hierarquia de fins é o seguinte: «serra-se a tábua

para construir a porta para pôr na casa para a manter segura para proteger a nossa saúde para nos mantermos vivos»[120]. A ideia sugerida é que estes padrões hierarquizados poderiam ser gerados a partir de uma única "semente", um conjunto pré-instalado de controlos para padrões simples de comportamento dirigido a fins, posteriormente "talhado" e especificado pelos processos de evolução no cérebro acima descritos. Diferentes sub-rotinas e sub-fins seriam instalados, diversificando os controlos para o comportamento dirigido por fins com que o sistema enfrentaria o ambiente. Assim, descrevendo metaforicamente o resultado, «a iniciação aferente do comportamento produz uma pressão de actividade eferente que procura libertar-se por várias vias, as quais controlam várias tentativas de atingir o fim»[121].

Noções como informação, ambiguidade, redundância, controlo, e evolução de conexões aferentes/eferentes no cérebro são essenciais na parte da teoria do conteúdo que é feita a partir de dentro e de baixo. Ligando-a com a teoria do conteúdo feita a partir de fora e de cima, é fácil ver que a pertinência das descrições intencionais ou centralistas do sistema aumenta na razão directa da distância dos eventos neuronais descritos relativamente à periferia do sistema. Por exemplo, não se ganharia nada atribuindo o conteúdo "Contrai agora, músculo!" a impulsos motores, enquanto que, quando se trata do cérebro e das chamadas funções cognitivas superiores, descrições intencionais são frequentemente a única maneira de fazer sentido da actividade neuronal que controla o comportamento do sistema. Como se viu, a teoria do conteúdo envolve considerações acerca de funções e hipóteses acerca da evolução de estruturas funcionais no cérebro. É portanto da biologia e nomeadamente do evolucionismo que está suspensa a teoria do conteúdo de Dennett, à qual se chamará a partir de agora "teleossemântica". É o próprio Dennett que considera ter apresentado a teoria teleossemântica original, abrindo um caminho hoje explorado por filósofos como R. Millikan e F. Dretske. Os princípio gerais da teleossemântica proposta em C&C são os seguintes:

(1) A discriminação de estímulos aferentes de acordo com a sua significação (o seu "conteúdo") *é pura e simplesmente identificável com a produção de efeitos eferentes*. Não faz sentido pensar que antes da produ-

[120] DENNETT 1969: 70.
[121] DENNETT 1969: 71.

ção de efeitos eferentes em resposta a eventos aferentes o cérebro descriminou-como-alguma-coisa o que quer que seja.

(2) *Deve-se assumir princípios behavioristas em relação a cérebros. Isto significa que os cérebros* (ou outros sistemas de controlo de comportamento inteligente) *são máquinas de estímulo-resposta e não máquinas semânticas.* Poder-se-ia dizer que os cérebros, além de serem cegos, não pensam: apenas respondem e reagem. Para Dennett este não é um ponto epistemológico ou metodológico e sim um ponto lógico ou conceptual. Não faz sentido supor que a discriminação de um estímulo pela sua significação se faz unicamente do "lado aferente" do cérebro: aquilo que um evento neuronal "significa" para um organismo depende necessariamente do que o organismo faz com ele.

(3) *O conteúdo a adscrever* a estados, eventos ou estruturas neuronais de sistemas físicos (se é que algum conteúdo é adscrito) *depende de dois factores*: a. A *fonte normal* dos eventos/estruturas em termos de estimulação, b. Os *efeitos eferentes* apropriados que os eventos/estruturas têm. Evidentemente, estes factores não são determináveis exclusivamente a partir da estimulação e da resposta motora, considerando o sistema em isolamento. Se a níveis suficientemente baixos da actividade aferente a questão da "referência" de estados neuronais é facilmente resolvida – os eventos neuronais referem-se às condições que causam a sua ocorrência (é o caso dos neurónios da rã no caso atrás referido) – a situação é tanto mais complicada quanto mais centrais forem os eventos em causa.

De acordo com a teleossemântica, a abordagem do processamento inteligente de informação através da adscrição de conteúdo só pode ser feita tendo como referência a *ligação comportamental*, por mais mediada que esta seja: o carácter adaptativo do comportamento é uma condição necessária da inteligência. Evidentemente aparece assim um limite para a adscrição de conteúdo: quando a ligação comportamental é de alguma forma inadequada, pode tornar-se impossível adscrever conteúdo a eventos neuronais, mesmo que estes tenham efeitos comportamentais. O exemplo de Dennett é o caso de Fido, o cão, que não foi alimentado todo o dia e que, quando se lhe dá um bife, reúne uma pilha de palha, põe o bife no centro e se senta em cima dele[122]. Dado que Fido não se comporta apropriadamente, não é possível adscrever aos seus estados neuronais o conteúdo "isto é comida", mesmo que os dados sobre os

[122] DENNETT 1969: 77.

estados neuronais aferentes o parecessem indicar. No entanto, não é simples atribuir um conteúdo alternativo. O comportamento parece poder corresponder aos conteúdos "Eu sou uma galinha e isto é um ovo", mas igualmente aos conteúdos "Se eu fingir que sou uma galinha, amanhã dão-me o dobro do bife" ou "Vale a pena morrer de fome para dar cabo da cabeça aos psicólogos"[123]. Como qualquer comportamento de uma entidade complexa como Fido é adequado a uma variedade de crenças e desejos, o único traço que pode suportar a hipótese correcta de atribuição de conteúdo é a fonte aferente da estrutura que governa o comportamento. Mas esta favorecerá uma das hipóteses apenas no caso de o comportamento ser adequado às condições da fonte: noutras palavras, quando não existem ligações apropriadas, o comportamento não faz sentido. Nos casos de comportamento apropriado, a adscrição de conteúdo traduz-se no uso de linguagem mentalista. Esta é uma forma de lidar com sistemas cognitivos que não envolve qualquer "fantasma na máquina" mas faculta o acesso ao que os sistemas estão a fazer, por contraste com aquilo que (neles) acontece.

Expostos os princípios básicos de uma teoria teleológica do conteúdo Dennett tem pela frente muitos problemas. É possível por exemplo objectar desde logo que a teoria envolve uma insustentável circularidade entre a noção de Sistema Intencional e a adscrição de conteúdo: afirma-se por um lado que Sistemas Intencionais são sistemas físicos aos quais convém adscrever conteúdo de modo a prever o seu comportamento, por outro que o conteúdo é uma interpretação ou descrição intencional de determinados sistemas físicos. Por outro lado, a questão do conteúdo cruza-se com a questão da informação e armazenamento físico desta, uma questão que não é propriamente psicológica, sendo duvidoso que a informação possa ser tratada de maneira exclusivamente interpretativa. Outro problema da teoria diz respeito à individuação de conteúdos. A adscrição de conteúdo não é, para Dennett, local mas global: apenas se pode atribuir estados com conteúdo a sistemas considerados como totalidades, uma vez tendo retraçado (no caso de sistemas físicos dotados de sistemas nervosos) as causas normais dos eventos neuronais na periferia aferente e os seus efeitos eferentes. Ora, esta regra aparentemente impossibilita a adscrição de conteúdo a eventos estados e estruturas que relativamente centrais num sistema nervoso de

[123] DENNETT 1969: 77-78.

grandes dimensões, o que é exactamente o caso da maioria das "crenças" humanas. No entanto e apesar destas dificuldades que são, aliás, localizadas por Dennett, algo se torna desde logo óbvio: Dennett pensa que a "informação" visada com a adscrição de conteúdo (adscrição que é feita usualmente, pelo menos por nós, humanos, de forma linguística) não é fisicamente armazenada sob forma linguística. É essencial para Dennett acentuar que a forma linguistica não é a quintessência dos veiculos de significação. Assim, embora seja incontornável no nosso tipo de descrição interpretativa do mental, a linguagem não é um traço profundo da natureza do mental. Apesar dessa decisão, não há forma de escapar a problemas específicos que surgem quando se trata de atribuir significação a eventos neuronais por meio da linguagem, nomeadamente: (1) o problema de saber quando passar a exprimir o conteúdo em termos de objectos e estados no mundo exterior, por oposição a acontecimentos nos trajectos sensoriais, (2) o problema do grau de especificação do conteúdo atribuído (as linguagens naturais humanas facultam um "grão demasiado fino" para a expressão de inúmeros fenómenos mentais, nomeadamente não humanos[124]).

Outro grande problema que a linguagem traz ao filósofo da psicologia é o facto de ela provocar interferências entre o tratamento psicológico e o tratamento epistemológico do mental. Se ao filósofo da psicologia cabe explicitar a não fundamentalidade da linguagem nos fenómenos mentais e evitar conscientemente o risco da especificidade excessiva nas atribuições de conteúdo mental feitas por meio da linguagem, na medida em que se assume que expressões verbais não são os veículos fundamentais da significação, pois só têm significação enquanto utilizadas em afazeres (que não são em última análise linguísticos) de sistemas, de um ponto de vista epistemológico não é possível ignorar que qualquer explicação teórica depende do uso de linguagem. A relação entre as duas perspectivas, psicológica e epistemológica, é ainda complicada pelo facto de as descrições intencionais não possibilitarem previsões estritas, embora sejam usadas precisamente para fazer previsões. Isso acontece porque previsões do comportamento feitas ao nível intencional, supondo a racionalidade, não são previsões de eventos nos sistemas nervosos nem são de modo algum previsões seguras. *A*

[124] Será que, por exemplo no caso de Fido, se poderia afirmar que ele tem o conceito de "bife", por oposição a algo de muito mais geral, como "comida"?

coerência racional não é um requisito do funcionamento neuronal, no entanto inferências feitas a partir de descrições intencionais pressupõem uma racionalidade ideal, algo do qual não existe qualquer garantia. Este elemento de pressuposição não é psicológico no sentido empírico, descritivo, ficando em aberto a sua natureza. Quanto a uma possível e mais profunda "natureza de linguagem" dos fenómenos mentais – que será postulada na futura e influente hipótese, de J. Fodor, de uma Linguagem do Pensamento – Dennett considera em C&C que a hipótese implicaria uma regressão infinita, embora admita não estar fora de questão que a analogia com a escrita cerebral tenha utilidade. No entanto ela parece «meramente substituir o homúnculo no cérebro por um comité»[125]. A regressão só pode parar com sistemas que armazenam, transmitem e processam informação cuja forma não é sintáctica. Resumindo a sua posição acerca das relações entre mental e linguagem, Dennett afirmará no artigo *Brain Writing and Mind Reading*[126] que não existe linguagem mais profunda do que aquela que utilizamos "cá fora".

1.2.4 *Os níveis pessoal e sub-pessoal de descrição e explicação. Como tratar o nível pessoal. A dor como exemplo.*

A teoria da adscrição de conteúdo é ilustrada em C&C com um curioso exemplo: a dor (o exemplo é curioso pois parece envolver a consciência, uma questão adiada para a segunda parte da teoria da mente). A dor ilustra a diferença entre o nível pessoal de descrição e o nível sub-pessoal de explicação. Apenas existe dor (como apenas existem pensamentos) ao nível pessoal, o nível descrito em "linguagem da mente". No entanto, os termos da linguagem da mente não referem e portanto nada do que é descrito em linguagem da mente (dores, pensamentos) existe individuadamente. Na prática isto significa que a resposta à pergunta "será a dor alguma coisa a mais para além daquilo que deve ser referido numa descrição sub-pessoal neurofisiológica?" deve ser negativa. Num fluxograma como os utiliza-

[125] DENNETT 1969: 87. Dennett virá a renegar esta crítica, não porque aceite a Hipótese da Linguagem do Pensamento mas porque a crítica ao comité deixa de parecer pertinente, com a importância que os "agentes" foram assumindo na sua teorização. Como será dito em *Brainstorms*, «Homunculi are boggeymen only if they duplicate entire the talents they are rung in to explain. If one can get a team or commitee of relatively ignorant, narrow-minded, blind homunculi to produce the intelligent behavior of the whole, this is progress» (DENNETT 1978: 123).

[126] Cf. DENNETT 1978c.

dos habitualmente em psicologia cognitiva e neurofisiologia para exibir a decomposição funcional dos circuitos nervosos, existirão "caixas" para, por exemplo, nociceptores, estruturas do sistema límbico, estruturas do córtex somatossensorial, centros de controlo, comandos efectores, etc. No entanto, *não existirá uma caixa mais* (situada por exemplo entre a análise perceptual de alto nível e os centros de controlo) para a dor. A dor sentida ela própria não pode aparecer num diagrama que pretende ser um diagrama de causas e efeitos. A relação entre as caixas é causal e não é em linguagem causal que podemos falar de dor.

Considerar que a dor "existe" apenas ao nível pessoal é, segundo Dennett, uma aplicação das lições de Wittgenstein e Ryle. Descrevendo a situação à maneira do Wittgenstein das *Investigações*, temos a situação seguinte. O sistema A, descrito pelo diagrama, produz a elocução "Sinto dor". Qual será o estatuto desse pronunciamento? Se alguém perguntar a A "Porque tiras a mão do fogo?", A responderá "porque dói". E apenas isso: não se trata de relatar um controlo, um controlo do comportamento iniciado *ex nihilo* por A . "Sinto dor", "dói" são *Äusserungen*, partes do comportamento de A e não relatos de observação de factos internos. A não é o seu próprio neurofisiologista, não sabe responder causalmente, em termos de eventos nervosos, não sabe, a partir da experiência pessoal, dizer o que está no diagrama das relações causais. Perante mais perguntas feitas à maneira de Wittgenstein, como por exemplo "Como é que distingues uma sensação dolorosa de uma outra que não o é?", A só poderia responder dizendo "Distinguindo". A não sabe explicar como distingue ou porque distingue uma sensação dolorosa duma que não o é: ele não sabe o que faz, apenas sabe fazer. Uma pessoa não vive ao nível de si própria que lhe permitiria produzir descrições com conteúdo epistémico dos processos causais responsáveis pela sua fenomenologia. Para quem diz "Sinto dor", a explicação pára cedo e não é possível perguntar mais nada na linguagem dos processos mentais acedidos em primeira pessoa. Não obstante, quando a descrição pessoal pára, inicia-se a explicação sub-pessoal, a qual apenas pode ser mecânica. O nível pessoal é definido por contraste como aquele cuja descrição não é mecânica, científico-natural. Ao nível pessoal não podem ser postulados legitimamente episódios mecânicos (de facto, as próprias expressões "estados" e "processos" mentais, difundidas na filosofia da mente, seriam incómodas para Wittgenstein e Ryle).

Em suma, as dores existem para pessoas e não para cérebros. No entanto, distinguir dores como dores não é uma *actividade* de pessoas. Seja o que fôr que se passa no sistema nervoso para que uma dor apareça e seja descrita a nível pessoal, não é a pessoa que faz isso ou que sabe disso: «O poder da pessoa de discriminar dores é um facto bruto, não sujeito a mais perguntas e respostas»[127]. A asserção "Sinto dor" exemplifica o estatuto das asserções feitas em linguagem da mente. Ao nível pessoal, o tratamento daquilo que é dito só pode ser holista e não-referencial. Dado o holismo não poderão ser legitimamente supostas individuações como "o pensamento p da pessoa A no instante t" nem poderá ser colocada a questão da identidade ou não identidade do mental com o físico individuadamente. Não fazem portanto sentido perguntas como "O que é que há no cérebro que é esta dor?" ou "O que é que há no cérebro que é este pensamento?"" pois pensamentos e dores não existem ao nível sub-pessoal. Feita a distinção entre níveis, importa sublinhar que o nível pessoal não é o único nível que interessa ao filósofo da mente, embora essa posição possa ter sido defendida a partir de interpretações de Wittgenstein e Ryle. Parte da tarefa do filósofo da mente é precisamente relacionar os níveis[128].

1.3 *O tratamento do problema da consciência em* Content and Consciousness*: inícios de uma teoria deflacionária da consciência.*

Na medida em que o programa de Dennett consiste na eliminação do fantasma na máquina e respectivas observações, um primeiro passo que se impõe é saber em que redunda a nossa prática de introspecção, já que a introspecção parece ser uma situação que alguém se observa como ser mental e se "relata". Em C&C a análise da introspecção mostra que pela atenção ao nível sub-pessoal se torna claro que a "consciência" não é um traço da mente mas vários. O problema da introspecção consiste em saber o que é que esta pode ser se não for uma "observação do interior". Para Descartes, por exemplo, seja o que fôr o mundo, eu tenho autoridade e certeza quanto àquilo que penso que ele é (i.e. quando "viso" den-

[127] DENNETT 1969: 93.

[128] «The recognition that there are two levels gives birth to the burden of relating them and this is a task that is not outside the philosopher's province» (DENNETT 1969: 95).

tro de mim o mundo enquanto mundo pensado tenho autoridade na descrição daquilo que encontro). Mesmo que eu pense uma inexistente quimera, estou numa situação de certeza acerca do seu ser pensado[129]. Esta situação cartesiana de certeza quanto aos conteúdos dos nossos pensamentos conduz à defesa da autoridade e incorrigibilidade do pensador ou meditador. Estas características aparecem, aliás, naturalmente numa descrição intuitiva daquilo que pensamos ser enquanto pensantes; a teoria científica da mente instala-se sempre num possível interregno de cada entidade consciente e pensante de si a si, admitindo que afinal não é a própria pessoa que pode descrever melhor, ou mais completamente, aquilo que é o caso quando pensa.

1.3.1 *O funcionalismo segundo H. Putnam. A certeza introspectiva de um ponto de vista funcionalista.*

Em C&C o esclarecimento do acesso introspectivo baseia-se nas ideias funcionalistas de H. Putnam, o que significa desde logo que o problema da consciência é interpretado como dizendo respeito ao acesso a si de um sistema, nomeadamente o *auto-acesso que é ocasião de relatos linguísticos*. Os relatos introspectivos são para Dennett – que exclui, assim, à partida a possibilidade de o verdadeiro problema consistir no facto de «que o introspectado seja fenomenal»[130] – a matéria prima da teoria da consciência. O auto-acesso é analisado mediante a distinção entre dois sentidos de apercebimento (*awareness*). O primeiro nomeia a possibilidade de *expressão linguística de uma interioridade*, o segundo nomeia o *controlo por um sistema das actividades próprias*. A estes sentidos do apercebimento chama-se em C&C apercebimento-1 e apercebimento-2. A intenção de Dennett é mostrar que não existe um terceiro apercebimento, que seria a consciência, a qual teria sido deixada de fora nas explicações do auto-acesso.

[129] Cf. DESCARTES [1641]: 136 «Todavia admiti anteriormente como absolutamente certas e manifestas muitas coisas que, entretanto, depreendi depois serem duvidosas. Que coisas foram estas? A Terra, o Céu, os Astros, e todas as outras que recebi pelos sentidos. Mas o que compreendia eu delas com clareza? As próprias ideias ou pensamentos de tais coisas, que se apresentavam ao meu espírito. *E ainda agora não contesto que essas ideias estão de facto em mim.* Mas outra coisa era o que eu então afirmava que compreendia claramente pelo hábito de crer, embora na verdade não compreendesse: que havia certas coisas fora de mim, de que procediam estas ideias e às quais eram totalmente semelhantes».

[130] PINTO 1999: 118.

Parece-nos que temos um acesso infalível ao conteúdo dos nossos pensamentos. A descrição cartesiana da intuição dos habitantes da *res cogitans* pelo meditador é estranhamente natural, mesmo que a particular justificação cartesiana dessa capacidade – a existência de ideias na alma – não o seja. Ora, segundo Dennett, autores como Wittgenstein, Ryle e Anscombe deram um primeiro passo acertado para a avaliação da nossa aparentemente cartesiana condição ao considerarem que se as asserções de experiência interior são infalíveis é porque elas não são relatos factuais. Qualquer relato factual pode estar errado, e as asserções de experiência interior não podem estar erradas. Os autores referidos estão certos quando defendem que asserções de experiência interior não são relatos factuais primeiro que tudo porque não há relatos sem relator, e no interior não existe esse relator. Eles sugerem ainda a Dennett a possibilidade de o atestar introspectivo de estados mentais ser não um conhecimento mas uma condição prática, uma capacidade, um comportamento. As asserções de experiência interior só podem ser teoricamente manejadas como sendo elas próprias comportamentos: não existem passos cognitivos intermédios (que seriam identificações de estados mentais) para a sua expressão em asserções exteriores.

No caso específico de Wittgenstein, o argumento segundo o qual os relatos factuais podem precisamente e por definição, estar errados e os relatos introspectivos não podem estar errados, não sendo portanto os relatos introspectivos relatos factuais[131], foi um ponto importante para o abandono do solipsismo e do verificacionismo que chegou a defender. É (posteriormente) a essa constatação que Wittgenstein sugere, nas *Investigações,* que a invulnerabilidade ao erro dos relatos de dor se deve ao facto de esses relatos substituirem o grito, em lugar de o descreverem. As asserções de experiência interior não são portanto relatos com conteúdo cognitivo mas manifestações comportamentais, como gritar. "Eu tenho dores" *não é para Wittgenstein uma proposição genuína mas uma Ausserung*, uma exteriorização que não tem conteúdo epistémico e que não resulta de qualquer observação da realidade (interior).

Ora, o que perturba Dennett nesta argumentação é o facto de ela conduzir à negação da utilidade epistemológica dos relatos de

[131] Cf. DENNETT 1969: 100. As passagens de Wittgenstein citadas são os parágrafos §244, §367 e §370 das *Investigações*. Cf também ZILHÃO 1993: 55, acerca do abandono do verificacionismo por Wittgenstein, precisamente devido ao problema da verificação das asserções da experiência interior. Para que a verificação tivesse sentido no âmbito da experiência interior teria que ser possível cometer erros.

experiência interior. Para Dennett, embora os relatos do auto-acesso não possam ser tomados como uma expressão da "intuição-clara-e-distinta-daquilo-que-é", eles devem ter alguma utilidade epistemológica e as posições de Wittgenstein, Ryle e Anscombe têm algo de implausível devido à rejeição de tal possibilidade. A implausibilidade torna-se mais clara quando se passa de relatos de dor para relatos introspectivos mais elaborados, dificilmente considerados como *Äusserungen* e *avowals*. Se é certo que não é possível considerar que um eu qualquer, na introspecção, observa ou conhece a sua experiência interior, esta consideração não deve conduzir-nos a negar o óbvio, por exemplo que relatos de dor são relatos de *dor*, que podem como qualquer relato ser verdadeiros ou falsos: quando eu digo a um médico que me dói ali não estou apenas a gemer sofisticadamente, quero informá-lo. As exteriorizações, confissões e habilidades, de Wittgenstein, Ryle e Anscombe devem por isso poder ser vistas, de alguma outra maneira como susceptíveis de verdade e falsidade. A diferença da posição de Dennett em relação às posições dos três autores referidos é determinada pela atenção da teoria da mente ao nível sub-pessoal. De facto, e como os três autores defendem, ao nível pessoal as explicações páram. Eles estão no entanto enganados quando pensam que o problema da incorrigibilidade da introspecção tem solução ao nível pessoal. Dennett sugere que é ao nível sub-pessoal que a incorrigibilidade da introspecção deve ser investigada. Para isso utiliza a distinção putnamiana entre estados físicos e estados lógicos.

Um estado lógico de um sistema é, por definição, aquilo que é em virtude das suas relações com inputs e outputs e com outros estados, e não em virtude de quaisquer características físicas específicas. Assim sendo, a situação na qual um sistema se assegura que está num estado lógico só pode ser concebida como a situação em que o sistema se *comporta como se* estivesse nesse estado lógico. E se o sistema se comporta como se estivesse nesse estado lógico é porque o sistema *está* nesse estado lógico, já que estar num estado lógico consiste exactamente em "comportar-se" de uma dada maneira. Então, quando um sistema relata que está num estado lógico, esse relato sai directamente do estar nesse estado lógico, não implica computação ou inspecção de indícios para chegar à conclusão de que está nesse estado lógico. Se alguém perguntar como é que o sistema sabe que está num estado lógico, a única resposta possível é, como diz Dennett na sequência da exposição de Putnam em *Minds and Machines*, "degenerada": o sistema

sabe que está num estado lógico determinado estando nesse estado lógico. O contraste que Dennett quer realçar é o seguinte: os humanos, e também as máquinas, são capazes de alguma monitorização dos seus estados internos, nomeadamente dos seus estados *físicos* internos. Quando um humano monitoriza ou sonda os seus estados físicos internos e relata as condições a que acede, faz sentido querer saber como é que ele se assegura daquilo que reporta. O humano poderá responder a uma questão com esse teor dizendo qual foi a sucessão de estados por que passou. No entanto quando o estado relatado pelo sistema é um estado lógico, i.e. funcionalmente individuado segundo Putnam, a mesma questão não faz qualquer sentido. Nos exemplos originais de *Minds and Machines*, Putnam compara, para mostrar a diferença nos erros possíveis, a máquina que relata "Estou no estado A" com a pessoa que diz "Sinto dor" e contrasta esses casos com os casos em que a máquina relata "O tubo de vácuo 312 falhou" e a pessoa diz "Tenho febre".

De entre os relatos de estados sub-pessoais feitos por uma pessoa, são os relatos de estados lógicos de um sistema que são invulneráveis a erros, não os relatos de estados físicos. É exactamente esse o caso dos relatos introspectivos humanos. A incorrigibilidade deve portanto ser explicada pelo estatuto dos estados lógicos. Como Putnam viu, certos relatos são incorrigíveis precisamente porque são relatos de estados lógicos. Os estados lógicos não podem ser *erradamente* identificados, simplesmente porque não têm que ser (nem podem ser) *identificados*. De modo a estabelecer este ponto putnamiano, Dennett imagina, como Putnam, uma máquina capaz de fazer *prints* de relatos dos seus estados internos[132]. Dennett quer mostrar que essa máquina, mesmo que fosse uma máquina "quase-skinneriana" (i.e. uma máquina que produzisse por hipótese respostas verbais skinnerianas a estímulos visuais, deixando sair com toda a simplicidade – sem mentiras, sem decisões, sem perguntas – relatos verbais daquilo que "vê") partilharia com os huma-

[132] DENNETT 1969: 104. As especulações acerca da máquina que percepciona e faz relatos introspectivos são as primeiras de muitas outras descrições de dispositivos cognitivos imaginários na obra de Dennett. As especulações acerca da produção de linguagem serão continuadas em DENNETT 1991, *How Words do Things with Us*. As intuições são basicamente as mesmas: por exemplo a máquina de perceber aí proposta teria câmaras de TV como orgãos sensoriais mas nela não teria que haver écrans. Dennett salvaguarda que, ao considerar este exemplo, é preciso cuidado para não pensar que relatos saem directamente de estados lógicos interessantes de humanos, como os estados lógicos estudados por linguistas como Chomsky, por exemplo.

nos uma característica muito importante: ela não poderia estar enganada acerca dos seus estados mentais (i.e. dos estados lógicos). Poderia no entanto falhar a relação input-relato de duas maneiras diferentes (além de, evidentemente, como qualquer máquina, poder funcionar mal em qualquer uma das suas partes). Por um lado, poderia haver falhas devidas a um mau funcionamento anterior ao input do centro da fala[133], fazendo com que o output do centro da fala fosse enganoso relativamente ao mundo externo. Por outro lado, poderia haver erros posteriores à análise no centro de fala, aos quais Dennett chama erros de expressão. O que é importante no exemplo é notar que, porque a máquina não tem outro interior que não seja o seu próprio funcionamento (exactamente como o nosso cérebro), ela não pode enganar-se acerca daquilo que lhe *parece* ser. Ela pode "estar errada" acerca do que é o mundo, pode cometer erros "verbais", mas não pode identificar erroneamente o output dos processamentos internos, pois o output dos processamentos internos simplesmente é o mesmo estado lógico que o input do centro de fala. Logo, ela não pode ser "responsabilizada" por input fraudulento ou por erros de análise de input.

A identidade entre o output dos processamentos internos e o input do centro da fala é aqui obviamente postulada, assumindo já um ponto de vista funcionalista. É no entanto essa identidade que traz a imunidade ao erro do meditador que pratica introspecção. A incorrigibilidade não resulta portanto da ponderada apreciação dos estados mentais próprios num retiro interior mas do que se passa no cérebro. A situação do auto-acesso de um humano que produz um relato introspectivo não é uma apreciação ponderada dos estados interiores, não é uma descrição de eventos nervosos, é uma expressão do conteúdo desses eventos. Isto é assim porque «o conteúdo de um evento, ou de um estado lógico do qual um estado físico é a realização, não é uma questão de qualidades físicas intrínsecas que possam ser relatadas ou descritas, mas de capacidades funcionais, incluindo a capacidade funcional de iniciar (excluído o mau funcionamento) a enunciação da qual se dirá que exprime o conteúdo em alguma linguagem»[134]. Se se chamar pensamentos àquilo

[133] O centro da fala é a estrutura cognitiva que permite o relato linguístico. O postulado deste centro de fala funcionalmente caracterizado e não fisiologicamente localizado é essencial na teoria da consciência de C&C. Ele é no entanto completamente polémico.

[134] DENNETT 1969: 112.

que é relatado por um humano (entenda-se: mesmo relatado para si próprio) a partir do auto-acesso, esses pensamentos não poderão ser identificados com o que quer que seja ao nível sub-pessoal. Não existem entidades determinadas (os pensamentos) que estariam a ser relatadas ("expressas" no sentido de estarem anteriormente no interior sendo depois exteriorizadas). O facto de as pessoas serem capazes de dizer o que estão a pensar não obriga a teoria da mente a um compromisso com a existência de pensamentos. De outro modo, diz Dennett, se atribuíssemos pensamentos a pessoas, teríamos que os atribuir também à máquina de perceber atrás referida e não pedimos aos engenheiros que os pusessem lá.

O facto de as pessoas serem capazes de dizer o que estão a pensar é assim explicado através do estatuto de estados lógicos no nível sub-pessoal, nível no qual não existem nem pessoas nem pensamentos. É isto que por exemplo R. Rorty[135] considera um defeito da teoria de Dennett. De acordo com a teoria apresentada, a capacidade de apercebimento humano sofisticado é ligada à capacidade de fazer relatos introspectivos. Ora, Rorty simplesmente duvida que a introspecção possua tal estatuto exemplar no pensamento humano. O conhecimento não inferencial, i.e. incorrigível, não é necessariamente introspecção. Rorty considera mesmo que Dennett amalgama em C&C o apercebimento humano (que é ou pode ser linguístico) com a capacidade de falar acerca da mente própria (i.e. de levar a cabo introspecção linguística). É esta última que é identificada por Dennett com a capacidade de fazer relatos não inferenciais. No entanto, segundo Rorty, a incorrigibilidade não tem nada a ver com o nível sub-pessoal: o apelo de Dennett à conexão directa para explicar a incorrigibilidade é um cartesianismo infeliz. Esta diferendo entre R. Rorty e D. Dennett acerca da importância ou de algo exterior ou de algo interior para explicar a incorrigibilidade não terá mais desenvolvimentos em C&C, onde o modelo descrito é a última palavra.

1.3.2 *Os sentidos de apercebimento. Apercebimento e controlo (apercebimento-2), apercebimento e expressão verbal (apercebimento-1).*

As relações entre acesso introspectivo e controlo do comportamento são complexas. Nem sempre aquilo que pode ser objecto de

[135] RORTY 1972a e RORTY 1972b.

relato por um sistema num momento dado é aquilo que está a ser relevante para o controlo do comportamento do sistema nesse momento. Há aliás ocasiões em que uma interferência introspectiva perturba o controlo do comportamento (como poderia acontecer com um pianista que, durante a execução de uma peça bem conhecida, começasse a ficar fascinado com a independência dos movimentos dos seus próprios dedos). Para Dennett, a correcta concepção destas relações envolve uma decisão teórica. Embora apercebimento (*awareness*) e consciência (*consciousness*) sejam usualmente considerados termos sinónimos, esse é um mau uso: a teoria da mente não deve ater-se à linguagem comum, podendo e devendo servir-se de reformas e precisões artificiais. Dennett retem o termo apercebimento para o explorar teoricamente, nomeadamente pelas relações do apercebimento com o controlo do comportamento e com a expressão linguística. Um primeiro sentido de apercebimento diz respeito à percepção ou registo por um sistema daquilo que há no seu entorno, de modo a poder mover-se e comportar-se apropriadamente (é isso que nos faz dizer que uma abelha tem que ter *apercebido* a árvore para a contornar, um pássaro tem que ter *apercebido* o gato que salta para fugir). Mas isto não deve comprometer-nos a afirmar que a abelha apercebe conscientemente a arvore como "árvore" ou que o pássaro apercebe conscientemente o gato como "gato". Por outro lado, os humanos produzem frequentemente relatos introspectivos em que expressam controlos comportamentais (nomeadamente como resposta à questão "O que estás a fazer?"). Mas não existe certamente uma relação necessária entre o facto de um apercebimento ser objecto de relato introspectivo e o facto de ele ser responsável pelo controlo comportamental. Aliás, o apercebimento introspectivo tanto parece ser uma condição necessária para o controlo comportamental como perfeitamente separável deste (ou mesmo impeditivo do controlo). Dennett pretende fazer ver antes de mais que o controlo comportamental não está necessariamente ligado a qualquer apercebimento explícito, clarificando em seguida o que se entende por apercebimento explícito. O exemplo para esta separação é o homem que pensa "devo ter apercebido o copo a chegar aos meus lábios ou não teria bebericado". Ele não está a fazer introspecção mas a especular, pois não tem mais provas dessa afirmação do que um observador exterior teria. Um certo apercebimento é condição necessária do controlo bem sucedido do comportamento, no entanto, o controlo do comportamento que

está a efectuar-se pode não estar disponível para introspecção e relato. Observações como estas constituem razão para a seguinte distinção entre apercebimento-1 e apercebimento-2.

> Apercebimento-1 O sistema A apercebe-1 que p no instante t SSE p é o conteúdo do estado de *input* do centro de fala de A no instante t
>
> Apercebimento-2 O sistema A apercebe-2 que p no instante t SSE p é o conteúdo de um evento interno em A no instante t que guia efectivamente o comportamento actual de A[136]

Em C&C, o termo apercebimento (*awareness*) tal como é usado comummente (ao nível pessoal, como sinónimo de consciência) é substituído por estes dois termos artificiais que tomam o sistema global como referência mas para os quais existem critérios sub-pessoais. As duas definições apresentadas são fundamentais para distinguir a consciência que um humano tem de um objecto da consciência que um animal tem do mesmo objecto. Nomeadamente, estipula-se que um animal só tem apercebimento-2, ficando o apercebimento-1 ligado, por definição, à posse de linguagem. As máquinas cibernéticas podem ter apercebimento-2 (ou mesmo apercebimento-1). Além destes dois tipos de apercebimento, não existe um terceiro que seria distintivamente humano. Isto significa que a fenomenologia de um sistema é supostamente capturada pela noção de apercebimento-1. Sem apercebimento-1, haveria, num sistema que se comporta, apenas controlo e muito menos (ou nada de) interior.

Aplicando a distinção entre apercebimento-1 e apercebimento-2, Dennett propõe-se analisar de novo o sistema-que-apercebe atrás referido. Este será dividido por uma linha imaginária em duas partes funcionais, um analisador aferente e um centro de fala. Esta linha imaginária, a linha do apercebimento (*awareness line*), é obviamente muito pouco clara no cérebro de uma pessoa. No futuro,

[136] DENNETT 1969: 118. Dennett virá a abandonar a distinção estrita entre apercebimento-1 e apercebimento-2 aqui proposta, sugerindo a existência de um contínuo (cf. DENNETT 1994f, *Get Real*, Resposta a D. Rosenthal) ou indistinção ou mesmo "contradição", apoiado nomeadamente em experiências psicológicas de A. Marcel nas quais são pedidas aos sujeitos respostas redundantes a um mesmo estímulo, por exemplo dizer "Sim", premir um botão, assinalar de forma convencionada significar que "sim". Estas respostas não são sempre coincidentes (cf. também DENNETT 1991: 248). O problema é evidentemente *o excesso de unificação ligado à noção de apercebimento-1*. Se é a linguagem que sustenta o apercebimento-1, Dennett virá a assumir que os modelos de produção de linguagem não permitem supor uma tal unificação.

Dennett virá aliás a afirmar que é um erro sequer postular uma *awareness line* (este seria aliás o principal erro da teoria da consciência proposta em C&C). A ideia da linha do apercebimento surge da constatação de que os anéis de *feed back* – que supostamente corrigem por exemplo maus funcionamentos nos vários níveis de produção de asserções linguísticas e que estariam presentes no centro de fala – não poderiam estender-se ao analisador aferente.

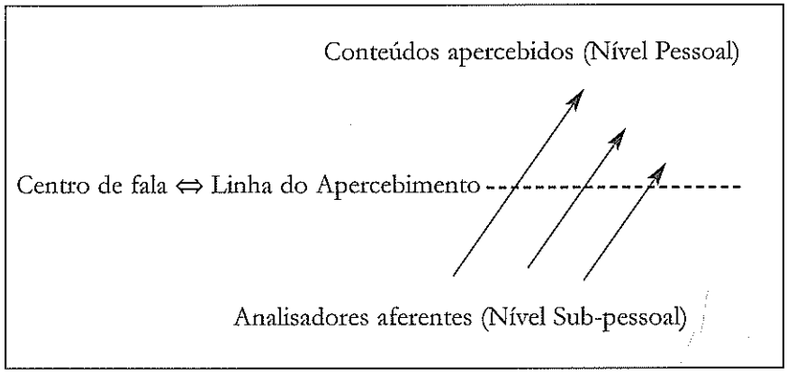


FIG. 1 O modelo de apercebimento de *Content and Consciousness*

O bordo ou fundo do centro de fala assim definido seria a linha do apercebimento. Há apercebimento-1 de alguma coisa quando a linha é cruzada. Se um sinal não cruza a linha, o seu conteúdo não pode ser expresso. Não haverá então apercebimento-1, mesmo que o evento esteja presente no controlo comportamental. Quanto à vantagem do apercebimento-1, considerando a hipótese de ele não ser unicamente um produto lateral do funcionamento do sistema, Dennett começa por notar que o apercebimento-1 é importante para o controlo de alto nível do comportamento, que envolve a correlação de informação de diversas fontes, correlação esta que é uma condição essencial da acção complexa, e que é muito importante por exemplo em termos éticos, como se verá no Capítulo 4. Se é verdade que sem centro de fala, não haveria apercebimento-1 nem introspecção relatante, nenhum dos dois sentidos de apercebimento é, apesar disso, o mais importante ou o único no modelo de C&C. Como ideia central da teoria da consciência fica o seguinte: um sistema não tem que aperceber-1 um objecto para o reconhecer e lidar apropriadamente com ele, mas tem certamente que

apercebê-lo-1 para dizer que o reconhece e para dar a razão do seu comportamento. É certo que toda a teoria está suspensa da noção (imprecisa) de centro da fala, da qual depende o apercebimento-1.

Dir-se-ia que com a apresentação da teoria funcionalista da certeza introspectiva e com a proposta de distinção de dois sentidos diferentes de apercebimento, não se iniciou ainda sequer o tratamento da consciência como estado de alerta, estado de vigília (*awakeveness*), aquele estado que falta quando um organismo está em coma, ou adormecido, ou inconsciente. A objecção intuitiva usual à teoria dennettiana da consciência é precisamente que nela falta alguma coisa e exactamente o mais importante. Esse é o primeiro sintoma de que se está perante uma teoria deflacionária. De certo modo, Dennett parece levar ao limite a perturbadora e influente afirmação de Karl Lashley segundo a qual «Nenhuma actividade da mente é alguma vez consciente» (*"No activity of mind is ever conscious"*)[137]. Com as noções de apercebimento Dennett explica processos vulgarmente considerados como característicos da consciência ou acompanhados sempre de consciência e no entanto esses processos parecem poder acontecer na ausência de consciência fenomenal. No entanto, com a sua proposta Dennett pretende precisamente negar a especificidade da consciência fenomenal, considerada como algo que não seja meramente a especificidade de um certo tipo (certamente humano, mas não necessariamente apenas humano) de processo ligado ao apercebimento-1 e por ele esgotado. O tratamento da consciência proposto produz vários tipos de desilusão, mais ou menos justificada. Por um lado, Dennett não respeita o privilégio da linguagem comum, antes estipula. Por outro lado, não respeita o mandamento ryleano que ordena que não se façam especulações mecânicas, defendendo precisamente uma teoria cognitiva sub-pessoal mecânica. Dennett defende no entanto que a transgressão destas duas regras tácitas de uma certa metodologia filosófica é essencial para o desenvolvimento da filosofia da mente. Finalmente, em termos da própria teoria, aquilo que se ganha com a transgressão das regras é um conceito de apercebimento-1 absolutamente austero. Toda a teoria da consciência redunda numa afirmação única e minimalista: tudo o que podemos fazer quanto aos eventos resultantes do processamento sub-pessoal de informação que cruzam a linha do apercebimento é expressá-los ou não os expressar. De acordo

[137] Citado em DENNETT 1969: 128.

com a teoria, a distinção entre a consciência de humanos e a consciência de animais como cães ou gatos depende da capacidade humana de produzir relatos linguísticos introspectivos, nomeadamente relatos dos estados internos responsáveis pelo controlo, por oposição ao caso em que o controlo simplesmente acontece.

Resta saber se Dennett não estará a confundir coisas diferentes na noção de consciência como apercebimento-1. De facto, na noção de apercebimento-1 juntam-se duas coisas: (a) *o facto de os humanos serem capazes de falar* e (b) o *facto de os humanos serem capazes de falar de uma certa maneira sobre os estados mentais próprios*. No quadro de C&C, a diferença entre o apercebimento de que são capazes por um lado os humanos e, por outro lado, os animais fica, como se viu, a cargo da diferença entre a capacidade humana de produzir relatos linguísticos introspectivos directos acerca de si e o simples controlo comportamental. O carácter especial do apercebimento humano advém portanto da capacidade de fazer relatos *introspectivos*. Mas talvez não haja razão para essa ligação: a introspecção não está por toda a parte na maneira humana de pensar, ao contrário do que Dennett propõe. Quando alguém relata o que faz, mesmo que essa pessoa esteja envolvida numa tarefa intelectual sofisticada, não se descreverá necessariamente o relato da sua tarefa intelectual como uma prática de introspecção. Dennett é levado a essa tese por identificar sob o título "linguagem" a capacidade de falar e a capacidade de falar de uma certa maneira sobre os estados mentais próprios. Tem aliás que o fazer para considerar que é possível explicar a incorrigibilidade humana através de uma teoria funcionalista da produção de linguagem e de uma teoria mecânica (i.e. sub-pessoal) da introspecção, apoiada na noção putmaniana de estados funcionais. Por exemplo R.Rorty[138] contrapropõe que a incorrigibilidade associada ao facto de os humanos falarem é uma questão de prática social e não do estatuto de estados sub-pessoais. Repare-se que se a razão da incorrigilidade fôr assim exteriorizada, já não será legítimo apelar à incorrigibilidade considerando que se oferece uma teoria da consciência.

Basicamente, a noção dennettiana de incorrigibilidade como conexão directa é, como o próprio Dennett virá a admitir, ainda cartesiana. A noção depende de outras duas noções – "centro da fala" e "linha do apercebimento" – que não são evidentemente mais do

[138] RORTY 1970 e RORTY 1972a.

que metáforas. Neste quadro, note-se, a consciência acaba por depender do que conta como fala e o único critério de fala que Dennett oferece é o facto de alguma coisa ser produzida pelo centro da fala. Apesar desta circularidade, toda a noção de apercebimento-1 depende da noção de centro da fala. É o centro da fala na sua ligação com a linha do apercebimento (que sustenta a noção de apercebimento-1) que separa a análise cognitiva sub-pessoal da expressão pessoal. É no entanto completamente arbitrário no interior da própria teoria decidir o que conta como fala, sendo por exemplo bastante duvidoso que seja necessário e suficente que um acontecimento seja linguístico para que ele conte como acto de fala produzido por um humano. Além do mais, dada a não restrição, de acordo com a teoria, do mental ao linguístico não foi considerado o que numa mente não humana poderia ser análogo à fala. O caminho que fica aberto, e que será explorado em CE, passa por supor que o papel da linguagem na consciência é muito mais interessante e indirecto.

1.3.3 *Imagens, qualia, preenchimentos e cores. Percepções de ausência ou ausência de percepção.*

C&C foi escrito antes dos revolucionários estudos acerca de imagens mentais[139] associados especialmente aos nomes dos psicólogos R. Shepard e S. Kosslyn[140]. De acordo com esses estudos, imagens internas são geradas para aparição interna, podendo ser inspeccionadas, manipuladas e transformadas quando seres humanos imaginam. Pretensões semelhantes ganharam progressivamente legitimidade dos anos 70 até ao presente. Por essa razão, Dennett virá a considerar que o capítulo VII de C&C *(Mental Imagery)* foi quase inteiramente posto de lado por investigações empíricas subsequentes. Ao contrário do que Dennett pensava quando escreveu o referido capítulo, o "mito" das imagens mentais não está prestes a perder o seu poder. Pelo contrário, a questão das imagens mentais (*imagery*) tornou-se uma respeitável área de investigação da ciência cognitiva. Ora, a questão das imagens mentais é importante para

[139] Para um enquadramento histórico destes estudos e um resumo dos problemas teóricos envolvidos, cf. ANDERSON 1998: 204, *Imagery and Spatial Representation*, GARDNER 1985: 233, *Mental Imagery: A Figment of the Imagination?*, KOSSLYN 1995, *Imagery* e STILLINGS, WEISLER, CHASE, FEINSTEIN, GARFIELD & RISSLAND 1995:42 e segs.

[140] Cf. SHEPARD & METZLER 1971 e KOSSLYN 1978.

uma teoria da consciência cognitivista e intelectualista[141] como a de Dennett na medida em que a existência de imagens, trazendo consigo as questões do espaço fenomenal e dos *qualia*, aparentemente se ergue contra uma definição de consciência estritamente comprometida com a linguagem. A questão é aliás também importante para o destino do cognitivismo simbólico em geral, i.e. para a ideia segundo a qual a possibilidade de mente seria suportada por estados e processos com características de linguagem: a controvérsia entre os psicólogos Z. Pylyshyn e S. Kosslyn acerca da existência ou não existência de um *medium* específico para a representação mental de imagens, analisada no Capítulo 3 deste trabalho, virá a ser nesse sentido reveladora.

No capítulo de C&C em que se trata de imagens mentais Dennett aborda as questões dos *qualia* (nomeadamente a questão das cores, exemplos usuais de propriedades reais e emergentes que exigem um espaço mental fenomenal) e do preenchimento (*filling-in*) de detalhes em falta naquilo que é imaginado e percebido. O ponto geral do capítulo é a crítica da identificação, na teoria da percepção em geral (quer se trate de percepção do mundo exterior ou de imagens interiores geradas), da ausência de percepção com a percepção de uma ausência. Dennett aborda a questão das imagens mentais com a intenção de pura e simplesmente abandonar a noção de imagem, que seria tão homuncular como a ideia de escrita cerebral (tanto para "*image-watching*" como para "*brain writing*" e "*brain reading*" deveria haver alguém a observar). Esta decisão vem a revelar-se exagerada e o recuo de Dennett é um caso interessante para verificar até que ponto dados empíricos podem refutar princípios gerais em filosofia da mente.

Dennett define *imagem* como a *representação de alguma coisa em virtude da posse* (pela imagem) *de uma qualidade ou característica em comum com aquilo que é representado* (por exemplo forma ou cor, no caso de imagens visuais). Uma imagem é o tipo de representação que deve por definição assemelhar-se àquilo que representa, por oposição à representação que representa o representado pelo facto de, meramente, deter um papel em algum sistema. A questão das imagens redunda então em investigar se existirão elementos na percepção actual (exterior ou internamente gerada) que representam em virtude de se assemelharem àquilo que representam. Ora, Dennett defen-

[141] Os termos, certeiros, são usados por Dretske em relação a Dennett em DRETSKE 1994.

de que não há lugar para imagens nesse sentido na explicação cognitiva sub-pessoal das experiências imagéticas do sujeito. Nomeadamente na percepção visual a última imagem é a imagem retiniana. Os exemplos com os quais Dennett pretende apoiar a sua conclusão quanto ao nível sub-pessoal são os seguintes:

Exemplo 1[142]: Para que uma imagem funcione como imagem deve existir "alguém" que a reconheça como tal. Daí o absurdo de uma situação em que alguém colocaria uma câmara de televisão no seu automóvel, ligada a um receptor sob o *capot* de modo a que o motor "possa ver para onde vai". O ponto deste exemplo é evidenciar que embora uma imagem tenha sido disponibilizada, nenhuma coisa análoga a um "perceptor" existe, logo nada "funciona como imagem". A conclusão é que para uma imagem funcionar como imagem na percepção ela teria que ser a matéria prima e não o produto final da percepção, caso contrário teriam que ser supostos sucessivos "perceptores" até ao infinito. Se se considera que "debaixo do *capot*" é o análogo do interior de um cérebro humano compreende-se porque pensa Dennett que é errado considerar as imagens como produto final. Só uma testemunha permitiria a existência de imagens sub-pessoais. A ideia de imagem é portanto homuncular, devendo ser abandonada.

Exemplo 2[143]: Assim como não é razoável considerar que a diferença entre as palavras "*ox*" e "*butterfly*" consiste no facto de uma ser mais pesada e mais feia do que a outra também não é razoável considerar que a diferença entre uma representação neuronal de um quadrado e a representação neuronal de um círculo seja uma diferença na forma geométrica de eventos neuronais que seriam imagens. Após o último lugar onde é legítimo falar de semelhança imagética no processo físico da percepção (a retina, no caso da percepção visual) a análise aferente continua, e haverá elementos que são representações, i.e. que devem ser teoricamente tratados como representações, mas em virtude do seu papel num sistema arbitrário, e não em virtude de qualquer característica de forma.

Da situação exemplificada pelos casos 1 e 2 Dennett conclui que *não há lugar para o apelo a imagens na explicação sub-pessoal*. Em relação ao nível pessoal, a conclusão geral é também eliminativista. O

[142] DENNETT 1969: 134.
[143] DENNETT 1969: 135.

exemplo 3 pretende evidenciar as condições não apenas da imaginação como também da percepção visual.

Exemplo 3: Imagina-se um homem alto com uma perna de pau. Ele terá cabelo loiro ou não?[144]. De acordo com Dennett, não existe nem pode por princípio existir, uma resposta para esta pergunta. O cabelo do homem simplesmente não é mencionado na "imaginação de homem". Ora, a não-menção de uma característica x do objecto imaginado não é identificável com a percepção da ausência dessa característica. Não há justificação para supor que qualquer preenchimento é necessário se não há ninguém a ver ou a questionar. A suposição de saturação exaustiva é de acordo com Dennett homuncular e intelectualista. O princípio geral é o seguinte: se não há ninguém (ou nada) a pedir uma resposta (se não há "apetite epistémico"), a resposta não tem que estar disponível e a sua ausência não será notada. Aliás, não se trata propriamente de uma ausência: só se trataria de uma ausência se um determinado elemento tivesse que estar presente, o que não é o caso. O mesmo princípio deve ser observado na própria concepção da percepção visual que não é internamente gerada. Retome-se um exemplo de Dennett.

Exemplo 4[145]: Alguém está num quarto a tarde toda, sem se aperceber da cor das paredes. Não é capaz de dizer de que cor eram as paredes, quando lho perguntam. No entanto, se elas fossem vermelhas, a pessoa tê-las-ia concerteza apercebido conscientemente como vermelhas. Mesmo considerando o contrafactual verdadeiro, nada nos obriga a aceitar que a pessoa tenha necessariamente tido que "aperceber-se sem se aperceber de que apercebe" de que as paredes eram beige, como de facto eram.

Em C&C, como futuramente em CE, Dennett coloca estas conclusões eliminativistas sob a égide de Wittgenstein. A situação que se pretende corrigir é a seguinte: quando se pensa na percepção, há uma tendência para supor que um lugar disponível tem que ser ocupado de qualquer modo. No entanto a ausência de percepção não é uma percepção da ausência mas mais propriamente uma *não-menção*,

[144] O exemplo standard de Dennett para esta questão do preenchimento ou menção é o exemplo do que acontece quando alguém vê um papel de parede coberto com Marylins (Cf. DENNETT 1991: 354 –355).

[145] DENNETT 1969: 140.

o que aparenta as supostas imagens mentais mais a descrições linguísticas do que a figurações. Para o tratamento das questões dos *qualia* e dos *zombies* na teoria da consciência impedir a identificação da ausência de percepção com a percepção de uma ausência é crucial. No entanto, o preenchimento é um princípio explicativo do qual os cientistas cognitivos não desejam abdicar. Embora Dennett não venha a abandonar esta sua posição quanto ao preenchimento, ela tem sido atacada com base em análises psicofísicas[146], o que pressupõe que a questão é empírica e não conceptual. Os exemplos em torno da ausência de percepção e percepção de ausência reaparecerão em força em CE onde o tratamento do problema filosófico da consciência, nos termos preferidos pelos filósofos (o quadro dos *qualia* e dos *zombies*) é precisamente iniciado por uma concessão às descoberta da psicologia cognitiva acerca de imagens mentais.

Em C&C a eliminação das supostas características imagéticas de representações é continuada com um argumento acerca de cores. Deixar de admitir representações imagéticas é deixar de poder considerar as qualidades destas, entre as quais cores. Mas as cores parecem exemplos inamovíveis de qualidades fenomenais reais, emergentes a partir das "qualidades primárias". Mesmo se as cores não são qualidades privadas, inefáveis e incomunicáveis (admitindo que é isso que Wittgenstein prova com a sua análise do uso de termos para cores), nem se encontram entre as "qualidades primárias" permitidas pela física, não sendo à partida portanto nem privadas nem objectivas, um longo passo separa a admissão destes factos da afirmação de que a experiência de cor não existe. A discussão inicia-se com a consideração das cores como propriedades de objectos externos (capacidades reflexivas de superfícies), surgindo o problema do facto de não se conseguir encontrar correspondências *unívocas* entre comprimentos de onda e experiências de cor. O problema das cores é que elas não se comportam como géneros naturais (*natural kinds)*[147]. Segundo Dennett as cores não são, obviamente, propriedades físicas primárias, mas também não são qualidades privadas, fenomenais e emergentes: «as cores são (...) propriedades funcionais. Uma coisa é vermelha se e só se quando é vista em condições normais por humanos normais lhes parece vermelha. Isto significa que há utentes dos termos para cores que são demonstravel-

[146] Cf. CHURCHLAND & RAMACHANDRAN 1993, *Filling In: Why Dennett is Wrong.*

[147] I.e. distinções que existiriam na própria natureza.

mente não excêntricos e que eles afirmam sinceramente que alguma coisa parece vermelha. Que eles digam isso não depende de alguma qualidade interna, mas da sua percepção do objecto, do facto de aperceberem-1 que a coisa é vermelha»[148]. Porque o tratamento da cor faz parte da exposição do modelo de consciência de Dennett que será neste trabalho considerado exemplar (o modelo de CE), será adiada até ao Capítulo 3 uma análise mais completa do tratamento da cor na teoria da consciência, desde já adiantando que esse tratamento faz apelo a considerações evolucionistas acerca da utilidade da visão de cor.

1.4 *A Intenção: pensar e agir voluntariamente.*

A decisão austera de identificar consciência com "apercebimento-1" e de investigar as condições deste a nível sub-pessoal parece eliminar da noção de consciência não apenas *qualia* e imagens como também a ideia de um espaço de trabalho e de iniciação de pensamentos e acções. O problema para Dennett é saber quão voluntários e controlados são os aparentemente criativos movimentos de pensamento, i.e. até que ponto devem raciocínios teóricos (envolvidos por exemplo na resolução de problemas lógicos) e práticos (conducentes por exemplo a decisões morais) ser considerados como *acções* pessoais.

Se raciocinar – entendendo o termo como a aplicação do espírito a um problema específico, que pode demorar muito tempo, ser cansativa, difícil – é alguma coisa que *fazemos* por contraste com experiências que subitamente apercebemos em nós (como quando nos apercebemos de que cheira a fumo), significará isso que somos os controladores directos do nosso raciocínio, e que este consiste em actos propositados de pensamento? A voluntariedade ou actividade aqui em causa não é comparável por exemplo com a crença na existência do mundo exterior. O sentido que interessa aqui é a voluntariedade do trabalho mental de resolução de problemas, o trabalho de ajuízar, ponderar, deliberar, calcular para chegar a algum fim ou solução ou conclusão[149]. Alguma coisa se passa quando raciocinamos. No entanto o que acontece não é caracterizável através da imagem de uma "arena debaixo do olhar mental para a qual vão proposições sobre as quais se age em seguida por meio de operadores lógicos". Existem certamente, passagens, quando são resolvidos problemas teóricos e práticos, mas não passagens directamente agenciadas. Procurar saber o que e como serão essas passagens, saber se elas são involuntárias e inconscientes, é afinal procurar saber se pensamos sem pensar [150].

A solução ryleana para o problema em *The Concept of Mind,* segundo a qual pensar é uma actividade social, é demasiado pouco internalista para Dennett já que desconsidera totalmente o papel causal de cérebros ou outros sistemas físicos. Há no entanto algo de verdadeiro na análise ryleana: a ideia de que pensar é uma actividade pessoal e não um conjunto de acontecimentos neuronais. Isto significa que o pensador não acede às transições e estados sub-pessoais pelos quais pensa: o trabalho de pensamento aparece ao pensador que é o seu autor mais próximo, não surge por *fiat* do autor quando este deduz, induz, julga, abstrai ou conclui. Como diz Ryle, termos como "deduzir" e "concluir" pertencem «não ao vocabulário das biografias mas ao vocabulário das recensões de livros»[151]. Enquanto "biógrafos fenomenólogos" sabemos suficientemente bem que os verbos quase-lógicos das auto-descrições de resolução de problemas (como deduzir, induzir, concluir) não referem actos nossos. Isto quer dizer que afirmar que as pessoas "*chegam* a conclusões" é afirmar que elas dão por si tendo chegado a conclusões, não que estão envolvidas directamente nelas. Assim, não tem sentido perguntar-lhes como fizeram para lá chegar. Já Ryle, em *The Concept of Mind,* pretendera mostrar o quanto a ideia de operações privadas de pensamento discursivo, de preparação agenciada por trás do cenário para que o trabalho intelectual aparecesse na sua forma publicada, fazia parte do mito cartesiano e intelectualista do palco privado. Chegar a uma conclusão num raciocínio é portanto, de alguma maneira, uma coisa que (nos) acontece, perante a qual nos vimos a encontrar. Nas palavras de Ryle, as conclusões de um trabalho de pensamento «pertencem ao estádio de depois de chegada e não a algum dos estádios da viagem para lá»[152].

[148] DENNETT 1969: 146.

[149] «Intelligent expression-using (...) done as work» é a caracterização de Ryle em RYLE 1949: 267. Ryle tem em mente o pensamento como trabalho no qual uma pessoa está empenhada num certo momento, que ocorre por episódios, por oposição aos pensamentos verdadeiros ou falsos daí resultantes. É esse também, aqui, o objecto de Dennett.

[150] Basicamente a resposta terá que ser sim (é o único caminho não homuncular) e esse é o interesse do problema.

[151] RYLE 1949: 269.

[152] RYLE 1949: 280.

Mas Ryle, em *The Concept of Mind*, termina a sua teorização do trabalho de pensamento reportando-o ao discurso didáctico, para si próprio ou para outros (i.e. elimina o suposto palco privado do acontecer agenciado do pensamento mas não quer saber do cérebro). Dennett, no entanto, pretende esclarecer as condições da situação caracterizada por Ryle (o facto de as pessoas não saberem o que fazem quando pensam, e de certo modo chegarem atrasadas às conclusões do seu próprio trabalho intelectual) e para isso utiliza de novo a distinção entre níveis pessoal e sub-pessoal. A primeira constatação que se impõe é que a noção de pensamento enquanto aplicação do espírito a um problema, exactamente como a noção de acção, só tem sentido a nível pessoal. As pessoas pensam e raciocinam mas dos seus cérebros não se pode afirmar o mesmo. Os cérebros devem evidentemente funcionar para que o pensamento aconteça. O problema é que nem esse funcionamento é o pensamento que aqui se procura explicar nem a alternativa reside em considerar que os cérebros são voluntáriamente comandados pelas pessoas que pensam. O pensamento de uma pessoa não começa com a ordem "pense-se" ou "que seja pensado" enviada ao cérebro.

Para compreender melhor a situação, Dennett procura conceber as pessoas enquanto agentes racionais sob o prisma dos agentes racionais artificiais. Toma essa decisão porque, por um lado, processos análogos aos que se passam num computador devem ocorrer numa pessoa quando ela pensa. Claro que quando uma pessoa pretende contar como fez para chegar a uma dada conclusão, verifica que é incapaz de analisar o que se passa ou passou para além de um certo nível de simplicidade. Por exemplo, eu pergunto a mim própria: "Como resolvi a tarefa de raciocínio a?". E respondo: "Somei A+B, dividi..., somei de novo, etc" Mas se continuo a questionar-me "E como somei?" não saberei responder: somei porque sei somar. É um caso clássico de "barragem" wittgensteiniana. Em suma, não temos acesso introspectivo a todos os processos que devem estar a constituir as conclusões que serão objecto de apercebimento-1, embora tenhamos acesso a algumas operações. Há um limite à profundidade do acesso às tarefas cognitivas próprias nos humanos, o que aliás também acontece nos programas de computador. Isso não significa que ao atingir o limite se isolem processos atómicos do acto de pensar, mas apenas que a partir daí a actividade em curso é não-analisável para o (seu) "sujeito", o que não quer dizer que seja não-analisável por princípio.

Dennett propõe que esta situação seja comparada com algo que se passa na simulação artificial da cognição. Certos programas incluem instruções para imprimir uma a uma as operações cumpridas na resolução de problemas, com o propósito de comparar esta publicitação das operações efectuadas, os rastros da máquina (*machine traces),* com os protocolos de sujeitos humanos[153]. O problema é evidentemente saber quais são os pressupostos dessa comparação e que razões se teria para falar da confirmação de uma simulação comparando protocolos humanos com rastros da máquina envolvida na simulação[154]. Imaginando que a tarefa é a "mesma", é frequente nos rastros da máquina aparecerem muitos passos (especialmente passos ligados a processos de tentativa e erro explícitos e exaustivos) que não aparecem nos protocolos dos sujeitos humanos, e que estes podem até negar ter cumprido. Para Dennett, estas dissemelhanças entre os rastros da máquina e os protocolos dos humanos não permitem concluir que os computadores e os humanos usem métodos cognitivos muito diferentes – ao contrário do que um crítico da IA como H.Dreyfus pretende[155] - nem provam por si só a não validade do modelo. Quer no caso dos protocolos humanos quer no caso dos rastos da máquina há limites à forma e à profundidade daquilo que é reportado, e esses limites à profundidade podem ser diferentes. É possível que a capacidade de um humano de publicitar aquilo que apercebe-1 tome como atómicas operações a cujo decurso a máquina tem um

[153] Este ponto gera o artigo, DENNETT 1968, *Machines Traces and Protocol Statements*, escrito em parte contra as críticas de H. Dreyfus à IA, feitas no relatório (RAND memo) "Alchemy and Artificial Inteligence", que esteve na origem do célebre livro de H. Dreyfus, *What Computers Can't Do*. Cf. DENNETT 1995j, em BAUMGARTNER & PAYR 1995, onde Dennett relata a história da sua aproximação à IA e a "encomenda" do artigo de oposição a H. Dreyfus. Dennett pensou que Dreyfus estava errado e respondeu a A.Newell e ao próprio Dreyfus com esse artigo, procurando examinar as pressuposições subjacentes às interpretações dos resultados da simulação computacional de processos cognitivos, como os do GPS de Newell e Simon; «As pessoas da IA ficaram contentes por terem um filósofo do seu lado. Uma relação interessante começou a desenvolver-se e continuou ao longo dos anos» (DENNETT 1996: 59-62).

[154] Para não mencionar a dificuldade de isolar, num humano que resolve problemas, o que vai ser simulado (porque não os "uhms.. ." e "ahms.....", bater com o lápis, enrolar os cabelos, "deixa-me pensar...", por exemplo?). A questão levanta desde logo o problema tayloriano quanto ao que constitui "comportamento de".

[155] As comparações que Dreyfus faz entre humanos e computadores acentuam como características especiais do pensamento humano a intuição, a habilidade de ignorar o irrelevante, a tolerância da ambiguidade, as margens da consciência, a existência de um corpo com um papel na organização e unificação da experiência e nos propósitos e necessidades da acção, as possibilidades de aborrecimento, fadiga e falta de motivação, etc.

acesso mais especificado. No entanto, na máquina também existe um limite, determinado pela linguagem em que ela exterioriza (*prints-out*), e que não é sequer a linguagem-máquina, mas uma linguagem de alto nível. Do facto de um humano não aperceber-1 determinados processos, i.e. não os poder relatar, não se pode concluir que os processos que ele apercebe-1 sejam os processos atómicos da sua tarefa cognitiva. Aliás, a "intuição", a experiência de AHA!, pode ser pensada neste contexto. A possibilidade de experiências de *insight* abrupto nos humanos é usualmente contraposta pelos críticos da IA aos métodos exaustivos, brutos, de computação. No entanto é possível que o apercebimento de uma intuição ao nível pessoal se deva meramente ao facto de ter ficado disponível para publicitação o resultado de um processo com características inacessíveis ao sujeito, o qual poderia até ser um método de busca exaustiva. A intuição corresponderia por hipótese a uma circunstância em que a publicitação é feita numa linguagem muito afastada das operações básicas (seria como o resultado do funcionamento de um programa sem rastros da máquina, cuja única publicitação fosse um comentário à solução do problema com a forma "Veio-me à mente, é tudo").

Em suma, uma pessoa não tem acesso integral ao processamento de informação que lhe permite pensar. Além do mais, num humano a especulação e a racionalização interferem constantemente com o acesso introspectivo (circunstância que os psicólogos introspeccionistas do início do século XX muito lamentaram e que constitui uma diferença importante entre os humanos e as máquinas até agora existentes). Assim sendo, uma comparação entre protocolos humanos e rastros da máquina não pode rasurar o facto de a relação entre a introspecção e o processamento interno de informação num humano ser muito mais complicada do que a relação entre um programa e os "rastros da máquina". Antes de mais, isso acontece porque os protocolos humanos estão muito mais afastados do processamento de informação do que os rastros da máquina. Se por vezes um humano pode dizer quase exactamente porque está a fazer o que está a fazer, por vezes esse acesso não existe de todo, e a acção é exactamente a mesma (o exemplo de Dennett é a acção de descer à cozinha, de noite, e mordiscar uma maçã; a pessoa que faz isso pode ou não ser capaz de apresentar um relato elaborado acerca de razões e a acção e a informação processada terão sido aproximadamente as mesmas).

O processamento de informação que deve existir – caso contrário o controlo do comportamento seria inexplicável – não acontece ao nível pessoal e não tem de modo algum que ser consciente. Mesmo o controlo de rotinas rígidas (por exemplo de comportamentos animais) envolve processamento de informação. O exemplo dado por D. Dennett é o de um pássaro que "finge" ter uma asa partida para afastar predadores do ninho onde estão os filhotes. Um tal comportamento de logro (aparentemente inteligente mas de facto instintivo) é descritível de forma mentalista fazendo apelo a uma razão para o comportamento em causa. Isso não significa no entanto que existam processos mentais conscientes no pássaro que tenham como conteúdo explícito essa razão, tal como ela é linguísticamente e humanamente formulável. Aquilo que pode vir a ser descrito de forma mentalista como razão – controlos de comportamentos com finalidade – pode existir a nível subpessoal sem qualquer apercebimento consciente do agente. Do mesmo modo e no que respeita a controlos do comportamento de resolução de problemas, os protocolos humanos não fornecem necessariamente pistas válidas acerca do processamento de informação sub-pessoal. Não é portanto razoável supor que se prova ou desmente modelos de processamento de informação da cognição humana comparando estritamente os protocolos humanos com os rastros da máquina.

Além disso, Dennett defende que as operações cognitivas que controlam comportamento de sistemas como os humanos não são as operações lógicas dos lógicos, ao contrário do que pode passar-se com a programação de máquinas, em que o *designer* pode e deve assegurar-se à partida de quais operações farão parte do programa. É notório que os humanos podem pensar mal e ilogicamente, e qualquer teoria adequada do raciocínio deve enquadrar esse facto. As simulações artificiais da cognição humana visam portanto um nível alto, o nível da resolução de problemas, não existindo uma correspondência entre as operações lógicas da máquina nas quais a simulação se baseia e os acontecimentos no cérebro. A capacidade que os humanos têm de seguir aproximadamente as leis lógicas em raciocínios não tem a ver com o facto de os cérebros serem máquinas lógicas que executam sem falha operações (os cérebros não são máquinas lógicas semelhantes, ao contrário dos computadores) mas sim com o facto de que «parte da maneira como as coisas são é a maneira como as coisas logicamente são, e se o nosso

comportamento deve ser apropriado à maneira como as coisas são, ele tem que ser produzido de acordo com linhas logicamente sólidas»[156].

O objectivo final das análises do comportamento de resolução de problemas é mostrar que a explicitação dos raciocínios das pessoas, feita a partir de fora e com linguagem, pelo teórico, não tem por que corresponder a operações e eventos no cérebro e não tem evidentemente também que corresponder a um percurso de operações conscientes. A explicitação dos passos do pensamento, de qualquer pensamento, não é uma biografia de eventos mentais conscientes nem uma descrição de eventos cerebrais e sim uma caracterização intencional. A conclusão é ryleana em espírito mas é também guiada pela teoria da intenção apresentada por Elizabeth Anscombe em *Intention*[157].

1.4.1 *O nível pessoal e a teoria da acção: as razões na acção. A intenção segundo Anscombe e a sua adaptação à psicologia filosófica de Dennett.*

O raciocínio não é apenas necessário para resolver equações, resolver problemas ou construir argumentações jurídicas. Ele é também essencial para guiar e determinar outro tipo de acções, mais directamente ligadas aos movimentos dos corpos no mundo. Assim, «ao relatarmos o nosso raciocínio, nem sempre estamos a dizer como chegamos a uma certa solução ou conclusão, (...) frequentemente estamos a dizer porque é que decidimos fazer o que quer que estejamos a fazer»[158]. Podemos precisamente afirmar que existe uma acção intencional quando alguém é capaz de dizer porque fez aquilo que fez, i.e. quando alguém é capaz de dar a

[156] DENNETT 1969: 155. Aqui Dennett reporta-se a TURING 1950. Ruth Millikan (cf. MILLIKAN 1984) retoma esta sugestão.

[157] Em *Intention* (ANSCOMBE 1957), E. Anscombe investiga os critérios que presidem à classificação de acções como intencionais. Uma acção é intencional de acordo com Anscombe se à questão "Porquê?" o agente responde com uma razão para a acção. "Eu não me apercebi de que fazia isso" constitui uma rejeição da questão "Porquê?". Dennett discorda da conclusão geral da teoria da acção de E. Anscombe, segundo a qual, *não sendo fisicamente determináveis quaisquer características* em virtude das quais um particular movimento seria uma acção intencional, existiria um abismo intransponível entre a intencionalidade das acções e a causalidade do mundo físico. Dennett considera esta conclusão segundo a qual as acções estariam divorciadas do mundo das causas físicas dualista e não fundamentada (cf. DENNETT 1968: 238). No entanto, crê que E. Anscombe oferece precisamente uma valiosa descrição dos traços ou critérios que permitem chamar a uma acção intencional, a qual, como à frente se verá, se adequa a modelos de processamento de informação do cérebro.

[158] DENNETT 1969: 156.

razão de determinados movimentos, por oposição ao caso em que algo simplesmente acontece. Supostamente é pelo facto de a razão dada pelo agente ser a razão para a acção que a acção deve ser considerada intencional. A questão da acção não é de modo algum uma curiosidade marginal da teoria da mente. De acordo com Dennett, «A prática de pedir e dar as razões das acções é central nas nossas noções de acção e responsabilidade e, de facto, na nossa noção de pessoa»[159]. Esclarecer a natureza do raciocínio prático e a existência de acções ao nível pessoal é, assim, esclarecer a própria existência de "pessoas". Só a existência de acções por contraste com meros movimentos assegura o vértice da teoria da mente, erguido sobre a intencionalidade e a consciência, que é a pessoalidade.

A existência de acção supõe aparentemente a existência de razões para movimentos de entidade físicas e a intenção de executá-los. Quando alguém faz alguma coisa por alguma razão esse agente tem, utilizando a terminologia de D. Davidson, uma atitude-pró[160] em relação a acções de um certo tipo e acredita que a sua acção é desse tipo. Caracterizar a autoridade do agente relativamente às razões para a acção é no entanto uma tarefa complicada. Em primeiro lugar, pedir a um agente que dê as razões da sua acção não é pedir-lhe que reporte os seus pensamentos. Frequentemente as razões para um comportamento – entendendo aqui por comportamento movimentos aparentemente voluntários de um corpo – não estarão sequer entre esses pensamentos (para não mencionar o facto de esse reportar ser confiável apenas supondo que a memória está a ser confiável e que não está a existir racionalização). O reportar normal que por exemplo as pessoas fazem das suas razões para a acção é duplamente inferencial, envolvendo não apenas a interpretação do recordado como a também racionalização, i.e. atribuição a posteriori de razões que não guiaram de facto a acção. Para além disso, uma pessoa pode ter conhecimento das causas do seu próprio comportamento, sem que a este comportamento se possa chamar com propriedade uma acção da pessoa. Em *Intention,* Anscombe dá um exemplo de conhecimento de um comportamento próprio cujos motivos são dados pela pessoa mas que não

[159] DENNETT 1969: 156.

[160] Esta é uma expressão geral que se aplica a desejos, princípios estéticos e morais, finalidades, etc. Cf. DAVIDSON [1963]: 3-4.

constitui uma acção dessa pessoa: é o caso em que alguém repentinamente salta para trás, e quando se lhe pergunta porque o fez, responde que foi o grito e a sacudidela de um crocodilo que o fizeram fazer isso[161]. É certo que nenhuma outra pessoa dispõe do acesso privilegiado que permite a essa pessoa fazer o relato. No entanto não é esse o tipo de conhecimento – o conhecimento das causas de movimentos que partem da própria pessoa mas que não são intencionados por ela – que interessa o teórico da acção. Esse conhecimento não interessa por um lado porque não é o conhecimento de um acto guiado por uma razão, por outro lado porque é duvidoso que não exista inferência na observação emitida pela pessoa. A pessoa possivelmente não apenas recorda os conteúdos apercebidos como também interpreta. O problema da acção propriamente dito diz respeito às condições em que *agentes são capazes de ter algum conhecimento não inferencial e não observacional das sua razões para a acção, razões essas que devem reger a acção "em acto"*, e não ser reconstruídas a posteriori. O que o caso de Anscombe mostra, segundo Dennett, é que «Não tenho acesso, privado ou outro, aos meus processos cerebrais, mas apenas ao meu apercebimento e à sucessão de mensagens que lá chegam; não tendo outra explicação do salto, e tendo visto outros saltarem quando se encontraram perante visões súbitas e estranhas, eu infiro (consciente ou subconscientemente) que a visão espantosa causou o salto»[162]. No entanto, num caso como este, embora o agente conheça sem observação o seu próprio movimento (e este é o traço que, como se verá, caracteriza aquilo que é voluntário) um neurologista poderia vir a dizer melhor do que a própria pessoa o que causou o seu comportamento. Assim, para que algo constitua uma razão para a acção não basta que seja causa de um comportamento próprio do agente.

O problema da acção não envolve apenas o problema geral que consiste em especificar o que distingue acções de meros acontecimentos e mesmo de movimentos próprios, envolve também o problema de saber se razões e causas se excluem mutuamente ou não. Numa primeira alternativa a existência de razões para as acções está de algum modo separada das questões da causalidade (é possível aliás alegar que parte do que se quer dizer quando se diz que uma acção foi intencional é que ela não foi causada, que nada ou ninguém manipulou o agente). Numa segunda alternativa as razões podem ser causas e servir assim para explicar acções[163].

[161] ANSCOMBE 1957:15.
[162] DENNETT 1969: 161.

Como wittgensteiniana, Anscombe defende em *Intention* que ao falar de razões como causas das acções intencionais se incorre numa falácia naturalista. A relação entre razões e acções é uma relação interna, conceptual, descritiva. A intencionalidade de uma acção está segundo Anscombe ligada a uma descrição intencional e

[163] O lugar clássico do tratamento do problema das razões e das causas na acção era, quando Dennett escreve C&C, o artigo *Actions, Reasons and Causes* (DAVIDSON 1963). É importante para o esclarecimento da posição defendida em DENNETT 1969 compreender a oposição de Davidson aos wittgensteinianos que nesse artigo se desenha. Os wittgensteinianos, como Anscombe, defendem que tomar razões como causas de acções constitui uma falácia naturalista (as alegadas falácias naturalistas consistem em identificar conceitos éticos com conceitos naturais). Ao contrário, contra nomeadamente E. Anscombe e também contra G. Ryle, D. Davidson defende no artigo que as razões podem ser causas. Aqueles que defendem que é um erro categorial e uma falácia naturalista considerar que razões podem ser causas alegam que os vocabulários das razões e das causas são totalmente diferentes e que existe um abismo intransponível entre as explicações causais naturalistas e as descrições interpretativas através de razões. Pode-se chamar ao argumento dos wittgensteinianos um argumento da conexão lógica (cf. ENGEL 1998: 32). Esse argumento é o seguinte: se as razões fossem causas de acções, então a causa suposta (a razão) deveria ser separável do seu efeito (a acção ou comportamento) que seria extrínseco e contingente. Ora, não é possível descrever a razão de uma acção independentemente da própria acção e do seu resultado (a intenção de fazer A menciona A). Logo existe uma relação intrínseca e necessária entre a razão e a acção, que não tem as características de uma relação causal. Em *Actions, Reasons and Causes*, Davidson ataca este argumento. Do facto de as descrições das razões estarem conceptualmente ligadas às descrições da acção não se segue que os eventos que as descrições descrevem não entrem numa relação causal extrínseca. O wittgensteiniano sustenta que, sendo toda a explicação causal por princípio, nomológica, se as razões fossem causas deveria haver leis que ligasssem as descrições mentais das razões às descrições das póptias acções. Ora, não existem essas leis psicofísicas. Davidson concorda que não existem leis psicofísicas. Considera no entanto que da inexistência de leis psicofísicas não se segue que as explicações psicológicas comuns não sejam causais. A posição de Davidson resulta do facto de ele defender também que as únicas leis estritas (i.e. sem excepção) são as leis físicas e que portanto as generalizações causais que sustentam os laços entre razões e acções dependem das características não intencionais dos eventos. Os eventos descritos como mentais (como as razões) são também eventos, e como tal submetidos a leis causais. Como diz D. Davidson (DAVIDSON 1963: 19): «Why on earth should a cause turn an action into a mere happening and a person into a helpless victim? Is it because we tend to assume, at least in the arena of action, that a cause demands a causer, agency an agent? So we press the question: if my action is caused, what caused it? If I did, then there is the absurdity of infinit regress; if I did not, I am victim. But of course the alternatives are not exhaustive. Some causes have no agents. Among these agentless causes are the states and changes of state in persons which, because they are reasons as well as causes, constitute certain events free and intentional actions». A razão para considerar como causais as explicações através de razões está dada. Evidentemente toda a dificuldade reside em conceber qual é exactamente o poder causal das explicações psicológicas através de razões e a questão não fica resolvida por aqui. Para um confronto mais recente entre a teoria davidsoniana das razões e das causas e a psicologia de senso comum (*folk-psychology*) cf. ZILHÃO 1998/1999. O autor conclui que uma explicação filosófica como a que Davidson oferece da explicação comum de acções através de causas constitui uma teoria inapropriada da acção humana.

não a um factor causal isolável (a uma razão que seria a sua causa). Ora é precisamente a enumeração dos componentes da descrição intencional que dá conta da intencionalidade da acção que Dennett vai buscar a Anscombe. Mas apesar de concordar com a análise que Anscombe faz dos componentes das acções intencionais, Dennett defende, ao contrário de Anscombe e da mesma forma que Davidson, que é possível considerar simultaneamente causação e razões para a acção. Quando alguém faz alguma coisa por uma razão não apenas aquilo que faz é causado, como aliás «o que faz (de) uma razão a (...) real razão para fazer alguma coisa é que os eventos de processamento de informação que causam o que (é feito) têm entre eles um evento com o conteúdo da (...) razão, quer (se) aperceba-1 ou não esse conteúdo»[164]. Assim, Dennett adaptará a caracterização de Anscombe à sua teoria do apercebimento-1, considerando que essa caracterização diz respeito ao nível pessoal.

Em *Intention* Anscombe chega à seguinte caracterização das acções intencionais:

(1) Não é possível falar da intencionalidade de movimentos corporais *tout court*: *apenas movimentos corporais sob uma particular descrição podem ser ou não ser intencionais*. Ora as descrições dos mesmos movimentos corporais podem ser várias[165]. Assim, sob uma determinada descrição um conjunto de movimentos corporais pode ser uma acção intencional mas não sob outra descrição. Por exemplo, uma pessoa está a comer carne ao jantar e sabe que está a comer carne. Ela está também a comer a carne do seu próprio animal de estimação, mas não sabe que está a comer o seu animal de estimação. Neste segundo caso a acção não é intencional. O agente ou actor deve aperceber os movimentos sob uma particular descrição para que essa acção seja intencional, o que aqui não acontece. Uma acção não é intencional se o suposto agente apenas chega mais

[164] DENNETT 1969: 163.

[165] Um exemplo importante em *Intention* (ANSCOMBE 1957: 37) é o exemplo de um homem que move o braço para bombear a água para encher a cisterna de um edifício. Alguém arranjou maneira de contaminar sistematicamente essa água de modo a envenenar de forma progressiva e impossível de notar imediatamente os habitantes do edifício. Estes são dirigentes políticos, envolvidos em crimes, cuja eliminação trará a libertação do país. O homem está portanto não apenas a mover o braço e a encher a cisterna mas a envenenar os habitantes do edifício, a matá-los e a libertar o país. O problema é saber o que é que ele está a fazer *intencionalmente quando o seu braço se* move para cima e para baixo, mesmo tendo sido informado do envenenamento da água por quem o levou a cabo e das finalidades deste.

tarde ao seu próprio comportamento e observa ou vem a tomar conhecimento de que estava fazer alguma coisa[166].

(2) *Não é suficiente que exista apercebimento de comportamentos próprios sob uma descrição mentalista para que a acção seja intencional: o apercebimento tem que ser apercebimento sem observação*. Não basta que eu me observe visualmente fazendo x (por exemplo desenhando enquanto ouço uma conferência) para que aquilo que eu me observo fazer seja uma acção minha (embora eu aperceba tais movimentos sob uma descrição mentalista). Se foi apenas observando que eu tive conhecimento do que estava a fazer, aquilo que eu estava a fazer não era intencional. Para que movimentos sob uma descrição constituam uma acção minha, eles têm que ser conhecidos sem observação. Adaptando esta descrição à sua própria teoria Dennett afirmará que apenas existe acção pessoal por oposição a comportamento próprio neurofisiologicamente voluntário se «um sinal com conteúdo descritivo dessa acção atravessa a linha do apercebimento e esse sinal não é nem um sinal proprioceptivo dos músculos ou articulações nem um sinal provindo mediadamente dos orgãos sensoriais»[167]. Apenas sinais eferentes de alto nível (por oposição a sinais motores próximos da periferia) podem estar nestas condições, podendo ser descritos como tendo como conteúdo ordens para que acções se efectuem: «Por exemplo "abre a porta!" poderia ser o conteúdo de um estado ou eventos de nível relativamente alto que controlasse um número de diferentes sub-rotinas, todas desenhadas para que a porta viesse a ser aberta de uma maneira ou outra»[168]. Se esse sinal atravessasse a linha do apercebimento, o sistema aperceberia-1 o seu conteúdo, i.e. aperceber-se-ia do que se passa (os diferentes movimentos físicos) sob uma particular descrição. A diferença chave para a caracterização deste apercebimento é

[166] Outro exemplo, dado por D. Dennett (DENNETT 1969: 165) é o seguinte: se eu estou a bater nas teclas do computador ao ritmo de *Rule Britannia* enquanto escrevo, e alguém me diz que estou a fazer isso – não a escrever, pois eu sei que o estou a fazer é escrever, logo estou a escrever intencionalmente, mas a bater nas teclas com aquele ritmo – e eu exclamo "Ah! Estou a fazer isso!", o que eu estava a fazer não era intencional. Imagine-se ainda o movimento de um braço de um humano que enterra um objecto pontiagudo no peito de outro humano. O que existe é um movimento, uma sequência de eventos, por hipótese físicos, que só será uma acção mediante o apercebimento consciente de alguém. Se não existir apercebimento e apercebimento sob uma particular descrição não se tratará de uma acção, como será o caso se a pessoa esteve por exemplo hipnotizada e vem a si no exacto momento em que acabou de matar outra pessoa.

[167] DENNETT 1969: 163.

[168] DENNETT 1969: 166.

a fonte eferente dos eventos descritos: o conhecimento sem observação de Anscombe é, na linguagem de Dennett, *apercebimento-1 de controlos eferentes*. Se eu ajo intencionalmente, não será por vias aferentes – por exemplo vendo o meu corpo fazer movimentos – que eu me apercebo de que estou a tentar abrir a porta.

Anscombe chama também a este conhecimento não observacional conhecimento prático[169]. O conhecimento prático é comparado com o conhecimento que um homem com uma imaginação perfeita tem do projecto de construção de um edifício, projecto esse que dirige dando ordens concebidas até ao último pormenor mas sem poder ver o edifício. Poder-se-á objectar que esse conhecimento será correcto se e só se as ordens forem levadas a cabo. Caso contrário, existirá erro no conhecimento. No entanto segundo Anscombe, neste caso «o erro [estará] na performance e não no juízo»[170]. Sendo este conhecimento não observacional prático um conhecimento de ordens (comandos eferentes na linguagem de Dennett) existe a possibilidade de movimentos idênticos poderem ser controlados por comandos eferentes diversos. Nesse caso, como Anscombe viu, a pessoa, e só ela, poderá perceber a diferença. O caso que Anscombe apresenta é o seguinte «digo a mim próprio "Agora pressiona o botão A"– enquanto pressiono o botão B – isto é algo que pode certamente acontecer (...) e aí o erro é de performance e não de juízo»[171]. De novo Anscombe pretende fazer ver que o erro não está no que foi dito e sim no que foi feito, e que a partir de dentro e sem observação o agente apercebe esse erro. Ele seria capaz de fazê-lo, diz Dennett, não por reconhecer diferenças qualitativas nos comandos eferentes subpessoais, mas pela sua disposição para afirmar coisas diferentes. Evidentemente, toda a dificuldade da auto-análise do conhecimento prático não observacional deriva do facto de este nunca existir isolado no comportamento próprio. O conhecimento prático não observacional do comportamento próprio está sempre intimamente ligado à propriocepção e mesmo à percepção visual do

[169] Objectar a uma concepção intelectualista do conhecimento prático é aliás uma das finalidades de *Intention*. Cf. ANSCOMBE 1957: 82 e seguintes. Segundo Anscombe as suas análises confirmam a teoria do conhecimento prático de Tomás de Aquino: «Practical knowledge is the "cause of what it understands" unlike speculative knowledge, which is "derived from the objects known"» (ANSCOMBE 1957: 87).

[170] ANSCOMBE 1957: 82.

[171] ANSCOMBE 1957: 57.

que se está a fazer[172]. A informação quanto ao que se está a passar comigo tem a cada instante proveniências diversas: a propriocepção e a percepção pelas vias aferentes normais juntam-se ao conhecimento prático dos comandos eferentes. Daí, aliás, a importância do apercebimento-1 na integração e por consequência no controlo do comportamento global. Anscombe aponta ainda uma terceira característica das acções intencionais.

(3) *Uma acção será intencional apenas se o agente puder oferecer a razão para a acção.* Caso contrário, a acção pode mesmo ser voluntária no sentido neurofisiológico, o agente pode ter conhecimento prático dessa acção, mas ela não será intencional.

Esta análise apoia a pretensão de Anscombe segundo a qual uma acção não é considerada intencional em virtude de algum característica adicional e isolável que estaria presente quando a acção é levado a cabo. Essa conclusão é justificada da seguinte maneira: (1) uma caracterização intencional da acção não pode ser dada sem a descrição pela qual a acção é intencional (a mesma acção pode ser intencional debaixo de uma determinada descrição e não intencional debaixo de outra descrição), (2) nada na pessoa considerada por si própria pode determinar o conteúdo dessas descrições. Como se viu, Dennett não aceita o ponto 2, oferecendo precisamente uma hipótese acerca do que na pessoa considerada por si própria poderia determinar o conteúdo dessas descrições, nomeadamente o apercebimento-1 do controlo do comportamento próprio. Existem comandos eferentes que são ao mesmo tempo causas de movimentos e suceptíveis de descrições intencionais que os caracterizam como acções regidas por razões. Evidentemente, a actividade de dar razões não é de modo algum uma actividade na qual os agentes sejam incorrigíveis. Existe, além disso, por exemplo no caso dos humanos, um fundo do comportamento do qual não faz sentido pensar que se podem dar razões, embora existam razões, na medida em que se trata aí dos fins do tipo de ser biológico que somos. No entanto e na medida em que controlos eferentes podem gerar movimentos susceptíveis de muitas descrições diferentes, o que se conclui é que a acção intencional ou voluntariedade só pode estar ligada à especificidade dos conteúdos atribuíveis aos comandos eferentes intencionalmente apercebidos pelo "ac-

[172] Pense-se num esquema complicado e detalhado que se está a fazer numa folha de papel. Nunca se conseguiria continuá-lo sem observação do que foi sendo feito.

tor" (e não a objectos a mais chamados volições, outro dos disfarces do homúnculo). É este apercebimento que corresponde à ideia geral de Anscombe sobre descrições intencionais. O silogismo prático aristotélico é o exemplo do estatuto destas descrições: «Se a descrição aristotélica do silogismo prático fosse suposta reportar processos mentais actuais, ela seria em geral bastante absurda. O interesse da descrição é que ela relata uma ordem que está lá quando quer que as acções sejam feitas com intenção»[173].

A descrição da ordem intencional feita por Anscombe e adaptada à teoria do apercebimento-1 não permite, é claro, considerar a acção regida por razões como algo que começa *ex nihilo* num centro irradiante. Mesmo admitindo em C&C, a linha do apercebimento e falando de comandos eferentes, Dennett afirma que essa linha não dá passagem para nenhum centro de controlo pessoal. A iniciação de actividades pelo cérebro, que determinará o comportamento, não tem que ser acompanhada necessariamente por um pensar para si próprio do agente. Os fenómenos do querer intencional e racional estão ligados portanto por um lado a comandos eferentes que são causas de movimentos próprios voluntários e por outro – ou melhor, ao mesmo tempo – a descrições mentalistas dos movimentos próprios voluntários envolvendo razões. O que faz a diferença das acções intencionais é o apercebimento-1 do conteúdo dos comandos. O apercebimento-1 não é a origem das intenções ou da iniciação do controlo: acções intencionais são apenas movimentos sob particulares descrições dos quais o agente tem conhecimento prático e dos quais ele está preparado para dar razões. Em suma, quando se fala de razões para a acção e de razões na acção trata-se mais propriamente de assunção e integração linguística – possível através da auto-estimulação e da recuperação de informação – do que de iniciação *ex nihilo*. O apercebimento-1 não é a origem das intenções ou da iniciação do controlo.

O problema é que parece, assim, que o que acontece nas acções intencionais é que o agente vai "atrás" do que já está "decidido",

[173] Cf. ANSCOMBE 1957: 80. Anscombe faz notar como seria absurdo considerar que uma pessoa pensa para si própria, enquanto age, de forma totalmente explícita (no exemplo de Aristóteles) "Os alimentos secos convêm aos seres humanos/ Este alimento é seco/ Eu sou um ser humano/ Esta é uma porção desse alimento/ Este alimento convem-me" (ANSCOMBE 1957: 58). Pensar assim é fazer como se o silogismo prático fosse uma demonstração, dadas as premissas, de uma conclusão, conclusão essa que seria uma acção. De resto, como nota Anscombe, no exemplo de Aristóteles nem sequer parece dever seguir-se alguma acção de todo o processo.

i.e. iniciado no seu cérebro ao nível sub-pessoal. Embora Dennett não considere esta a melhor descrição da situação, o que é certo é que sem volições nem centro de controlo não existe uma distinção absoluta entre movimentos (descritos como) intencionais e movimentos (descritos como) não intencionais. Aliás Dennett não aceita que o caso elaborado por Anscombe em *Intention* prove que a noção de acção deve ser divorciada do mundo das causas físicas: se fosse assim «o meu braço levantar-se-ia e o meu dedo premiria o gatilho como parte de um universo determinista mas eu reteria a livre escolha de esta ser uma acção minha (...) ou meramente algo que me aconteceu»[174]. Não existindo distinção entre o intencional e o não intencional do ponto de vista da determinação causal, resta apenas uma distinção de graus de apercebimento. Na prática isso fará com que a distinção verdadeiramente importante para a ética seja uma distinção quanto ao tipo e à sofisticação do apercebimento, sofisticação essa relacionada por Dennett com a importância que a estimulação verbal tem no controlo do comportamento próprio.

A linguagem tem então, pelo menos no caso humano, um papel central no controlo fino e sofisticado do comportamento próprio. Evidentemente, o que se procura controlar verbalmente deve ser controlável (um comportamento como tremer, por exemplo, não o é). Para decidir perante que caso se está, vale o critério que precisamente Anscombe propõe em *Intention*: alguma coisa fica estabelecida como razão susceptível de controlar o comportamento se se pode objectar contra ela. Objectar contra as razões que controlam o comportamento de uma pessoa é argumentar com ela. Nos termos de Dennett, argumentar com uma entidade será precisamente tratá-la como uma pessoa, um agente racional e auto-consciente, o qual é, por definição, uma entidade que de algum modo é movida a razões. A responsabilidade por uma acção é assim, como as razões para a acção intencional, uma questão de nível pessoal e não uma questão sub-pessoal relativa a antecedentes causais. As questões éticas da liberdade e responsabilidade na acção estarão relacionada com a possibilidade de controlo do comportamento por meios racionais, pelo próprio ou por outrém. A aplicação à ética destes pontos de teoria da mente será explorada no Capítulo 4.

[174] DENNETT 1968: 242.

1.5 *A linguagem, o entendimento e o nível pessoal. O que fica estabelecido em* Content and Consciousness.

Uma vez que o mental tal como é caracterizado em C&C não se restringe aos humanos, resta neste momento clarificar o estatuto de um tipo especial de fenómeno, o conhecimento linguísticamente enunciado, que é, esse sim, restrito a um tipo especial de mentes. É esse o teor do último capítulo de C&C, *Language and Understanding.*

A TSI tem como princípios que (1) a adscrição interpretativa de conteúdo a sistemas físicos é neutra em relação à linguagem e que (2) a adscrição interpretativa, linguística, de conteúdo a sistemas cognitivos não tem, nem deve pretender ter, precisão. A impossibilidade de individuar conteúdos de forma precisa transpõe-se inclusivamente para o nível das crenças e conhecimentos explicitáveis linguisticamente pelo próprio sistema intencional. O conhecimento de uma pessoa não consiste em (1) uma lista dos factos conhecidos por essa pessoa, especificados com toda a precisão, nem em (2) crenças necessariamente verdadeiras, explicitamente justificadas por essa pessoa para si própria. Aliás, dado o que foi dito anteriormente, é claro que uma pessoa não tem acesso introspectivo à diferença entre "crença" e "conhecimento" (considerado como crença verdadeira justificada). Quando alguém afirma taxativamente que conhece algo que é (a partir de fora e reconhecidamente) falso, isso não poderia dever-se a uma má identificação do seu estado interior, que é impossível. É neste ponto que se revela a diferença de interesses entre o filósofo da psicologia e o epistemólogo. O centro do interesse do epistemólogo são as crenças verdadeiras justificadas da melhor maneira possível (na comunidade de investigação e na espécie em geral e ao longo do tempo, poder-se-ia dizer). Pelo contrário para o psicólogo e o filosófo da psicologia, é óbvio que "ser tomado por verdadeiro" não significa, ao nível das crenças de animais ou de indivíduos humanos, "não ser sustentado senão após justificação explícita ou testes rigorosos". De facto, num indivíduo, qualquer crença pode constituir-se em razão para a acção, não apenas as "crenças verdadeiras". Estabelecida a diferença, e excluído do âmbito da filosofia da psicologia o estudo normativo das crenças, o problema para o filósofo da psicologia consistirá em saber se num indivíduo a informação armazenada e a sua utilização inteligente pode ou não ser objecto de uma caracterização precisa. Para Dennett, o conhecimento distingue-se do mero armazenamento de informação na me-

dida em que envolve entendimento, i.e. capacidade de utilização inteligente da informação para controlar e guiar a acção. Este entendimento é na terminologia ryleana mais um *know how* do que um *know that*, envolve seres activos no mundo, com capacidades comportamentais para além do facto de serem depósitos de informação. O entendimento assim caracterizado estaria, segundo Dennett, de momento ausente em máquinas. Um teste da posse de entendimento por um sistema não pode, assim, visar apenas informação armazenada. Deve visar ainda aquilo que "cerca" essa informação e aquilo que cerca o sistema. Saber o que está envolvido no entendimento para além de recuperação de informação, saber se o entendimento é discretizável exaustivamente em parcelas, se admite ou não graus, é parte da tarefa de todos os teóricos da cognição.

Dennett avança algumas propostas de análise nessa linhas, nomeadamente: (1) não pode haver um sistema que tenha apenas o entendimento isolado de um facto, tal como "Jones está aqui". Um facto que não pode ser usado não pode ser compreendido. Crenças são algo que não pode existir singularmente, isoladamente. Mas se, por exemplo no caso humano, não pode existir uma pessoa que entenda apenas uma frase, não é também possível marcar o limite exacto entre o estado da pessoa que não sabe alemão e que foi colocada numa esquina a responder "Der Banhof ist links um die Ecke" quando ouve "Wo ist der Banhof?" e a pessoa que sabe alemão. Se grande parte do conhecimento de uma pessoa é habilidade, *know how,* é no entanto impossível estabelecer a priori limites quanto à informação que tem necessariamente que estar envolvida na compreensão de uma simples frase em língua natural. O entendimento de informação particular depende do que o sistema pode fazer com esta, o que por sua vez depende de outra informação de que o sistema dispõe. De qualquer modo, a conclusão geral é que a informação não é contabilizável de forma discreta. Não é também possível considerar a informação como uma quantidade abstracta em relação directa com estados de coisas no mundo, pois esse posicionamento esquece o mediador, o entendedor, assim como a sua limitação (estados de coisas são captados, captados como importantes, relevantes, salientes, para particulares entendedores que são entendedores de um dado tipo). Alguma coisa só constitui informação, quando tem *funções* de ordenação num contexto, e esse contexto é determinado, num sistema intencional, pelo seu comportamento. Mesmo na teoria da informação, quando se mede a quanti-

dade de informação, esta é determinada pela diminuição do grau de incerteza no receptor. Sobretudo, previsivelmente na TSI, se há informação num sistema A, que procura transmiti-la a B, de modo a fazê-lo armazenar a informação, não se pode daí concluir a existência de veículos de informação com um conteúdo idêntico – um particular fragmento de informação — em A e em B, e passando de A para B, nomeadamente porque o que exactamente é transmitido depende do contexto em que é recebido.

Todas estas considerações se combinam com a ideia central da TSI, segundo a qual o conteúdo é função da capacidade de controlar actividade eferente. Dennett conclui que «o facto de que a capacidade da linguagem de ser portadora de informação depende dos efeitos da linguagem numa pessoa, no que eu chamei um sistema intencional, é o mesmo facto, ao nível pessoal, que o facto, ao nível sub-pessoal, de o centralista não ser capaz de adscrever conteúdo preciso aos eventos e estados de um tal sistema»[175]. Não pode haver precisão e individuação na adscrição de conteúdo a sistemas devido ao holismo da adscrição e ao carácter teleológico dos sistemas aos quais é atribuído conteúdo, o qual se traduz no facto de estes usarem informação para fins, sendo a "identidade" desta determinada em função desse uso.

É com estas observações sintomáticas do interpretativismo fundamental da teoria que Dennett termina C&C. Como se verá, o interpretativismo é a razão pela qual Dennett não apenas não aceita a teoria da identidade, como também não aceita o funcionalismo puro, que identifica estados mentais com estados lógicos ou funcionais salientes e individuáveis. Esse realismo intencional será a opção de J. Fodor, na linha directa de H. Putnam. A TSI é de Dennett é pelo contrário, em relação ao funcionalismo simbólico ortodoxo proposto programaticamente por Putnam e desenvolvido por Fodor, um funcionalismo enfraquecido. No próximo capítulo far-se-á uma análise comparativa das duas grandes alternativas quanto à natureza da intencionalidade que são o interpretativismo e o realismo intencional, exponenciado pela obra de Jerry Fodor, de modo a avaliar a sustentabilidade da teoria da mente proposta por Dennett. Esta, como se viu, consiste num centralismo que aparentemente é anti-behaviorista mas que de facto se mantém behaviorista em duas frentes: numa frente interna (cerebral) e numa frente

[175] DENNETT 1969: 189.

fenomenológica (entendendo por fenomenologia descrições como as de Ryle e Wittgenstein).

J. Fodor afirma frequentemente que um behaviorista lógico não pode oferecer uma teoria dos processos da inteligência: é de supor que a teoria "relacional" da mentalidade proposta por Dennett sofra de idêntica deficiência. Por seu lado T.Nagel, perante a teoria da mente apresentada em C&C, afirmou que este livro mostra que «É possível dizer muita coisa verdadeira sobre a mente e a sua relação com o corpo, ao mesmo tempo que se deixa o problema mente-corpo fundamentalmente intocado»[176]. Sabendo que para Nagel é a consciência fenomenal que torna o problema mente-corpo verdadeiramente importante, «intratável e interessante»[177] é compreensível que a teoria da mente apresentada em C&C lhe pareça insatisfatória: nenhuma das duas noções em que a teoria da mente culmina, o apercebimento-1 e o apercebimento-2 é uma noção fenomenal. Nagel comentará aliás, numa futura recensão ao livro maior de Dennett sobre a consciência, *Consciousness Explained* que «é notável como a visão de Dennett mudou pouco», continuando os seus livros a apresentar teorias empiricamente ricas e estimulantes sobre a mente mas que deixam o problema mente-corpo intocado. Nagel prevê então que «o desacordo entre nós acabará no túmulo, se acabar aí».[178]

[176] NAGEL 1995b: 85.

[177] Cf. NAGEL [1974]: 165, para a célebre abertura do artigo *What is it like to be a bat?* (NAGEL 1974).

[178] NAGEL 1995b: 87. De facto, como nota Rorty (RORTY 1993:188) «Nagel and Dennett beg all the questions against each other». De resto, a disputa é em si própria esclarecedora, uma vez que gira em torno da noção de "intrínseco" na teoria da mente. Para Dennett não existe "intrinsecamente" mentalidade (i.e. toda a mentalidade é relacional) enquanto que para Nagel o mental "é intrinsecamente" (consciente).

Segunda Parte: O Modelo

Capítulo 2

A posteridade do funcionalismo de Putnam: diferendos acerca da natureza da psicologia.

«These properties — call them stance properties – exist only by grace of our talking about them. Shades of Berkeley! In the intentional stance we play God to objects in the quad!», Fred Dretske[179]

2.1 *Dos anos 70 aos anos 90: A teoria teleológica do conteúdo, as suas implicações e os seus opositores* – Brainstorms *(1978)*, The Intentional Stance *(1987)*, Brainchildren *(1998)*.

De acordo com a ordenação defendida pelo próprio Dennett tratar-se-á neste capítulo da teoria do conteúdo desenvolvida ao longo das últimas décadas deixando para o próximo capítulo a teoria da consciência. Assume-se portanto a independência da primeira relativamente à segunda.

2.1.1 *Linhas de análise da teoria do conteúdo: a formulação da Teoria dos Sistemas Intencionais (TSI), a oposição à TSI, especialmente da Teoria Representacional da Mente (TRM), os princípios behavioristas da TSI, a sua ligação ao evolucionismo e a uma teoria geral do design.*

Procurar-se-á esclarecer a particular teoria do conteúdo defendida por Dennett a partir de três linhas de análise: (1) as formulações sucessivas da Teoria dos Sistemas Intencionais (TSI), (2) a

[179] DRETSKE 1988b: 511.

oposição à TSI, exemplificada especialmente pela Teoria Representacional da Mente (TRM) de Jerry Fodor e (3) a enunciação dos princípios behavioristas da TSI, intimamente ligados com o evolucionismo biológico.

Relativamente à primeira linha de análise, ao longo da sua obra Dennett especifica pontos particulares sem alterar a posição básica apresentada em C&C, de acordo com a qual o conteúdo é uma atribuição interpretativa global a sistemas físicos que manifestam comportamento adaptado ao ambiente. Dos anos 70 aos anos 90 três artigos constituem, sucessivamente, a referência canónica da teoria: *Intentional Systems* (1971[180]), *True Believers* (1981)[181] e *Real Patterns* (1991))[182]. Os três artigos constituem o núcleo da exposição da TSI neste capítulo.

A segunda linha de análise da teoria dennetiana do conteúdo – o contraste com a TRM proposta por Jerry Fodor[183] – pode parecer relativamente arbitrária. Porquê escolher Jerry Fodor, de entre todos os filósofos da mente contemporâneos, como aquele perante quem qualquer posição relativa a intencionalidade se estabelece? É no entanto o próprio Dennett quem considera que compreender os atractivos da posição de Fodor quanto à intencionalidade e quanto às explicações psicológicas constitui uma espécie de iniciação[184]. A posição de Fodor será aqui considerada a posição cognitivista exemplar e, de resto, a oposição implícita ou explícita a Fodor e à TRM encontra-se em praticamente todos os escritos de Dennett sobre teoria do conteúdo. Em *Brainstorms*, o artigo *A Cure For The Common Code*[185] visa essencialmente a Hipótese da Linguagem do Pensamento, enquanto no artigo *Brain Writing and Mind Reading*[186] se defende que não é razoável esperar que alguma coisa que se vá encontrar dentro da cabeça (as representações mentais de Fodor) torne preciso aquilo que Quine deixou indeterminado (o significado, o mental). Também *True Believers*[187], o texto canónico de exposição da teoria do conteúdo nos anos 80, termina com a avaliação (e afastamento) da Hipótese da Linguagem do Pensamento. Em *Real Patterns*[188], a posição de Fodor, aí chamada um realismo intencional fortíssimo, é apontada como o extremo da escala na discussão da intencionalidade entre teóricos que partilham a mesma evidência empírica e que aceitam o naturalismo. No extremo oposto da escala encontra-se, naturalmente, o materialismo eliminativo de Paul Churchland[189]. A TSI situa-se portanto entre ambos. Também em *The Intentional Stance*, a obra que Dennett dedica, na sua totalidade, à teoria do conteúdo, a hipótese extrema de Fodor estará sempre em causa[190].

Relativamente à terceira linha de análise, imputar princípios behavioristas à TSI pode parecer estranho, já que a filosofia de Dennett nasce da revolução cognitiva. Dennett crê no entanto que a vitória do cognitivismo sobre o behaviorismo é em muitos aspectos ilusória. A crítica cognitivista foi essencial para libertar linguagens e procedimentos metodológicos de constragimentos artifi-

[180] Inicialmente publicado no *Journal of Philosophy*, LXVIII, 4, 1971, o artigo *Intentional Systems* vem a inserir-se em *Brainstorms* (1978).

[181] O artigo *True Believers* vem a ser incluído em *The Intentional Stance* (1987), tendo sido originalmente uma conferência em Oxford (1979) e tendo sido publicado em Heath (ed), *Scientific Explanation*, 1981. Tomo, com alguma arbitrariedade, a data desta publicação como referência.

[182] A texto vem a aparecer em *Brainchildren* (1998), mas foi originalmente publicado no *Journal of Philosophy* LXXXVIII, 1, em Janeiro de 1991.

[183] Cf. FODOR 1968 *Psychological Explanation*, FODOR 1975 *The Language of Thought*, FODOR 1981 *Representations*, FODOR 1983 *The Modularity of Mind*, FODOR 1987 *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, FODOR 1990 *A Theory of Content*, FODOR 1994 *The Elm and the Expert* e FODOR 1998 *Concepts*. Como guia para a obra de Fodor foi utilizado sobretudo o volume *Meaning in the Mind*, editado por Barry LOWER e Georges REY em 1991.

[184] DENNETT 1987: 130.

[185] DENNETT 1978f. O artigo foi inicialmente uma recensão a FODOR 1975.

[186] DENNETT 1978e.

[187] DENNETT 1987g.

[188] DENNETT 1998d.

[189] Cf. CHURCHLAND 1981, *Eliminative Materialism and the Propositional Atitudes*. Paul Churchland e Patricia Churchland (cf. Patricia CHURCHLAND 1986) são os mais conhecidos defensores do materialismo eliminativo. O materialismo eliminativo é a hipótese segundo a qual as entidades postuladas pela *folk psychology* ou psicologia do senso comum (as crenças, desejos, etc) serão eliminadas das explicações psicológicas e não "reduzidas", nomeadamente reduzidas a entidades admitidas nas teorias neurofisiológicas. Na prática isto significa que uma psicologia científica será neurociência e que não será possível identificar correlações entre estados mentais e estados neuronais. O materialismo eliminitivo é a razão da salvaguarda admitida quando se afirma que crenças e desejos são *quase generalizadamente* evocados para explicar o comportamento de agentes inteligentes: de facto, o materialista eliminativo não evoca crenças e desejos para explicar tal comportamento. É necessário sublinhar que o materialismo eliminativo é uma posição acerca da *explicação* em psicologia (uma posição internalista, i.e. uma posição que defende que nas explicações psicológicas apenas deve ser utilizada informação acerca do indivíduo e não informação acerca do mundo) e não uma posição acerca do conteúdo de estados mentais. A TRM de Fodor é aliás também ela uma posição internalista. No entanto a teorização de Fodor não se restringe à explicação psicológica, alargando-se ao conteúdo de estados mentais, como se verá na segunda parte do presente capítulo.

[190] Nomeadamente em artigos como *Three Kinds of Intentional Psychology* (DENNETT 1987h), *Making Sense of Ourselves* (DENNETT 1987i), *Beyond Belief* (DENNETT 1987j), *Styles of Mental Representation* (DENNETT 1987n). Nestes artigos, como à frente se explicita, Dennett procura enquadrar no âmbito da TSI a resposta a questões como as da relação entre a psicologia do senso comum *(folk psychology)* e a psicologia científica, da racionalidade e do contraste entre representações explícitas e implícitas. A TSI opõe-se à TRM relativamente a todas estas questões.

ciais, mas o behavorismo tem forças insuspeitadas. Antes de mais, os princípios behavioristas são perfeitamente sãos, apenas foram mal compreendidos e perniciosamente difundidos por alguns behavioristas e particularmente por B.F.Skinner[191]. Um dos propósitos da TSI é dar continuidade às boas intuições behavioristas – que não são senão os princípios do evolucionismo biológico aplicados à cognição – afastando as simplificações abusivas do behaviorismo skinneriano. O princípio behaviorista segundo o qual acções seguidas de recompensa se repetirão enquanto outras, não recompensadas, desaparecerão – a "lei do efeito", na terminologia de E. Thorndike, o "Princípio do Condicionamento Operante" na terminologia de B.F.Skinner, ou a "Lei do Reforço Primário" na terminologia de C. Hull[192] – é uma forma do princípio biológico segundo o qual os mais aptos se propagarão, enquanto os não aptos se extinguirão. É certo que as populações de organismos são substituídas, na teoria da cognição, por "populações" de conexões em sistemas de controlo de entidades físicas inteligentes. No entanto, em ambos os casos se trata de evolução por selecção natural a partir de um processo de geração e teste, a única via para a explicação sem petição de princípio de conformações teleológicas, sejam elas organismos, comportamentos ou estruturas cognitivas.

Em *Brainstorms*, os artigos *Why the Law of Effect Won't Disapear* e *Skinner Skinned*[193] encarregam-se do reajustamento do behaviorismo e de mostrar que o erro central de B. F. Skinner na sua meta-teoria do comportamento foi o facto de se ter iludido quanto ao âmbito de aplicação do seu próprio método, ao tomar o organismo como ponto de referência. A solução para esta situação encontra-se no próprio Skinner, numa afirmação de cuja importância este, segundo Dennett, não se apercebe, a afirmação segundo a qual «a pele não é uma fronteira assim tão importante»[194]: De facto, os princípios behavioristas aplicam-se igualmente ao comportamento interior de um sistema físico. Em *Darwin's Dangerous Idea* (DDI)[195]

[191] Cf. para exemplos desse behaviorismo, SKINNER 1957, *Verbal Behavior* (o livro que provocou a reacção de N. Chomsky que é apontada frequentemente como um manifesto da ciência cognitiva), SKINNER 1971 *Beyond Freedom and Dignity*. Para uma análise comparada do behaviorismo psicológico e do behaviorismo filosófico, cf. BYRNE 1994: 132.

[192] Cf. GARDNER 1985.

[193] DENNETT 1978c e DENNETT 1978d.

[194] SKINNER 1974, *Behaviorism at fifty*, citado em DENNETT 1978:78.

[195] DENNETT 1995.

Dennett virá a abordar extensamente os problemas levantados pela relação entre cognição e evolução. O princípio geral a que a reavaliação do behaviorismo conduz está, no entanto, há muito definitivamente instaurado: sob a teoria interpretativa do conteúdo, "fundamentando-a" a partir do interior dos sistemas cognitivos, encontra-se uma teoria evolucionista do *design*.

Ainda relativamente à terceira linha de análise, se, como se verá, o realismo intencional exponenciado por J. Fodor é um dos adversários maiores da TSI, o outro adversário maior é a ideia de intencionalidade intrínseca[196]. Ora, é em grande parte devido aos seus princípios evolucionistas e behavioristas que a TSI se oporá por princípio à distinção entre intencionalidade original ou intrínseca de sistemas e intencionalidade meramente atribuída. Esta distinção é fulcral na discussão da intencionalidade: a aceitação ou não aceitação da noção de intencionalidade intrínseca divide claramente as posições em filosofia da mente, e uma das razões para tal é o facto de a ligação da intencionalidade intrínseca com a consciência impedir a independência da questão do conteúdo relativamente à questão da consciência[197]. No entanto, nem todos os defensores da intencionalidade intrínseca consideram as questões do conteúdo e da consciência inseparáveis: por exemplo Fodor e Dretske não as consideram inseparáveis e no entanto defendem que existe uma diferença importante entre intencionalidade intrínseca e intencionalidade meramente atribuída. Na medida em que a ideia de intencionalidade intrínseca interfere com a continuidade entre princípios evolucionistas aplicados a organismos e princípios behavioristas aplicados à cognição, a questão da intencionalidade intrínseca será inserida na terceira linha de análise. As três linhas de análise da TSI apontadas serão a partir de agora exploradas de forma entrecruzada, sendo evidentemente apontadas de forma explícita.

[196] A distinção entre intencionalidade genuína e intencionalidade atribuída é analisada em HAUGELAND 1981 e vai tornar-se muito importante por exemplo nas discussões da experiência mental do Quarto Chinês de J. Searle (SEARLE 1980). A experiência mental do Quarto Chinês, apresentada no célebre artigo de J. Searle, *Minds Brains and Programs* é uma das principais apologias da intencionalidade intrínseca na literatura da filosofia da mente e Dennett dedica-lhe vários comentários. Em *The Intentional Stance*, é no artigo *Fast Thinking* (DENNETT 1987l) que são resumidas as críticas ao argumento de Searle. Dennett considera no entanto que os truques que fazem o argumento funcionar teriam sido já anteriormente expostos por si e por D. Hofstadter. O problema é que embora Dennett pense que uma refutação definitiva foi há muito apresentada, a experiência do Quarto Chinês não deixa por isso de converter novos adeptos.

[197] Considerar-se-á que esta é a posição de J. Searle: não existe intencionalidade genuína que não seja consciente.

2.1.1.1 *Jerry Fodor, a Teoria Representacional da Mente e a Hipótese da Linguagem do Pensamento.*

Na medida em que a TSI se define por oposição ao realismo intencional defendido por Jerry Fodor, começar-se-á por caracterizar este último. A primeira das muitas virtudes de Fodor, discípulo de H. Putnam, colega e interlocutor de N. Chomsky no MIT nos anos 60, em plena revolução cognitiva[198], e detentor, como Dennett sublinha, de credenciais únicas para se lançar numa empresa como a filosofia da ciência cognitiva[199], é pôr a nu o cerne do problema filosófico da ciência cognitiva. Em 1975 Fodor publica um livro, *The Language of Thought,* onde afirma aquilo que ninguém ousara explicitar, mais especificamente que os muitos e bem sucedidos estudos empíricos da linguagem e da cognição, que consideram os processos mentais como computacionais, pressupõem um certo tipo de realismo – um realismo intencional – bem como a possibilidade metafísica de leis intencionais. Qualquer teórico da cognição que admita explicações psicológicas computacionais, deve assim admitir um *medium* das computações, um sistema interno de representações simbólicas, i.e. uma "linguagem do pensamento". Como Fodor virá a dizer posteriormente, com a sua habitual frontalidade, pode parecer «cru, ofensivo e não biológico» supor, para explicar a intencionalidade e a racionalidade, que as pessoas «têm frases na cabeça»[200]. São no entanto teorias cognitivas já aceites que obrigam a essa suposição. Para Dennett a Hipótese da Linguagem do Pensamento é uma «hipótese de engenharia, convincente mas cheia de problemas, segundo a qual todos os sistemas intencionais sofisticados devem partilhar pelo menos um traço de *design*: têm que ter um sistema interno ou linguagem de representação mental»[201]. A disputa entre Dennett e Fodor é uma disputa acerca do estatuto da explicação psicológica e dos objectos da psicologia científica. De acordo com Dennett, a psicologia científica não tem o estatuto que Fodor pretende atribuir-lhe.

[198] Fodor permaneceu no MIT até 1986. Foi um dos primeiros divulgadores das ideias de Chomsky, pondo em relevo a sua importância filosófica. Fez também aí investigação em psicolinguística.

[199] DENNETT 1978: 91. Dennett considera aliás FODOR 1968, *Psychological Explanation*, «um dos textos definidores do funcionalismo» (DENETT 1987: 345).

[200] Fodor 1990a, *Fodor's Guide to Mental Representations,* foi publicado na revista *Mind* em 1985, e é aproximadamente contemporâneo de um texto de Dennett com uma intenção análoga, *Mid Term Examination: compare and contrast* (DENNETT 1987f).

[201] DENNETT 1978: xxi.

Quando Fodor inicia a sua cruzada em defesa do realismo intencional, o behaviorismo aparecia – pelo menos na psicologia e filosofia da psicologia americanas – como uma posição quase incontestada, e como aliado natural do ponto de vista empirista e naturalista em epistemologia. Por isso, «Muitos filósofos e alguns cientista [*pareciam*] sustentar que os tipos de teorias agora em geral defendidas pelos psicólogos não podem iluminar o carácter dos processos mentais (...) pois Ryle e Wittgenstein mataram este tipo de psicologia mais ou menos em 1945, e não vale a pena especular acerca das perspectivas da falecida»[202]. Contra esta ortodoxia, a explicitação do compromisso realista da investigação sobre cognição é o núcleo da psicologia especulativa de Fodor. A psicologia especulativa corresponde para Fodor ao seguinte trabalho teórico: «"Queremos dizer que se a nossa psicologia está em geral correcta, então, a natureza da mente deve ser aproximadamente esta...." e preencher o espaço»[203].

É esse intuito que conduz Fodor ao realismo intencional e à elaboração de uma teoria da explicação psicológica. Aos olhos de Fodor, a filosofia da psicologia aparecia como filosofia da ciência em estado pouco sofisticado, devido à enorme influência dos preconceitos behavioristas. Apoiado no realismo intencional, Fodor propõe-se justificar o estatuto de ciência especial da psicologia. Uma ciência diz-se especial por contraste com a ciência básica, a física, cujos conceitos e leis se aplicam sem excepção. As taxinomias das ciências especiais (como a psicologia, a geologia ou a economia) não correspondem a taxinomias físicas e as suas generalizações não são sem excepção. Estas disciplinas não deixam no entanto por isso de ser ciências. A psicologia, especificamente, é uma ciência especial que formula leis (i.e. generalizações verdadeiras, confirmadas pelas suas instâncias e que suportam contrafactuais) em termos intencionais. Também a explicação psicológica, como as outras explicações científicas, se faz por subsunção a leis. É para sustentar esta tese que Fodor argumenta a favor da existência de uma Linguagem do Pensamento.

É evidente que a Linguagem do Pensamento que Fodor propõe para fundamentar a explicação psicológica é de alguma maneira uma linguagem privada caracterizada por processos causais internos. O interiorismo assumido da posição de Fodor permite prever que dois dos inspiradores de Dennett, filósofos para quem a ideia

[202] FODOR 1975:2.

[203] FODOR 1975: ix.

de estados mentais reais e dentro da cabeça é quase herética, Ryle e Wittgenstein, serão alvo do seu ataque[204]. Tal ataque deve-se à confessada perturbação de Fodor perante a total ausência de uma teoria do pensamento no behaviorismo lógico: de acordo com Fodor, nenhum behaviorista lógico alguma vez apresentou (ou poderá apresentar) qualquer proposta relativa à natureza dos *processos* mentais nem quanto ao que poderia *causar* sequências de estados mentais. Ao contrário, a Hipótese da Linguagem do Pensamento fundamenta precisamente uma teoria dos processos mentais, e, desmentindo uma interpretação mais imediata, fá-lo sem comprometer os seus defensores com a ideia de que o pensamento num indivíduo humano se realiza numa língua natural particular ou com a ideia de que a fenomenologia humana é proposicional. Aliás, a Teoria Representacional da Mente (TRM), associada à Hipótese da Linguagem do Pensamento, não é sequer imediatamente uma teoria do conteúdo ou significado (*meaning*). Esse é outro (grande) problema para Fodor. A TRM é uma teoria sintática da natureza das entidades representacionais. A afirmação elementar da TRM de Fodor é portanto (e vale a pena enunciá-la na sua formulação original): «*No representations, no computations. No computations, no models*»[205].

Qualquer modelo cognitivista é essencialmente uma teoria da natureza de estados e processos mentais em sistemas físicos, que interpreta transformações físicas, causais, como transformações de representações. O realista intencional tem, ao contrário do behaviorista lógico, alguma ideia acerca da natureza desses estados e processos. Para Fodor, e como A. Turing definitivamente propôs, processos mentais são computações de representações. Estas são por sua vez instanciações de símbolos. Tudo isto pode ser analisado deixando em suspenso a questão do significado. Apenas propriedades de símbolos permitem compreender como é que um sistema físico pode ser intencional e racional, como é que os seus estados podem ser "acerca de" e como é que as transições cognitivas podem preservar a verdade. A teoria da mente deve portanto envolver estados mentais simbólicos e processos internos sintácticos, logo depende especificamente da ideia de uma *linguagem* interna e privada[206].

O tabu das linguagens privadas foi instaurado por Wittgenstein. No entanto, nas cândidas palavras de Fodor, «Vale a pena mencionar que, o que quer que Wittgenstein tenha provado, não pode ter sido que uma linguagem privada, no sentido em que a linguagem da máquina de um computador é privada, é impossível, pois há coisas tais que são computadores, e o que é actual é certamente possível»[207]. Convem repetir que a posição básica de Fodor em teoria da mente é acerca de sintaxe "vazia", e não acerca de semântica. O problema do conteúdo é outra questão, que Fodor vem a considerar directamente quando mais tarde e face aos apelos de Putnam ao externalismo[208], defende o solipsismo metodológico como estratégia de investigação para a teoria psicológica[209]. De acordo com o solipsismo metodológico, estados psicológicos podem por princípio ser individuados sem olhar à sua avaliação semântica. O solipsismo metodológico é, assim, uma reafirmação do internalismo que

[204] E de facto o behaviorismo lógico por um lado (i.e. Wittgenstein e Ryle) e o reducionismo biológico por outro são os alvos assumidos em FODOR 1975 (cf. *Introduction: Two Kinds of Reductionism, Logical Behaviorism and Physiological Reductionism*). Na recensão que faz do livro de Fodor, *A Cure for the Common Code* (DENNETT 1978f), Dennett pretende defender Ryle. O mal entendido entre Fodor e Ryle resumir-se-ia ao facto de Fodor ler Ryle como afirmando que materialismo e mentalismo são mutuamente exclusivos. Mas Dennett pretende também defender Fodor de ataques demasiado simples que acusam o uso de representações na teoria da cognição de regressão infinita. Dennett partilha com Fodor a crença na necessidade de uma teoria materialista da representação. O que Dennett não pode aceitar na proposta de Fodor, em virtude dos seus princípios quineanos, é a ideia de uma "escrita cerebral literal" (esta é a expressão que Dennett utiliza frequentemente para nomear a Linguagem do Pensamento). O problema da ideia de uma linguagem do pensamento é que ela envolve representações explícitas. Ora, como se verá, de acordo com a TSI nada é intrinsecamente a representação de alguma coisa. Isto exclui a possibilidade—que Dennett considera central para Fodor—de afirmar que "Só representações explícitas são representações", pois é óbvio que a ideia supõe uma descontinuidade clara entre representações e não representações. Relativamente ao outro inspirador da filosofia de Dennett, Quine, a intenção polémica de Fodor é diferente. Maugrado os ataques sempre repetidos de Fodor ao holismo em teoria da mente, uma posição cuja enorme popularidade teria tido origem em Quine, Fodor dirá em *Concepts* (FODOR 1998: 145) que frequentemente se vê a afirmar aquilo que Quine deveria ter afirmado se não fosse o seu empirismo.

[205] FODOR 1975: 31. A centralidade desta afirmação no cognitivismo é apresentada de forma exemplar em PLYSHYN 1984. No prefácio, o autor afirma explicitamente a sua dívida para com Fodor, que lhe teria mostrado para que serve a filosofia.

[206] Cf. FODOR 1975: 55, *Private Languages, Public Languages*. Os passos da "demonstração" de Fodor intitulam-se *Why There Has to Be a Language of Thought, How There Could Be a Private Language, What the Private Language Must be Like*. Cf. também FODOR 1975: 157, *The Structure of the Internal Code* e a conclusão (FODOR 1975:197) .

[207] FODOR 1975: 68.

[208] Cf. PUTNAM 1975, *The Meaning of Meaning*. O externalismo é uma posição acerca de explicação psicológica e de conteúdo de estados mentais definível por oposição ao internalismo até aqui referido e segundo a qual a referência ao que está "fora da cabeça" – nomeadamente informação acerca do mundo - é indispensável. Foram as experiências de pensamento propostas em PUTNAM 1975 (a Terra Gémea) e BURGE 1979 (a artrite) a origem do debate entre internalismo e externalismo. Ambas fazem apelo a intuições a favor do externalismo, e nomeadamente à ideia segundo a qual é o que está fora da cabeça que determina o significado.

[209] Cf. FODOR 1981a, *Methodological Solipsism as a Research Strategy*. O artigo de Fodor é uma resposta a PUTNAM 1975.

caracterizava já a Hipótese da Linguagem do Pensamento, e que estará na raíz da introdução da distinção entre conteúdo estrito e conteúdo lato no pensamento de Fodor[210]. O que justifica, segundo Fodor, o solipsisimo metodológico é a necessidade de admitir representações explícitas para se poder falar processos mentais computacionais e de explicação psicológica. Estas representações constituem conteúdo estrito e não conteúdo lato, no entanto o facto de "estarem na cabeça" garante a possibilidade de causalidade mental (local)[211]. Admite-se ao mesmo tempo sem problemas que a semântica poderá vir a ter que ser dada à parte. Por outras palavras, e com a usual perspicácia de Fodor: «o meu ponto não é, claro, que o solipsismo seja verdadeiro; é apenas que a verdade, a referência, e o resto das noções semânticas não são categorias psicológicas. O que elas são é: modos do Dasein. Eu não sei o que é o Dasein, mas tenho a certeza que existe muito por aí»[212].

Em geral, Fodor pensa que a maneira como a psicologia de senso comum (*folk psychology*) "vive" de explicações intencionais é sintomática. Explicações intencionais supõem estados semanticamente avaliáveis que têm ao mesmo tempo poderes causais, e a psicologia de senso comum pressupõe já aquilo que virá a ser "confirmado" ao nível da psicologia científica, a existência de leis intencionais. Se não houvesse verdade no que a psicologia de senso comum supõe (i.e. que existem leis intencionais) seria «o fim do mundo»[213]... A explicação psicológica supõe portanto, de acordo com Fodor, leis. Dennett, pelo contrário, sempre considerou a ideia de leis intencionais extraviada e a obsessão de Fodor por leis um sintoma de "inveja da física".Os pontos mencionados (o realismo intencional, o papel das leis intencionais na explicação psicológica, o solipsismo metodológico, a ideia de que a psicologia é uma ciência especial) são alguns dos pontos centrais do pensamento de Fodor, que têm vindo a ser reiterados de forma relativamente inalterada nos últimos trinta anos. Adoptar-se-á mais adiante neste capítulo como referência a formulação recente do programa fodoriano para a psicologia especulativa, a formulação de *Concepts*[214].

[210] Em geral o conteúdo estrito diz respeito a representações internas, i.e. à parte do conteúdo mental que depende apenas do indivíduo. O conteúdo lato diz respeito ao significado destas representações, que não pode ser caracterizado sem referência ao mundo, ao meio do sistema, e que portanto depende de factores exteriores ao indivíduo. O conteúdo estrito permaneceria, de acordo com uma caracterização simples, inalterado apesar das eventuais transposições de um sistema de um ambiente para outro.

[211] O conteúdo lato não pode ter um papel causal e portanto não pode ser utilizado na explicação psicológica. O chamado argumento causal (cf. FODOR 1987) conclui, assim, pela necessidade de um conteúdo estrito.

[212] FODOR 1981: 71.

[213] FODOR 1990: 56.

Não é difícil imaginar que do ponto de vista de um autor como Fodor a TSI de Dennett, herdeira de Wittgenstein e de Ryle, herda a incapacidade básica da teoria da mente destes autores: não explica os processos mentais e não resolve o problema de Hume. Uma teoria da cognição sem representações, sem estados internos independentes, salientes e individuáveis é impossível para Fodor e a partir do momento em que a TSI considera o mental como uma atribuição, nada resta que possa ser causalmente responsável pelos processos mentais: se não existe genuína intencionalidade, não podem existir mecanismos para a racionalidade. Se não existem *realmente* representações, os processos mentais não podem ser processos causais internos de transformação de representações. Uma posição como a de Dennett, ao negar que as representações internas e a racionalidade *realmente existam*, ao mesmo tempo que utiliza as noções de forma interpretativa, é uma maneira de evitar os verdadeiros problemas práticos da intencionalidade e da racionalidade, para além de confundir indevidamente a questão da intencionalidade de estados com a questão da racionalidade de processos, ao fazer da primeira (a intencionalidade) função da segunda (a racionalidade), na situação da consideração global de sistemas físicos. Dennett, por seu lado, admite que a Hipótese da Linguagem do Pensamento, é uma tentadora extensão da necessidade, quase generalizadamente admitida, de atribuir estados mentais a sistemas físicos para explicar o comportamento inteligente. Mas se num sentido muito fraco a hipótese tem que ser verdadeira (i.e. existem representações mentais no sentido em que existe armazenamento e processamento de informação nos sistemas cognitivos), no sentido forte (segundo o qual existem representações *explícitas* e uma linguagem do pensamento nos sistemas físicos inteligentes bem como processos definidos sobre essas entidades, o que legitima uma hipótese geral acerca do *design* para a cognição), ela tem que ser falsa. De facto, Dennett pensa que os computadores vieram mostrar à psicologia o princípio de uma solução para o problema de Hume, e sabe que Fodor pensa algo de semelhante. Considera no entanto que é preciso não conceder demais a Fodor nas suas extrapolações. No que diz respeito à

[214] FODOR 1998.

hipótese geral de uma Linguagem do Pensamento o veredicto de Dennett é que com ela Fodor comete o pecado ryleano de confundir uma resposta conceptual com uma resposta causal.

2.1.1.2 *Instrumentalismo? – De* Intentional Systems *a* Real Patterns. *A Teoria dos Sistemas Intencionais e as três estratégias (Estratégia Física, Estratégia do Design e Estratégia Intencional).*

As teorias da intencionalidade de Dennett e Fodor são posições apenas definíveis após a caracterização putnamiana do funcionalismo nos anos 60. A primeira, apoiada numa teoria quineana do significado, inclina-se para o enfraquecimento do funcionalismo, a segunda é um realismo intencional. Tendo já iniciado a caracterização da versão realista forte do funcionalismo caracterizar-se-á em seguida a TSI – um funcionalismo enfraquecido frequentemente acusado de ser instrumentalista – nas suas sucessivas exposições e clarificações, desde *Brainstorms* (BS) *a Brainchildren* (BC).

2.1.1.3 *Princípios gerais.*

As noções centrais da TSI são as noções de Sistema Intencional e Estratégia Intencional *(intentional stance)*. Esta última expressão, que Dennett adopta definitivamente, substitui o "centralismo" de C&C. Na programática Introdução a *Brainstorms*, a TSI é apresentada como o momento, numa hipotética sequência de propostas fisicalistas em filosofia da mente, em que, após a contraposição do funcionalismo de Putnam à teoria da identidade se conclui que o funcionalismo das Máquinas de Turing é uma forma demasiado forte de funcionalismo. Não basta portanto a Dennett fazer o que Putnam faz, i.e. afastar a teoria da identidade e propor o funcionalismo. É preciso também afastar o funcionalismo radical puro, desenvolvido sobretudo pelo discípulo de Putnam convertudo em ideólogo da revolução cognitiva, J. Fodor. Contra esse funcionalismo e o que ele implica quanto à natureza explícita e sentencial das representações internas, a posição de Dennett é a seguinte[215]: (1) «Proponho isto: *(x) (x acredita que a neve é*

[215] Cf. DENNETT 1978, *Introduction*, retomado em DENNETT 1987h: 65, *Three Kinds of Intentional Psychology.*

branca ≡ pode ser atribuída a x a crença de que a neve é branca de uma forma que permite previsões)»[216]. Esta hipótese aparece como alternativa à hipótese (2): (2) «(x) (x acredita que a neve é branca ≡ x realiza alguma Máquina de Turing k no estado lógico A)»[217] . O funcionalismo das Máquinas de Turing expresso em (2), que afirma a identidade dos espécimes e o funcionalismo dos tipos, por sua vez aparecera contra a hipótese segundo a qual: (3) «(x) (Mx ≡ Fx)»[218]. A expressão (3) é a expressão da teoria da identidade, a variedade de fisicalismo que afirma a identidade dos tipos mentais e físicos e que considera que «para todo o predicado mental M existe um predicado F exprimível no vocabulário das ciências físicas de modo que uma criatura é M se e só se ela é F»[219].

Do ponto de vista adoptado neste trabalho, o que é notável nesta definição programática é que a TSI é definida como alternativa ao tipo de funcionalismo associável à TRM. É verdade que por outro lado – para a explicitação daquilo que está errado com a teoria da identidade relativamente à qual quer a TSI quer a TRM se apresentam como posições anti-reducionistas – Dennett assume a crítica elaborada por Fodor em *The Language of Thought*. Aí, no capítulo *Physiological Reductionism*, Fodor apresenta argumentos contra aquela que considera ser uma tese típica de uma (má) filosofia da ciência positivista, a tese segundo a qual ciências especiais como a psicologia deveriam vir a *reduzir-se* às teorias físicas. Fodor procura explicar por que razão considera essa redução impossível. *As ciências especiais – como a psicologia ou a geologia – são ciências especiais não devido à nossa relação epistémica com o mundo mas devido à maneira como este está organizado.* E o mundo está organizado de um modo tal que nem todos os tipos, nem todas as propriedades que aparecem em leis, correspondem a tipos físicos. Nomeadamente os tipos acerca dos quais são feitas as generalizações psicológicas não correspondem a tipos físicos. Daí que para Fodor a ideia de redução fisiológica das teorias psicológicas não decorra necessariamente da (correcta) consideração da física como ciência básica, embora a maior parte dos filósofos da ciência pensem que considerar a física a ciência básica é a mesma coisa que afirmar que as teorias das ciências especiais devem *reduzir-se* às teorias físicas.

[216] DENNETT 1978: xvii.
[217] DENNETT 1978: xvi.
[218] DENNETT 1978: xv.
[219] DENNETT 1978: xv.

Para Fodor isso supõe um equívoco. As leis das ciências especiais podem ser formuladas acerca de eventos cujas descrições físicas não têm nada em comum (e se têm alguma coisa em comum, isso é irrelevante para a verdade das leis). Daqui Fodor conclui que o programa reducionista em psicologia, associado à teoria da identidade, é, *por razões ontológicas*, um caminho falhado. É perfeitamente possível defender um fisicalismo dos espécimes (é a posição de Fodor) sem se ser um reducionista e um fisicalista dos tipos. O fisicalismo dos tipos – definido por Fodor como a doutrina segundo a qual toda a propriedade mencionada nas leis de qualquer ciência é uma propriedade física – não é, segundo Fodor, verdadeiro quanto às ciências especiais. Em *The Language of Thought* Fodor apresenta uma teoria das relações entre a ciência básica (a física) e as ciências especiais (nomeadamente a psicologia) que justifica o estatuto da psicologia como ciência especial. A teoria é solicitada, repita-se, pela possibilidade empírica de os predicados das ciências especiais – nomeadamente os predicados psicológicos – serem "instanciados" por uma disjunção não sistemática de predicados na ciência básica. Essa possibilidade não afecta a unidade da ciência um vez que se reconstrói a relação entre as ciências especiais e a ciência "básica" abdicando das correspondências entre os predicados das ciências especiais e os predicados da ciência básica e deixando de considerar as generalizações das ciências especiais como leis "sem excepção". Apenas as leis da ciência básica são leis sem excepção. As leis das ciências especiais também são leis, mas leis *ceteris paribus*. Permitindo que as leis das ciências especiais tenham excepções, preserva-se e legitima-se os tipos aos quais as generalizações se aplicam, admitindo no entanto a heterogeneidade possível das descrições físicas das instâncias desses tipos. Sublinhe-se de novo que, para Fodor, o estatuto das ciências especiais se deve à maneira como o mundo está organizado e não a uma qualquer razão superficial ou humana. É a organização do mundo que Fodor pretende capturar com a filosofia da psicologia que propõe. Aceitando que «nem todos os tipos são, ou correspondem a, tipos físicos»[220], nomeadamente os tipos que aparecem em generalizações psicológicas, poder-se-á considerar, como é desejável, que a explicação psicológica é nomológica (sendo portanto propriamente uma *explicação* e não apenas uma descrição). A psicologia é, assim, segundo Fodor, uma ciência especial cujas generalizações exprimem relações causais entre estados mentais intencionalmente caracterizados cujas realizações físicas não são o mais importante.

[220] FODOR 1975:25. Como Fodor diz «even if brains were out there where they could be looked at, we wouldn't know what to look for» (FODOR 1975: 11).

Na introdução a *Brainstorms* Dennett não tem, como se disse, quaisquer problemas em remeter para Fodor o estado da discussão do reducionismo em psicologia. No entanto, os problemas de filosofia da psicologia que trata não são de todo os problemas de Fodor. Dennett sempre considerou a ideia de *leis* intencionais irremediavelmente má e os candidatos ao estatuto de géneros naturais intencionais inexistentes. A pedra de toque da TSI é precisamente o facto de considerar que a posição enunciada em (2) – um fisicalismo dos espécimes associado a um funcionalismo dos tipos – e desenvolvida por Fodor é, na expressão sempre repetida por Dennett, demasiado forte. Não é plausível considerar que a "partilha" de estados mentais por duas entidades resulta necessariamente da realização dos mesmos estados lógicos ou funcionais. A "mesmidade" de predicados mentais de sistemas é reportável exclusivamente à possibilidade de descrição e previsão desses sistemas a partir da Estratégia Intencional. É isto que exprime a expressão (1): não apenas a realização física do estado não é importante como a própria individuação do estado intencional apelando para estados internos do sistema não é importante (não é sequer possível).

De acordo com as primeiras formulações da TSI em termos de EI e SI, os estados mentais são atribuíveis por razões pragmáticas a sistemas considerados de forma holista. A maneira correcta de considerar a questão da mente é assim perguntar como é que os idiomas mentalistas são *legitimáveis* e a proposta de Dennett é a utilização da noção de SI: «Os sistemas intencionais são supostos terem um papel na legitimação dos predicados mentalistas paralelo ao papel da Máquina de Turing abstracta ao estabelecer regras para a interpretação de artefactos como autómatos computacionais»[221]. Os fenómenos mentais são fenómenos caracterizáveis em termos de Sistemas Intencionais (*intentional-systems-characterizable*) no mesmo sentido em que procedimentos de cálculo são caracterizáveis como computáveis (*Turing-computable*). A ideia é, assim, reportar a noção intuitiva de "mentalidade" à noção técnica e mais precisa de Sistema Intencional. Um SI é um sistema cujo comportamento pode ser explicado e previsto através de adscrições de crenças e desejos.

[221] DENNETT 1978: xvii.

Neste contexto Dennett defende aliás que as noções básicas da TSI – as noções de SI e de EI – são neutras em relação ao fisicalismo, embora pretendam ser compatíveis com ele.

No artigo *Can Machines Think*[222] Dennett explora o parentesco entre a caracterização do mental em termos de SI e o Teste de Turing[223]. Para Dennett, o Teste de Turing é um interruptor de discussões (*conversation-stopper*) que impede uma interminável querela acerca da essência do pensamento, sugerindo uma estratégia prática. Além do mais, maugrado as acusações de operacionalismo que sempre lhe foram dirigidas, o Teste de Turing afigura-se neutro, já que incorpora a possibilidade de separar aquilo que importa acerca de inteligência – *o comportamento inteligente* – daquilo que é adjacente ou secundário, como as características biológicas de seres inteligentes. O Teste de Turing é, segundo Dennett, suficientemente eficaz (além de natural desde sempre entre os membros da espécie humana) para o seu propósito prático. O conceito de SI tem um estatuto análogo.

Note-se que, assim sendo, a caracterização e a previsão intencionais de sistemas físicos inteligentes, legitimadas pela TSI, não apontam, mesmo que grosseiramente, ao contrário do que Fodor pretende, quaisquer géneros naturais[224] acerca dos quais a psicologia científica formularia leis. Isto significa que não existe, para Dennett, qualquer passagem natural entre as explicações mentalistas (aparentemente) causais da psicologia de senso comum e as explicações da psicologia científica, passagem que constitui um ponto central do realismo intencional de Fodor. A justificação da psicologia de senso comum no âmbito da TSI deve ser outra.

[222] Cf. DENNETT 1998 a. O artigo *Can Machines Think?*, cujo título é evidentemente uma citação do início do texto de A. Turing (TURING 1950), é de 1985. As explicações que se seguem da fundamental aceitação por Dennett do posicionamento de Turing em TURING 1950 reportam-se a esse artigo, mas a proposta de aproximação entre caracterização em termos de SI e computabilidade foi feita pela primeira vez na introdução a *Brainstorms*.

[223] Cf. TURING 1950. O Teste de Turing, proposto pela primeira vez em TURING 1950, é uma situação a que se poderia chamar experimental em que *está em causa a atribuição de mentalidade e* inteligência a entidades a partir de critérios exclusivamente comportamentais, nomeadamente a presença de comportamentos linguísticos. O teste é um jogo de imitação, jogado entre um interrogador, um humano e uma máquina. O "objectivo" é em última análise, e pensando nas competições que actualmente acontecem em torno da ideia de Turing, iludir o interrogador humano acerca da natureza de máquina de uma máquina, fazendo-o pensar que interage com outro humano.

[224] Por oposição à ideia segundo a qual os termos mentais eventualmente corresponderiam a *natural kinds* ou géneros naturais Dennett propõe, na introdução a *Brainstorms*, uma comparação dos termos mentais com o termo *fatigues*, usado por uma sociedade de humanos sem a nossa fisiologia mas com linguagem semelhante, excepto por esta referência a *fatigues* (cf. DENNETT 1978: xix-xx).

Fazer sentido do pensamento e do comportamento próprio e de outrém a partir da psicologia de senso comum é uma actividade que constantemente ocupa os humanos, e saber como é que essa habilidade se relaciona com a psicologia científica é um problema inevitável na meta-teoria cognitiva. Ora, a posição de Dennett quanto à psicologia de senso comum é uma posição de compromisso, que muitos críticos consideram insustentável. Como um realista (por exemplo Fodor), Dennett não abdica das descrições intencionais. No entanto como um eliminativista (por exemplo Churchland[225]), Dennett não considera entidades e princípios da psicologia de senso comum explicativos na psicologia científica[226]. O estatuto da psicologia de senso comum é abordado no artigo *Three Types of Intentional Psychology*[227] através da explicitação das relações entre os "três tipos de psicologia intencional", a psicologia de senso comum, a própria TSI e a psicologia cognitiva sub-pessoal. No artigo Dennett defende a futura co-existência dos três referidos tipos de psicologia intencional.

A psicologia de senso comum é, para Dennett, uma habilidade constantemente utilizada, *um método abstracto e instrumental de interpretação e previsão cujo núcleo é a EI*. Não é de modo algum uma teoria, ao contrário do que defendem materialistas eliminativos como os Chur-

[225] Como já se afirmou, os Churchland (Paul Churchland e Patricia Churchland) são os mais notórios proponentes da eliminação das entidades da psicologia de senso comum. Note-se que é precisamente em relação à psicologia de senso comum ou *folk psychology* que o Materialismo Eliminativo se define (cf. GUTTENPLAN 1994: 104-106 e CHURCHLAND 1994: 308-316, onde P. Churchland define a psicologia de senso comum como «pre-scientific, common sense conceptual framework, that all normally socialized humans deploy in order to comprehend, predict, explain, and manipulate the behavior of humans and of higher animals»). O importante na definição de psicologia de senso comum avançada por Paul Churchland é a ideia de *teoria pré-científica*, que coloca as entidades postuladas pela psicologia de senso comum (i.e. os estados mentais como crenças e desejos) a par do flogisto, do éter e das bruxas, para retomar os exemplos habituais dos Churchland. Neste quadro, é evidentemente impossível uma explicação psicológica intencional legiforme como a que é defendida por Fodor. Churchland (CHURCHLAND 1994: 310) define o materialismo eliminativo como a posição que sairá vencedora se a psicologia de senso comum for irredutível por não corresponder à teoria materialista sua sucessora. A psicologia de senso comum será substituída na prática social e explicativa e a sua ontologia terá o destino do flogisto, do fluido calórico e das esferas de cristal (i.e. o destino de entidades postuladas por teorias que vieram a ser simplesmente afastadas e não reduzidas). Em termos científicos, o materialista eliminativo propõe que a cinemática e a dinâmica da cognição substituirão o modelo sentencial. Cf. CHURCHLAND 1981 para um manifesto do materialismo eliminativo. R. Rorty e P. Feyerabend são dois anteriores proponentes do materialismo eliminativo. Para a diferença entre materialismo eliminativo e teoria da identidade relativamente à questão da redução teórica, cf. CHURCHLAND 1984. Cf. DUMMETT 1988 para uma manifestação de incompreensão perante o que se entende por *folk psychology* nestas discussões.

[226] Cf. DENNETT 1987: 55: «o conceito de crença (...) não é atraente como conceito científico».

[227] DENNETT 1987h.

chland. A posição correcta do teórico da cognição perante a psicologia de senso comum é um behaviorismo lógico: o teórico deve considerar que afirmar que alguém tem a crença C é afirmar que esse alguém está disposto a comportar-se de determinada maneira. Certamente não se aprende o que são crenças nem se aprende o que se passa no interior dos sistemas físicos que prevemos e interpretamos mediante crenças quando aprendemos a utilizar a palavra "crença". O aspecto idealizador e normativo da psicologia de senso comum traduz-se no facto de atribuirmos usualmente aos sistemas cognitivos, tal como será explicado à frente, as crenças e a racionalidade que estes *deveriam* possuir. Essa atribuição idealizadora funciona suficientemente bem, no entanto nenhum sistema físico é idealmente racional, pela seguinte razão: a racionalidade de um sistema depende do seu *design* e como todos os produtos da evolução por selecção natural, agentes racionais biológicos resultam de um processo de *patchwork*, de acumulação por aproveitamento e reutilização, são produto de uma natureza cujos processos de "decisão" são eles próprios processos de *satisficing* (na expressão de H. Simon[228]), processos de decisão que condu-

[228] Por exemplo em *The Sciences of the Artificial* (SIMON 1969), Herbert Simon analisa os processos de *satisficing* na resolução de problemas. Os processos de *satisficing* são processos de decisão que encontram soluções razoavelmente boas para problemas em situações muito complexas – situações em que a decisão deve ser tomada no desconhecimento e na incerteza – mas não soluções óptimas (estas são inalcançáveis, na medida em que não é possível medir todas as alternativas por uma função de utilidade comum) (cf. SIMON 1969: 11). O *satisficing* caracteriza a resolução de problemas pelos humanos em inúmeras circunstâncias e também "a resolução de problemas pela natureza", considerada a evolução por selecção natural como resolução de problemas, que conduz nomeadamente ao desenho de agentes racionais. H. Simon acentua aliás (e questiona) o facto de também a biologia evolucionista (além da economia, que é o objecto central da discussão) utilizar livremente suposições de optimização. O problema dos processos "utilizados" pela selecção natural para a criação do *design* do agente racional é muito importante pois põe em jogo a natureza (pura, impura) da razão real (i.e. da racionalidade de agentes físicos naturais). De acordo com H. Simon, o agente real é um *satisficer*, que «aceita alternativas suficientemente boas, não porque prefira o menos ao mais, mas porque não tem escolha» (SIMON 1969: 29). As suposições de optimização e maximização, comuns por exemplo no tratamento dos agentes económicos, seriam assim irrealistas. Para Simon existe uma diferença importante entre um agente que é um optimizador e um agente que é um *satisficer* no que diz respeito ao que será a sua escolha racional. Se no mundo real os agentes são *satisficers* o mundo real é caracterizado pela racionalidade limitada (*bounded rationality*). A racionalidade a que aqui se alude está tomada, pelo menos como ponto de partida, no sentido económico de *maximização da utilidade num processo de gestão de meios em ordem a fins*. É importante notar que em *The Sciences of the Artificial* Simon enuncia exactamente o mesmo princípio que Dennett relativamente ao estatuto da biologia e da psicologia: o papel da racionalidade nas ciências do comportamento humano é de acordo com Simon exactamente análogo ao papel da selecção natural na biologia evolucionista. A definição mais geral de racionalidade é assim a de "adaptação do sistema ao ambiente". Os fins adaptativos do sistema seriam então, retomando a definição económica atrás citada, utilizada agora por analogia, a maximização da utilidade ou benefício.

zem a resultados satisfatórios e não aos melhores resultados possíveis. Os requisitos da natureza quanto à racionalidade não são por isso os requisitos de teorias formais. Pelo contrário, é mesmo possível defender que existe uma pressão evolutiva favorável a métodos de raciocínio que se afastam dos *standards* formais de racionalidade[229]. Vários estudos empíricos da racionalidade feitos no âmbito da ciência cognitiva mostram – pelo menos aparentemente - que os humanos estão sempre bastante prontos a raciocinar mal[230].

A psicologia de senso comum é, em termos teóricos, um beco sem saída, e Dennett propõe-se substituí-la, em contextos teóricos por dois edifícios melhor construídos. Estes são por um lado a TSI, i.e. a *justificação filosófica da psicologia*, e por outro lado a psicologia cognitiva sub-pessoal, i.e. as teorias empíricas da cognição. É entre a TSI e a psicologia cognitiva sub-pessoal que se coloca de forma mais clara a questão da redução teórica e é aí que Dennett defende uma posição anti-reducionista. A TSI é *uma teoria idealizadora, holista e instrumentalista da competência*, uma teoria abstracta, tal como a teoria da decisão racional e a teoria dos jogos. Mesmo sendo uma teoria de caixa negra, que não pretende analisar os mecanismos responsáveis pelas funções dos sistemas, ela dá um sentido técnico ao

[229] Não é nada claro, por exemplo, que a lógica seja uma teoria da racionalidade: para uma teoria contar como teoria da racionalidade deverá considerar como central a noção de *agente*, o que não acontece em lógica (seria mais razoável considerar que a dedução tem esse papel central na lógica, estando em causa portanto estruturas formais e não relações agente/ambiente). *Standards* lógicos de racionalidade, a serem identificáveis de algum modo, seriam identificáveis com a validade das inferências, teorizada para cálculos formalizados e no âmbito do estudo dos sistemas formais. As questões pertinentes seriam questões como a consistência e a completude dos sistemas formais. Expressões (que podemos considerar estarem por pensamentos/afirmações) seriam "racionais" (obtidas racionalmente) se se seguissem de..., fossem sustentadas por..., no âmbito de cálculos. A ideia de *standards* lógicos de racionalidade acentua a fiabilidade formalmente controlada dos processos de obtenção ou geração de expressões. A ser alguma coisa do ponto de vista lógico, a racionalidade seria assim *validade* (de derivações). De qualquer modo, a teoria formal da racionalidade por excelência não é a lógica mas a teoria da decisão. A noção de racionalidade aí em causa é a racionalidade dos meios ou racionalidade instrumental: os agentes (caracterizados através de crenças, preferências, atribuição de utilidades a estados do mundo preferidos e probabilidades atribuídas à obtenção destes estados dadas as diferentes acções) são racionais se as suas escolhas *maximizam a utilidade esperada*.

[230] Para um panorama dos estudos empíricos e filosóficos da racionalidade cf. SMITH&OSHERSON 1995. Na psicologia, os trabalhos ligados ao nome de Amos Tversky constituem uma referência (cf. por exemplo TVERSKY & KAHNEMAN 1981, TVERSKY & KAHNEMAN 1982, TVERSKY & KAHNEMAN 1993 e SHAFIR & TVERSKY 1995). Cf. também, para uma exposição dos problemas das teorias filosóficas da racionalidade, NOZICK 1993, STICH 1990 e HARMAN 1995. Cf. ainda SUTHERLAND 1992, uma obra de divulgação feita por um psicólogo, e ZILHÃO 1998/1999 para a análise de alguns problemas gerados pelo confronto (e pelas incompatibilidades) entre concepções filosóficas e psicológicas de racionalidade.

vocabulário das "crenças e desejos". Dennett chama-lhe em *Three Kinds of Intentional Psychology*, um behaviorismo lógico *holista* (por oposição à impossível tentativa ryleana de caracterizar individualmente as crenças e os desejos como disposições comportamentais). A TSI não visa a implementação e é nesse sentido que ela é neutra relativamente ao fisicalismo. Lidar com a implementação é a tarefa da psicologia cognitiva sub-pessoal. Esta é *uma teoria da performance* a qual, no caso concreto dos humanos, deverá explicar o modo como o cérebro enquanto máquina sintáctica se aproxima do objecto impossível especificado pela TSI, a máquina semântica movida a crenças e desejos.

Em geral, e quanto às relações entre os três tipos de psicologia intencional, a ideia é que «se podemos legitimar, não temos necessidade de reduzir, e esse é o ponto do conceito de SI»[231]. Outros problemas podem no entanto surgir: o próprio Dennett nota a propósito da afirmação de N.Block em *Troubles with Functionalism*[232], segundo a qual as teorias funcionalistas tendem ou para o chauvinismo ou para o liberalismo, que a TSI é, de acordo com tais princípios, extra-liberal na atribuição de mentalidade a sistemas físicos. Fodor comentará perante este quadro que a grande virtude do instrumentalismo de Dennett é permitir que se fique com tudo o que é bom sem nada sofrer[233]. No entanto, como se verá, apesar das suas reticências quanto à possibilidade de explicações psicológicas, e perante as acusações gerais de instrumentalismo que lhe são dirigidas, Dennett vem a sentir a necessidade de defender a TSI como uma espécie de realismo. O suposto instrumentalismo da TSI seria antes um *realismo acerca de propriedades relacionais abstractas*, que relacionam o organismo com o ambiente, e que são evocadas por exemplo nas teorias da representação de Ruth Millikan[234], J. Gibson[235] e os gibsonianos, ou mais tarde, A. Clark[236].

[231] DENNETT 1987: 67.

[232] DENNETT 1987: 68, referindo-se a BLOCK 1978.

[233] FODOR 1990a, citado em DENNETT 1987:71.

[234] Cf. MILLIKAN 1984 e MILLIKAN 1993. A filósofa da psicologia Ruth Millikan é frequentemente apontada por Dennett como tendo explorado e sofisticado as suas ideias quanto à teleossemântica. Precisamente por Dennett considerar o pensamento de Millikan como um prolongamento do seu próprio pensamento acerca de semântica, é especialmente importante compreender as razões por que Millikan defende uma posição realista acerca da representação.

[235] GIBSON 1966, GIBSON 1979.

[236] CLARK 1997. Para uma referência a esta possibilidade de realismo, cf. por exemplo DENNETT 1987: 81, *Instrumentalism Reconsidered*. Dennett considera aí no entanto que a transformação do instrumentalismo em realismo acerca de propriedades relacionais abstractas não valeria o esforço...

2.1.1.4 *Intentional Systems.*

Intentional Systems é a primeira exposição explícita da TSI. A definição dos SI, sistemas cujo comportamento é descritível e previsível por meio da atribuição de crenças e desejos, é aí relacionada com três estratégias de abordagem: a Estratégia Física (EF), a Estratégia do Design (ED) e a Estratégia Intencional (EI). Para expor o tipo de funcionalismo fraco proposto, a que Dennett chega a chamar em *Brainstorms* um "funcionalismo dos espécimes de estados mentais", considerar-se-á antes de mais alguns dos exemplos práticos com os quais Dennett ilustra a teoria.

(1) Um exemplo central de *Intentional Systems* é o exemplo de um computador que joga xadrez[237]. O comportamento do computador a jogar xadrez, é de tal modo complexo que nem mesmo o próprio programador é capaz de lidar com o computador (i.e. jogar xadrez com ele) a partir da *Estratégia do Design*. A ED é definida como produzindo descrições de sistemas em termos de funções ou instruções (são disso exemplo as descrições feita pelo biólogo evolucionista ou pelo programador de computadores) e distingue-se da *Estratégia Física*, que produz a descrição mais básica (por exemplo em termos de partículas sub-atómicas) do mesmo sistema. Quando se trata de prever o comportamento do computador como jogador de xadrez, nem a EF nem a ED estão disponíveis: são ambas demasiado complexas e não permitem previsões em tempo útil. A única alternativa para a interacção com o sistema é tratá-lo como um ser inteligente. Esta capacidade de prever o comportamento de um sistema atribuindo-lhe crenças e desejos em função da racionalidade, que é completamente independente do conhecimento da física e do *design* do sistema, é a *Estratégia Intencional*.

(2) Um outro exemplo de *Intentional Systems* são criaturas de um planeta distante que se movem, multiplicam, reagem a eventos ambientes e entram em decadência[238]. Embora as criaturas sejam completamente diferentes de humanos, existiria, segundo Dennett, justificação para adoptar a EI em relação a elas e razão para esperar o sucesso das previsões elaboradas a partir desse ponto de vista.

(3) No artigo *Reply to Arbib and Gunderson*[239] aparece outro exemplo importante, o de dois sistemas artificiais inteligentes – progra-

[237] DENNETT 1978: 5.

[238] DENNETT 1978: 8.

[239] DENNETT 1978: 24.

mas para reconhecedores de rostos – que são "idênticos" (i.e., são caracterizáveis do ponto de vista da EI como idênticos na medida em que são boas aproximações do reconhecedor de rostos ideal) mas que são no entanto totalmente diferentes como máquinas e como programas.

(4) Finalmente, é considerado o exemplo das crenças atribuídas ao computador como adversário num jogo de xadrez. Imagine-se uma situação em que o computador é o nosso adversário no jogo e em que pensamos que ele pensa "Devo retirar a minha rainha o mais cedo possível" e age conformemente[240]. Esta crença é atribuível a partir da EI sem que exista qualquer instrução explícita no sentido de "Retirar rainha cedo".

Os quatro casos ilustram princípios que são, resumidamente, os seguintes.

(1) *Uma entidade é um Sistema Intencional em função de uma estratégia adoptada para a descrição e previsão do seu comportamento.* Não há nada de intrisecamente certo ou errado na adopção de tal estratégia: a decisão é puramente pragmática, nada na caracterização da EI obriga a supor que os SI *realmente* têm crenças e desejos. Apenas se estabelece que os comportamentos dos SI podem ser previstos "como-se" eles tivessem crenças e desejos. Não é necessário fazer qualquer suposição anexa quanto à matéria de que o sistema é feito, quanto à sua origem ou quanto à sua posição na comunidade de agentes morais. A suposição de racionalidade é uma suposição na medida em que apenas seria literalmente verdadeira de sistemas completamente racionais. O facto de ela ser "*apenas* uma suposição" significa que é contrária aos factos e que é acompanhada de uma meta-suposição, segundo a qual é "suposta-ser-falsa". É isso que significa considerá-la apenas "instrumental".

(2) *A Estratégia Intencional* (EI) de descrição e previsão de sistemas *deve ser considerada em relação com as estratégias alternativas que são a Estratégia do Design* (ED) e a *Estratégia Física* (EF). Uma descrição a partir da ED não "desce" tanto como uma descrição a partir da EF: para formular previsões a partir da ED basta capturar as funções de um dispositivo. O propósito do exemplo do computador que joga xadrez é mostrar que as diferentes estratégias poderiam ser, em princípio, utilizadas para a previsão dos lances do computador no jogo, mas que há uma "inclinação" para assumir a EI. A EF, «a

[240] DENNETT 1978a e DENNETT 1978c.

grandiosa e impraticável estratégia de Laplace para prever completamente o futuro de tudo o que há no universo»[241] é impossível na prática: «o número de variáveis críticas na constituição física do computador ultrapassaria o mais prodigioso calculador[242]» (o mesmo se poderia dizer em relação ao comportamento de qualquer humano). As previsões a partir da ED por sua vez partem do conhecimento do *design* funcional do sistema, neste caso o conhecimento do programa. Também ela pode ser na prática inacessível mesmo ao programador: o funcionamento da máquina, o seu jogo de xadrez, tornou-se demasiado complexo. A única maneira de lidar com a situação é adoptar a EI.

Não são no entanto possíveis descrições e previsões exclusivamente intencionais, ou permanentemente intencionais, de sistemas físicos. Os pontos em que a psicologia de senso comum abdica das explicações intencionais e racionais em relação a humanos são ilustrativos da alternância proposta entre estratégias explicativas: na prática comum da interacção humana abdica-se de descrições intencionais em casos de disrupção, como doenças mentais e lesões neurológicas. Do mesmo modo, quando a abordagem de um sistema pela EI não é possível, a TSI indica que se deve descer o nível da explicação, para a ED ou para a EF.

(3) *A EI distingue-se da EF e da ED pelo facto de envolver uma suposição de racionalidade.* Evocando o primeiro exemplo dado, isso significa que a melhor maneira de prever e descrever o comportamento do computador que joga xadrez é tratá-lo como uma entidade racional. A racionalidade de que aqui se trata não tem que ser mais do que um «*Design* óptimo em relação a uma finalidade, ou hierarquia de finalidades»[243] num determinado contexto (neste caso, no contexto do jogo de xadrez). O exemplo central de *design* cognitivo óptimo é o *design*-para-a-racionalidade de criaturas que resultam de selecção natural, i.e. a racionalidade instrumental[244]. A reportação da racionalidade ao *design* e do *design* à selecção natural trará vários problemas à TSI.

[241] DENNETT 1987: 16.

[242] DENNETT 1978: 5.

[243] DENNETT 1978: 5-6.

[244] Como afirma R. Nozick (NOZICK 1993: 133), o núcleo mínimo de acordo entre teorias da racionalidade é formulável em termos da teoria da decisão racional. A racionalidade instrumental consiste em procurar atingir eficazmente os fins que se tem de acordo com as crenças (acerca do mundo, acerca de si próprio) que se tem, mobilizando os meios de que se dispõe.

(4) A EI, i.e. o tratamento de sistemas físicos como SI, *é uma estratégia de antropomorfização justificada por razões pragmáticas*, nomeadamente pela impossibilidade de adoptar a EF ou a ED em relação a certos sistemas. Esta antropomorfização diz respeito apenas à racionalidade na acção, não envolve a importação de quaisquer outras características (por exemplo biológicas, como cópula ou nutrição) dos humanos. O ponto é exemplificado pelas hipotéticas criaturas num planeta distante que poderiam ser previstas do ponto de vista da EI: o facto de ser possível adscrever crenças e desejos a tais criaturas prevendo assim as suas acções mostra também a interdependência da atribuição de crenças e desejos: o único ponto de ancoragem da atribuição intencional são as regularidades comportamentais, nas quais essas crenças e desejos se "manifestam".

(5) *As crenças adscritas aos SI são em geral apropriadas ao ambiente.* A apropriação de uma crença ao ambiente consiste na sua verdade: «a capacidade de acreditar não teria qualquer valor de sobrevivência a não ser que fosse uma capacidade de acreditar em verdades»[245]. *Os SI têm portanto em geral crenças verdadeiras.*

Uma vantagem imediata da abordagem do mental através da noção de SI é o facto de ser muito mais simples, para o teórico da mente, decidir se um sistema é um SI do que decidir se uma entidade é consciente, auto-consciente ou capaz de moralidade. O conceito de SI é uma fonte de ordem em teoria da mente, não incorrendo em compromissos metafísicos pouco explícitos ou inauditos: a suposição de racionalidade que possibilita as descrições e previsões a partir da EI é a mesma que guia as explicações do comportamento animal e que guiaria as previsões do comportamento de entidades estranhas em cujos comportamentos fossem reconhecíveis padrões. É certo no entanto que os princípios enunciados passam totalmente ao lado da questão da consciência, deixando a TSI vulnerável perante a acusação de falhar na tentativa de capturar as características do mental não reportáveis à inteligência e à racionalidade (as *program resisting features of mentality*), nomeadamente a consciência (exactamente o ponto em que Ryle falhou com a sua análise anti-cartesiana do conceito de mente). A definição austera de consciência (como apercebimento-1, na formulação de C&C) não constituiria assim um instrumento suficiente para captu-

[245] DENNETT 1978: 17.

rar os episódios de experiência consciente[246]. A única saída de Dennett é reiterar que a TSI constitui de facto fundamento suficiente para uma teoria da consciência e que uma teoria das *program-receptive features of mentality* é o primeiro passo e o único caminho possível para uma teoria das *program-resisting features of mentality*. Embora Dennett admita que o reportar linguístico não exaure a riqueza do que se possa num instante dado numa vida mental e que o apercebimento-1 representa sempre uma amostragem parcial, ele defende que é a noção de apercebimento-1 que explica o que é ter uma fenomenologia. Assim sendo, de acordo com a TSI o núcleo da fenomenologia é apenas incorrigibilidade. É portanto a qualidade priveligiada do auto-acesso que constitui a subjectividade. Ter uma fenomenologia é ser incorrigível mas não infalível: a (suposta) infalibilidade envolvida na situação de acesso privilegiado é uma incorrigibilidade acerca do que se quer dizer, quer se o diga quer não, e não acerca do que se diz e é portanto muito mais restrita do que usualmente se supõe. A crítica à TSI que sublinha a insuficiência desta como teoria da consciência é deste modo afastada. Ela é no entanto possivelmente a mais séria crítica e será retomada no Capítulo 3.

2.1.1.5 *A suposição de racionalidade, o crente perfeitamente racional e o design da racionalidade. Das virtudes do behaviorismo skinneriano à TSI.*

A partir do momento em que se utiliza uma suposição de racionalidade, está de alguma forma criada a obrigação de explicar tal racionalidade. De um ponto de visto prático imediato, impõe-se por exemplo decidir quão racionais têm que ser os sistemas físicos para serem considerados como SI, sem que o intérprete cometa um erro táctico. Mas impõe-se sobretudo estabelecer se, de acordo com a TSI, os sistemas são *realmente racionais* ou se, pelo contrário, é apenas pragmaticamente útil tratá-los *como se* fossem racionais.

Considerando os humanos como o protótipo dos sistemas físicos racionais, a primeira verificação que se impõe é que, se eles são

[246] DENNETT 1978: 30. É a crítica de K.Gunderson, mas também de T. Nagel e de D. Chalmers. Para uma repetição mais recente da crítica, cf. CHALMERS 1996: 361: «Dennett (1969) distinguishes two kinds of awareness, the first associated with verbal reports and the second more generally with the control of behavior, although neither of these is a clearly phenomenal notion».

seres racionais, são seres racionais tais que são perfeitamente capazes de inúmeras irracionalidades, tais como acreditar em contradições, não acreditar nas consequências daquilo em que acreditam, tomar conjunções de condições como tendo uma probabilidade mais alta do que apenas uma das condições isoladamente considerada, etc. Em suma, os humanos não são crentes *perfeitamente* racionais mas antes sujeitos de uma racionalidade limitada, com recursos (tais como atenção, memória, tempo, informação) limitados, e por isso mesmo crentes propensos a adoptar estratégias heurísticas para fixar crenças e tomar decisões. Nestas circunstâncias práticas de limitação de recursos, não é aliás sequer possível, e esse é um ponto muito importante, afirmar taxativamente que a adopção de heurísticas e estratégias racionais imperfeitas é irracional.

É um problema de fundo da TSI ao mesmo tempo supor e não poder indefinidamente continuar a supor que a racionalidade é abstractamente caracterizável. Supor que a racionalidade é abstractamente caracterizável não é estranho: é algo comum em teorias normativas da racionalidade como a teoria da decisão e a teoria dos jogos. As teorias referidas são no entanto assumidamente abstractas e o estatuto da TSI não assim é tão nítido: apesar da professada neutralidade da TSI relativamente ao fisicalismo ela pretende ser no mínimo compatível com ele e portanto não pode permitir-se ignorar completamente a constituição física e funcional dos agentes racionais.

Um outro problema inerente à noção de SI e à correlativa suposição de racionalidade é o facto de elas se traduzirem num alargamento da noção de racionalidade muito para além do comportamento racional consciente de tipo humano (nomeadamente a animais e máquinas bem como a quaisquer outras entidades que venham a ser confrontadas e que justifiquem (comportamentalmente) ser tratadas como SI). Isto significa que, para Dennett, o pensamento e a intencionalidade de tipo humano são uma espécie do género dos sistemas físicos justificavelmente tratáveis como sendo racionais. Mas se o contrário fôr verdadeiro, i.e. se aquilo a que é possível chamar sem alargamento injustificável e no sentido próprio "racionalidade" fôr uma espécie no seio do tipo humano de uma intencionalidade que é mais alargada, terá sido uma má estratégia apelar à racionalidade para explicar uma noção que é, afinal, mais fundamental, a noção de intencionalidade. Esta deveria em alternativa, por exemplo, ser ligada, à percepção, à consciência ou a relações computacionais com representações internas, tal como a TRM propõe.

Como se disse, a TSI deve de algum modo justificar a decisão de tratar um sistema como racional. Para essa decisão são necessários critérios. Um critério possível seria por exemplo a "crença" do sistema em verdades lógicas. O problema da relação entre verdades lógicas e crenças de agentes cognitivos não é evidentemente um problema específico da TSI e é aliás duvidoso que a obediência a regras lógicas em raciocínios feitos por agentes envolva qualquer crença explícita em verdades lógicas. Este é no entanto um problema que Dennett considera directamente, assumindo que a possibilidade de prever o comportamento de um sistema através da atribuição de racionalidade não obriga o intérprete a atribuir a esse sistema qualquer crença em verdades da lógica. O que está em causa na imputação de racionalidade é apenas o seguimento de regras (vagamente) lógicas. Um outro ponto é desde logo também admitido: nem todas as regras seguidas por um agente cognitivo poderiam ser substituídas, sem regressão infinita, por crenças na verdade dessas regras. Regras tácitas de inferência são indispensáveis[247]. Não poderia portanto dar-se o caso de existir um sistema cognitivo físico no qual e para o qual as regras da sua própria racionalidade fossem totalmente explícitas.

Em suma, não apenas (1) não existem sistemas físicos genuína e perfeitamente racionais na medida em que a racionalidade é uma atribuição normativa, como (2) não existem sistemas físicos nos quais e para os quais todas as regras da racionalidade própria sejam explícitas, como (3) o estado normal da racionalidade actual de sistemas físicos é a vulnerabilidade (a interferências e disrupções provocadas por estragos no *design* ou na física do sistema) e a imperfeição. Do ponto de vista da TSI, o que é certo é que quanto mais irracionalmente se comportar o sistema, mais forçado é o movimento do intérprete em direcção a outras estratégias. De qualquer modo, sejam quais forem as limitações da racionalidade actual, a racionalidade de sistemas físicos, aparente quando estes são considerados do ponto de vista da EI, é possibilitada pelo *design*, cuja origem no caso dos sistemas físicos biológicos é a selecção natural. É o *design* dos sistemas que permite o sucesso na adopção da EI pelo intérprete.

[247] Cf. DENNETT 1978: 11 em que é referido, acerca deste assunto, o artigo *What the Tortoise Said to Achilles?*, publicado por Lewis Carroll/Charles Dodgson na revista *Mind* em 1895. Acerca do mesmo assunto cf. HOFSTADTER 1980: 43 e MILLIKAN 1993.

Se noutros contextos teóricos são contraídos empréstimos de inteligência análogos àquele que é inerente à adopção da EI – por exemplo em teorias económicas, quando se supõe a racionalidade dos agentes no mercado ou a maximização da utilidade esperada – sem que se considere necessário vir posteriormente a "amortizar" o empréstimo, sendo a estratégia perfeitamente legítima e eficaz na medida em que os indivíduos se comportam como aproximações suficientemente boas de agentes racionais, o teórico naturalista da mente não pode fugir à obrigação de amortizar o empréstimo de racionalidade. Não pode nomeadamente apelar à psicologia dos agentes e deixar essa psicologia para o psicólogo, já que lhe cabe esclarecer a própria possibilidade de psicologia. Ora, é precisamente para definir a sua posição perante os "empréstimos de inteligência" supostos pela TSI que Dennett assume e explicita princípios behavioristas. O empréstimo de racionalidade a sistemas físicos, i.e. a caracterização de sistemas em termos de crenças e desejos, o centralismo que Dennett defende desde C&C é, *prima facie*, profundamente anti-behaviorista: uma maneira rápida de caracterizar a postura behaviorista apresentá-la-ia precisamente como uma tentativa de não contrair qualquer empréstimo de inteligência na teoria do comportamento. Ora, se bem que a TSI defenda o empréstimo de inteligência a sistemas, Dennett pensa que o behaviorista está absolutamente certo ao temer os riscos de um tal empréstimo. A teoria da mente necessita do empréstimo de racionalidade porque o seu objecto são comportamentos e não meramente movimentos dos sistemas físicos. Não ter compreendido a dimensão da distância entre uns e outros foi o grande erro behaviorista. No entanto, o behaviorista está certo quando estabelece que a inteligência dos comportamentos não pode ser tomada como garantida, tendo que, pelo contrário, ser explicada.

É com essa intenção que Dennett propõe a utilização de descrições e previsões intencionais dos sistemas globalmente considerados e a transposição da "avareza" behaviorista em relação à inteligência para o interior dos sistemas. Obtem-se assim uma variedade sofisticada de behaviorismo que não pára na fronteira do organismo[248]. A Lei do Efeito, de acordo com a qual acções recompensadas serão repetidas sublinha a importância da selecção a partir de um processo de tentativas e erros no estabelecimento do controlo comportamento. Aliás, o princípio não é apenas inevitável no estudo do comportamento animal e humano, também surge na simulação da inteligência. Como H.Simon afirma em *The Sciences of the Artificial*[249] – e Dennett evoca frequentemente de forma aprovadora esta afirmação – a resolução de problemas, da mais elementar à mais brilhante, natural ou artificial, não envolve mais do que uma mistura variada de tentativas e erros e selectividade. A intuição behaviorista é portanto que para "inventar", ou seja, para criar o novo (seja o "inventado" uma espécie, uma rotina comportamental, um *insight* cognitivo, um novo conceito, etc) são precisos dois processos: um processo para fazer combinações, outro processo para escolher (entenda-se: e nenhum génio criativo). É este processo de geração, teste e selecção que coloca o behaviorismo na continuidade do evolucionismo biológico: a pele, ou melhor a fronteira de um sistema físico biológico, não é um limite metodologicamente relevante. De um lado ou de outro da pele, só um processo de evolução por selecção natural a partir de tentativas de combinação pode explicar as conformações teleológicas sem petição de princípio. Essa é a razão da continuidade entre o evolucionismo biológico e o behaviorismo psicológico. A explicação, sem petição de princípio, da teleologia de sistemas foi trazida à biologia pelo darwinismo e a intenção de Dennett é transpôr o darwinismo para o tratamento dos processos mentais. Ora, o darwinismo conduz a um anti-essencialismo relativamente ao estatuto das entidades resultantes de evolução por selecção natural (esta é uma das razões pelas quais Dennett afirma em *Darwin"s Dangerous Idea* que o darwinismo é um "ácido universal").

Devido a estas convicções behavioristas, Dennett procura, cuidadosamente, em *Brainstorms,* separar o conteúdo de verdade do behaviorismo de certas consequências perniciosas do behaviorismo recente. Para Dennett isto significa interpretar Skinner (é o que faz em *Skinner Skinned*[250]), procurando "extrair" de Skinner o que este deveria ter afirmado. Como é sabido, Skinner tinha muitos problemas com o mentalismo (embora, de acordo com Dennett, não soubesse exactamente porquê) e defendeu furiosamente a ideia segun-

[248] A posição de fundo de Dennett quanto ao behaviorismo pode resumir-se na seguinte observação: num certo sentido a palavra comportamento nomeia «tudo o que é intersubjectivamente observável como o comportamento do nosso coração ou do nosso RNA e ninguém se queixaria de que um modelo em ciência se limita a explicar o comportamento» (DENNETT 1987: 334).

[249] SIMON 1969.

[250] DENNETT 1978d.

do a qual não devem ser utilizados em psicologia (na expressão do seu amigo Quine) "idiomas intencionais". Ora, Quine apresenta argumentos – relativos à indeterminação da tradução, como se viu no Capítulo 1– a favor da depreciação dos idiomas intencionais e Skinner não repete esses argumentos. Ainda assim, Dennett pensa que o desespero de Skinner perante o mentalismo tem motivos válidos. Skinner percebeu o quanto as explicações mentalistas estão infectadas de "*virtus dormitiva*", i.e. o quanto explicações mentalistas em psicologia frequentemente *pressupõem* racionalidade e inteligência que deveriam estar a ser *explicadas*. O que Skinner não viu é que não é legítimo concluir daí que os idiomas intencionais não têm um lugar legítimo em psicologia. O "mentalismo" não tem necessariamente que ser uma efabulação acerca de eventos ocultos: pode ser apenas um atalho heurístico. De facto e ao contrário do que Skinner defendeu, o empréstimo de inteligência é incontornável na teoria da cognição. No entanto, isso não o exime de justificação. Mas a justificação pode ser conseguida dentro do próprio espírito skinneriano. Para não se envolver em aporias, Skinner deveria ter conduzido o seu behaviorismo mais longe, até ao interior dos organismos, adaptando-o às necessidades práticas de pressuposições mentalistas na ciência da cognição e utilizando-o para "amortizar" o inicial empréstimo de inteligência. Para Dennett, o erro de Skinner – supor que a psicologia poderia sequer existir sem idiomas mentalistas – está em parte ligado a uma espécie de dogma da individuação na teoria do comportamento: o organismo individual tem um estatuto excessivamente importante no behaviorismo skinneriano. Evidentemente, a inflação resulta do facto de o organismo individual ser o ponto de passagem da biologia à psicologia propriamente dita. Foi mesmo devido a este "dogma da individuação" que os behavioristas se viram obrigados a recorrer eles próprios a uma "*virtus dormitiva*" não mentalista, i.e. aos barrocos conceitos behavioristas cuja crítica feita por C. Taylor fora uma das inspirações de C&C. Para Dennett esses conceitos (tais como *drives* de curiosidade, punições internas, etc) só não são tão aberrantes como parecem porque cumprem uma função idêntica à da especulação evolucionista sobre a ascendência de organismos e de espécies. O problema é que Skinnner não tem consciência do estatuto especulativo de tais conceitos.

Além da interpretação de Skinner, que des-individua e interioriza a aplicação dos princípios behavioristas, é preciso segundo Dennett também separar esses princípios do anti-humanismo simplista

de Skinner, do «hábito exasperante (que Skinner tem) de identificar homúnculos (...) com o agente moral livre e responsável»[251] para eliminar em seguida ambos e afirmar que toda a explicação psicológica tem que ser um desmascaramento de noções como liberdade ou responsabilidade. Dennett admite que as caracterizações mentalistas têm um peso ético inabdicável já que é enquanto entidades mentais que entidades físicas se constituem como morais e pessoais. No entanto discorda da ideia de Skinner segundo a qual a possibilidade de uma explicação mecanicista do comportamento exclui toda a descrição mentalista, e nomeadamente a descrição moral, desmascarando esta como supersticiosa, arcaica e ignorante. Segundo Dennett, Skinner não mostrou que a psicologia não mentalista é possivel nem que a liberdade e dignidade humanas são convicções arcaicas e supersticiosas. Mas o facto de Skinner ter falhado nas suas intenções não prova que a Lei do Efeito seja um mau princípio. Pelo contrário, ela é um princípio central da teoria da cognição, um princípio que deve ser combinado com a estratégia "centralista", i.e. com a adopção da EI na teoria do comportamento. No entanto os riscos do empréstimo de inteligência, temidos com razão pelos behavioristas, só poderão ser evitados com a investigação dos mecanismos cujo funcionamento apropriado "garante" o conteúdo atribuído "a partir de cima e a partir de fora" pelas descrições mentalistas. Para Dennett esta é por exemplo a estratégia de investigação da cognição na IA, a que lhe serve constantemente de modelo.

2.1.1.6 *O valor de sobrevivência das crenças verdadeiras e a predominância do normal. A tensão entre racionalidade e incorrigibilidade. Racionalidade, holismo e indeterminação: as crenças nucleares e as outras crenças. Crenças e opiniões.*

O desenvolvimento de *design* por um processo de selecção natural deve permitir explicar não apenas a racionalidade como também as próprias crenças. Em *Intentional Systems,* Dennett defende que a crença em verdades tem valor de sobrevivência e que é um traço constitutivo das crenças o facto de elas serem na sua maioria verdadeiras. É essa, aliás, a raíz da normatividade inscrita na noção comum de crença. A noção de crença é normativa na medida em que "acredito

[251] DENNETT 1978d.

que p" parece implicar "deve-se acreditar que p". Este teor normativo não tem a ver com a voluntariedade do enunciador ou com uma ordem dada ao interlocutor mas sim, segundo Dennett, com a ligação íntima entre crença e verdade, a ser explicada em termos de *design*. Um sistema físico desenhado para acreditar em falsidades nunca teria evoluído por selecção natural: ele é uma entidade impossível (como um peixe solúvel, diz Dennett). Aliás, nem sequer é possível descrever coerentemente um sistema desenhado para ter crenças falsas: como qualquer funcionamento resultante de evolução, as crenças são adaptadas apropriadamente ao ambiente. O estado normal e apropriado de uma crença é portanto ser verdadeira.

Não apenas é "normal" que (1) a maior parte das crenças de um sistema sejam verdadeiras, como também é normal que (2) a maior parte das crenças reportadas por um sistema sejam reportadas com verdade (recorde-se o estatuto da incorrigibilidade, analisado no Capítulo 1). Estas condições são também as condições da adscrição de crenças a sistemas a partir da EI. A já referida dependência recíproca entre a atribuição de crenças e a atribuição de desejos evidencia também ela a predominância do normal: em geral as crenças os e desejos são aqueles que o SI deve ter dadas as circunstâncias, o que o fará fazer o que seria racional fazer dadas as crenças e os desejos. Em caso contrário, haverá disrupção e impossibilidade de adscrição mentalista: o comportamento de um homem que foge a correr para debaixo de uma árvore, por exemplo, é um indício comportamental de que ele acredita que está a chover apenas se fôr assumido que o homem deseja manter-se seco. Esta circularidade é também, segundo Dennett, uma maneira de contornar a suposta privacidade de crenças e desejos. Há no entanto aqui um problema: as pré-condições (1) e (2) para a adscrição de crenças a sistemas estão em tensão e podem entrar em conflito em casos particulares de interpretação de sistemas. São indubitavelmente frequentes os casos em que uma das duas suposições deve cair para que se possa fazer sentido de um comportamento de um sistema. Se, por exemplo, um homem fosse perfeitamente racional, não poderíamos confiar na precisão do seu reportar quando ele afirma possuir crenças que são contraditórias entre si ou que foram empiricamente infirmadas (como poderia um sistema perfeitamente racional ter chegado a ter tais crenças?). Por outro lado, se o seu auto-acesso é de facto privilegiado, não há razão para pôr em causa que ele tenha de facto as crenças contraditórias que reporta. Mas então deixará de ser possível considerá-lo racional! Nenhuma das duas posições é estável, e para Dennett, esta instabilidade caracteriza a situação do intérprete perante qualquer SI actual.

A situação é então a seguinte: de acordo com um sentido mínimo de racionalidade como consistência, o conjunto das crenças de um SI actual deve ser, por princípio, considerado como consistente. No entanto, de facto, alguma inconsistência é sempre permitida e apenas a ilogicidade em grande escala indica ou um defeito tão sério no sistema que o desqualifica como crente ou um defeito nas hipóteses de interpretação. Dennett deseja sobretudo mostrar que a possível inconsistência só pode, por princípio, ser ela própria pensada no âmbito de uma caracterização *holista* do sistema. Esta posição proporciona mais um ataque importante ao atomismo das representações mentais defendido pelo realista intencional.

No artigo *Brain Writing and Mind Reading*[252], dois casos-teste são utilizados nesse sentido. O primeiro é o caso de Sam, um crítico de arte honesto e competente que afirma que o seu filho, um pintor medíocre, é um pintor de génio e promove a sua pintura. Dennett pretende mostrar que não é por princípio possível decidir que uma hipótese de interpretação (escolhida nomeadamente entre as hipóteses (1) "Sam não acredita que os quadros sejam bons, mas por lealdade e amor decide ajudar o seu filho" e (2) "O amor de Sam pelo filho cegou-o quanto aos defeitos dos quadros e ele de facto acredita que os quadros são bons") é a boa hipótese, i.e. que ela é a verdadeira crença de Sam. A razão para a impossibilidade da decisão é quineana: não há *matéria de facto* que possa fazer com que uma ou outra hipótese seja verdadeira. Dennett faz notar que também Quine, no artigo *Quantifiers and the Propostional Attitudes,* analisando uma situação análoga à de Sam põe em dúvida que a questão "será que Ralph realmente acredita que Ortcutt é um espião ou será que Ralph realmente acredita que Ortcut não é um espião?" tenha sentido[253]. A única estratégia possível para a abordagem do caso de Sam

[252] DENNETT 1978c.

[253] QUINE 1966 a. É neste artigo que Quine propõe o exemplo que se tornará uma referência para o tratamento dos objectos de atitudes proposicionais e nomeadamente para o tratamento da quantificação no contexto dos idiomas de atitudes proposicionais, o exemplo de Ralph e Ortcutt. No exemplo de Quine, Ralph desconfia que um homem de chapéu que já viu diversas vezes em circunstâncias suspeitas é um espião. Ralph sabe também que existe um homem de cabelo cinzento, Ortcutt, que é um pilar da comunidade e que Ralph pensa ter visto apenas na praia. Ralph não sabe que os dois homens são um só. Dennett está interessado em afirmar a partir de Quine que não há razões para defender que Ralph acredita que Orttcutt é um espião.

consiste em considerar a sua biografia comportamental. Nenhum "neurocriptógrafo" poderia alguma vez ler no cérebro de Sam as crenças de Sam, incluindo a crença, possivelmente auto-enganadora (*self-deceiving*), segundo a qual o seu filho é um pintor de génio.

A intenção principal do artigo *Brain Writing and Mind Reading*, um dos artigos fundamentais de *Brainstorms*, é negar a existência de representações internas explícitas, *à la* Fodor, que poderiam ser recuperadas e decifradas individuadamente. É relativamente a essa intenção que se compreende a objecção à figura do neurocriptógrafo. Um neurocriptógrafo não poderia nunca, por princípio, ler qualquer coisa no cérebro de um indivíduo porque pura e simplesmente não existe escrita cerebral explícita. As crenças de Sam, ou de qualquer outro sistema, não podem ser pensadas como coisas individuadas, recuperáveis e decifráveis uma a uma. Ora, que essa fosse a forma de existir das crenças seria a condição necessária para a decisão quanto à existência de *uma* hipótese interpretativa verdadeira acerca da situação de Sam. A principal razão pela qual Dennett se opõe à ideia de escrita cerebral explícita é, assim, o facto de considerar que a abordagem dos SI é necessariamente holista. Crenças particulares não são isoláveis e identificáveis e por isso não poderia haver uma decisão de um neurocriptógrafo – decisão que poderia mesmo, por hipótese, pensar-se como independente do conhecimento de Sam (!) – quanto à presença ou ausência de uma particular crença na vida mental de Sam.

O holismo da abordagem intencional é ainda ilustrado em *Brain Writing e Mind Reading* através de um segundo exemplo. Procurando explicitar as características de uma hipótese defensável de escrita cerebral[254], Dennett propõe um caso que põe directamente em risco tal hipótese. Não se trata, como na situação de Sam, de decidir se alguém tem ou não tem uma particular crença, mas sim de imaginar uma situação em que uma crença isolada é artificialmente inserida numa vida mental. A crença (falsa) "Tenho um irmão em Cleveland" é inserida em Tom. A questão é saber se essa inserção perturbaria ou não a racionalidade de Tom enquanto crente. Considerando que Tom declara professar a crença referida, logo alguém poderia perguntar-lhe "E como se chama esse irmão?". Tom verificaria en-

[254] As condições para uma hipotética escrita cerebral explícita, i.e. uma Linguagem do Pensamento, seriam a generatividade do sistema de representações, a correspondência entre semelhanças e diferenças sintácticas em tal linguagem e semelhanças e diferenças físicas, a saliência física dos espécimes (*tokens*) das representações, a coerência biográfica das representações, a existência de um mecanismo de leitura ou *playback* e a consistência do "armazém de crenças" (cf. DENNETT 1978e).

tão não estar disposto a declarar o que quer que fosse. Em abstracto dir-se-ia que a questão seria saber se Tom diria então "O que é que eu disse? Eu não tenho nenhum irmão em Cleveland!" ou "Não sei", continuando disposto a professar as crenças inconsistentes "Sou filho único" e "Tenho um irmão em Cleveland". Mas para Dennett, o que o exemplo demonstra é que a ideia de uma crença isolada é indefensável: é um traço definitório da noção de crença o facto de as crenças viverem num meio de coerência biográfica e lógica em constante reestabelecimento; não é sequer concebível a inserção de uma crença isolada numa vida mental.

Continuando com as concessões ao hipotético sistema de escrita cerebral, admitindo provisoriamente a sua existência, colocar-se-ia a questão do espaço físico que este ocupa. A teoria do conteúdo trata de sistemas representacionais físicos e a questão do espaço ocupado torna-se mais séria quando se encara a quantidade (a infinidade, para Dennett) de crenças implícitas que por exemplo os humanos têm. Crenças implícitas são aquelas que os sistemas têm sem nunca as terem aprendido explicitamente como "factos". Provavelmente o sistema – um humano, no caso – nunca utilizará tais crenças, mas indubitavelmente estaria disposto a asserir se tal fosse necessário, que "o sal não é açúcar", que "Nova Iorque não é na lua", etc. Segundo Dennett, se essas crenças implícitas, potencialmente infinitas, estivessem representadas explicitamente criariam um problema de espaço de armazenamento no cérebro. O facto de o problema não se colocar constitui mais uma prova da inexistência de representações explícitas. Se, num cenário alternativo, existisse um mecanismo de extrapolar e deduzir a partir de uma biblioteca básica de crenças do sistema, caso em que existiria um núcleo de crenças a partir do qual seriam extraídas consequências apenas quando (e se) isso fosse necessário, o problema seria a necessidade de informação para a recuperação e análise da informação. Ainda assim se chegaria, segundo Dennett, a um idêntico problema de excesso. Dennett argumenta por isso que a ligação da noção de crença a qualquer noção de informação explícita conduz a um círculo vicioso e que a única forma de evitar a situação é considerar que a "escrita" (i.e. as representações explícitas, que são para o realista intencional as representações *tout court*) é apenas uma forma dependente de armazenamento de informação. No caso humano, o cérebro deve armazenar informação de algum outro modo que não representações explícitas. Mais especificamente, Dennett defende que existe uma distinção impor-

tante entre as representações disponíveis para uso pessoal consciente, próximas do modelo da "escrita" e ligadas ao uso das línguas naturais (neste caso estarão por exemplo os conhecimentos de factos) e as representações implícitas, que sustentam o sistema como o tipo de ser que é, sem que precisem de aparecer explicitadas ou individuadas à (hipotética) consciência.

É neste ponto da teoria da crença que Dennett introduz uma distinção à qual não deixará de recorrer, a distinção entre crença e opinião[255]. Considerando SI humanos, as coisas que, por um processo de busca, damos por nós prontos a dizer, e que são determinadas, individuadas e linguisticamente formuladas, não são as nossas crenças, mas os nossos *juízos ou opiniões* (o exemplo de Dennett é "Balzac casou-se em Berditchev"). Estes juízos ou opiniões são definíveis como estados com conteúdo de utentes de linguagem e distinguem-se das crenças *tout court*, estados com conteúdo, mas mais profundos e que existem também noutros animais, mas que, ao contrário das opiniões, não ocorrem nos sistemas cognitivos serialmente, individuadamente, articuladamente e explicitamente. Segundo Dennett, podemos ter em cada instante indefinidamente muitas crenças, mas só podemos estar a pensar um pensamento. Ora, nós temos autoridade sobre os nossos juízos, justificada por razões funcionalistas, mas não sabemos sequer se eles correspondem às nossas crenças. Este espaço entre juízo e crença é aliás o espaço lógico para o fenómeno, tão frequente nas vidas mentais humanas, do auto-engano (*self-deception*). É o descolamento ou desajuste entre juízos e crenças que justifica, finalmente, a afirmação de Dennett em *Brain Writing and Mind Reading* segundo a qual mesmo que o neurocriptógrafo viesse a descobrir que o último juízo pronunciado por Sam antes de morrer tinha sido "O meu filho é um grande artista" ele não poderia afirmar taxativamente que essa era a crença de Sam, pois tal opinião poderia ser um auto-engano.

[255] Ao longo do período em análise no presente capítulo, Dennett sempre defendeu e explorou essa diferença que tem obviamente um peso pragmático e ético (cf. nomeadamente DENNETT 1978q, *How to change your mind*, DENNETT 1987j, *Beyond Belief* e DENNETT 1998c, *Two Contrasts: Folk Craft versus Folk Science and Belief versus Opinion*).

2.1.1.7 *A oscilação entre instrumentalismo e realismo: o realismo enfraquecido, os verdadeiros crentes e os padrões reais.*

2.1.1.7.1 *True Believers*

Em *True Believers*, o ensaio nuclear de *The Intentional Stance*, empreende-se uma nova caracterização da EI. Dennett defende então (obviamente procurando demarcar-se de classificações da sua posição como insustentavelmente instrumentalista[256]) que embora a decisão de adoptar a EI seja livre, os factos relativos ao seu sucesso ou malogro são perfeitamente objectivos. Noutras palavras, os estados mentais são agora considerados padrões *reais* e não apenas interpretações úteis e a posição geral acerca de intencionalidade é apresentada como um realismo enfraquecido ou moderado (*a mild sort of realism*). Apesar da concessão realista, Dennett continua a defender que o facto de considerar reais os padrões envolvidos na intencionalidade não o obriga a supor que um «padrão real seja produzido por outro padrão real aproximadamente isomórfico a ele, no cérebro das criaturas inteligentes»[257]. I.e., os padrões, para serem reais, não têm que ser as representações internas explícitas de Fodor.

Em *Mid-Term Examination*[258], o artigo que conclui *The Intentional Stance*, Dennett procura situar este realismo moderado relativamente a teses concorrentes. As teses-limite da tipologia são o mentalismo brentaniano e o behaviorismo quineano e, da "geografia" apresentada, Dennett conclui com pouca modéstia, que «os filósofos, sendo aproximadamente sistemas intencionais racionais estão gradualmente a persuadir-se de que Dennett está certo»[259]. Isto significaria que, na metáfora organizadora de *Mid-Term Examination*[260], eles se aproximam de zonas (ainda temperadas) do Sul do Equador, afastando-se dos limites e levando a sério a ideia quineana de não levar a sério os idiomas intencionais continuando a utilizá-los. Dois filó-

[256] Cf. DENNETT 1993: 210: «I once made the mistake of aquiescing, in a cooperative spirit, when Ned Block suggested that I should let myself be called an instrumentalist».

[257] DENNETT 1987: 34.

[258] DENNETT 1987f: 339.

[259] DENNETT 1987: 350.

[260] Na geografia imaginária de *Midterm Examination*, o "Pólo Norte" é o mentalismo brentaniano, o Pólo Sul o behaviorismo quineano. Dennett considera que, de facto, todos os autores exploram territórios mais temperados. Fodor está, evidentemente, próximo do Norte. O bom rumo geral é para Dennett uma gradual migração do Sul para do Equador, definida pela condição de "levar a sério não levar a sério os idiomas mentalistas".

sofos cujas trajectórias são na geografia imaginária de *Midterm Examination*, bastante semelhantes, Davidson e Dennett, são conduzidos da indeterminação quineana da tradução à formulação da teoria da mente como uma teoria da interpretação. Dennett distingue-se no entanto de Davidson fundamentalmente pelo facto de este se ocupar mais directamente de ontologia, defendendo que cada evento mental é idêntico a um evento físico e procurando conceber a interpretação nesse quadro. O seu argumento depende de uma doutrina acerca de relações causais formulada em termos de eventos, que lhe permite manter o realismo acerca de crenças. Dennett admite não estar tão directamente interessado em ontologia como Davidson, nem ser tão realista acerca de estados mentais, ao ponto de os querer numa ontologia científica. Para Dennett, Davidson, ainda que considere acertadamente que a teoria do mental deve ser uma teoria da interpretação, leva as atitudes proposicionais mais a sério do que deveria, como quineano. Dennett reclama para si, por conseguinte, o título de verdadeiro quineano. O verdadeiro quineano não leva a sério as atitudes proposicionais, o que se traduz-se na ideia, que dá o título a *True Believers*, de que o verdadeiro crente será qualquer sistema cujo comportamento é previsível através da EI. Esse sistema é um crente tão verdadeiro quanto algum crente pode ser. Poder ser descrito e previsto a partir da EI é tudo o que é necesssário a um verdadeiro crente – não é legítimo procurar alguma intencionalidade genuína como ingrediente a mais do verdadeiro crente, que tanto pode ser um humano como um termostato, para evocar um célebre exemplo de *True Believers*. Esta posição conduz evidentemente à objecção segundo a qual todo o sistema pode ser um SI: até uma estante no escritório pode ser descrita como acreditando que está no centro do mundo civilizado e desejando permanecer nesse mundo[261]. É certo que perante objecções ao liberalismo excessivo nas atribuições de mentalidade, Dennett não pode controlar a extensão excessiva do conceito de SI apelando a uma intencionalidade intrínseca. Apenas pode fazê-lo apelando a razões pragmáticas, i.e. sublinhando a inutilidade de prever a partir da EI sistemas insuficientemente complexos.

Mesmo sublinhando o realismo, o realista moderado considera que a existência de crenças só pode ser confirmada pelo sucesso da

[261] DENNETT 1978: 23.

estratégia de atribuição. Esta não é uma consideração nova, mas uma das primeiras críticas dirigidas à TSI acusava esta de ser indiferente perante o sucesso obtido mediante previsões intencionais. No artigo *Reply to Arbib e Gunderson*[262], Dennett respondera a críticas nesse sentido. Tomando o exemplo atrás referido dos dois sistemas artificiais com a mesma descrição intencional (são reconhecedores de rostos), se de acordo com a TSI o que dois sistemas têm em comum quando se lhes adscreve um mesmo predicado mental não tem que necessáriamente ser um traço independentemente descritível do *design* ou da física do sistema, o problema é que as possíveis verificações das previsões nada confirmarão ou infirmarão quanto à descrição do sistema. Para Dennett, recorde-se, além de não haver razão para supor correspondências entre atribuições intencionais e traços mecânicos salientes, não se põe sequer o caso de uma atribuição mental isolada a um sistema poder ser comparada com um traço físico, na medida em que na EI estão sempre envolvidas indefinidamente muitas atribuições intencionais. É este o propósito do exemplo dos dois reconhecedores de rostos: aquilo que nos faz considerá-los como tal, como reconhecedores de rostos, é caracterizado intencionalmente. Trata-se de uma caracterização de função, e de acordo com a doutrina funcionalista, a "mesma" função pode ser cumprida de diferentes maneiras. Para M. Arbib, um dos críticos a quem Dennett responde em *Reply to Arbib and Gunderson*, isto significa que Dennett deveria admitir que não considera a precisão no relato intencional um objectivo científico importante, ou sequer um objectivo por si própria, pois não permite de facto caracterizar o sistema. Arbib considera a tese falsa e derrotista: a descrição intencional de sequências causais através de etiquetas funcionais não apenas é possível como caracteriza de facto os sistemas. Dennett admite a objecção, mas contorna-a da seguinte maneira: ele considera que o discurso intencional é de facto inescapavelmente idealizante, mas «pode ser tornado honesto pela excelência do *design*»[263].

A reafirmação e aperfeiçoamento dos princípios da TSI em *True Believers* cobre inicialmente os seguintes pontos:

(1) *Os verdadeiros crentes acreditam principalmente em verdades* (Dennett reitera o facto de esta ser uma ideia defendida também por Quine, Putnam e Davidson[264]). É óbvio que a regra da atribui-

[262] DENNETT 1978b: 26.
[263] DENNETT 1978: 27.
[264] Cf. DENNETT 1987: 19.

ção de crenças maioritariamente verdadeiras a sistemas peca por (i) atribuição excessiva e (ii) não capturar as crenças falsas. Estas, para Dennett, requerem genealogias especiais, nas quais serão encontradas essencialmente crenças verdadeiras: «as crenças falsas crescem num meio de cultura de crenças verdadeiras»[265]. No caso dos humanos, a distinção supõe a diferença acima mencionada entre crença e opinião: são as crenças que são maioritariamente verdadeiras, e não certamente as opiniões ou juízos. Estes últimos são estados cognitivos sofisticados envolvendo o compromisso com a verdade de uma particular asserção linguisticamente formulada. Por exemplo, as "crenças" para-científicas do atomismo de Demócrito (i.e. as "opiniões" de Demócrito) podiam ser completamente erradas, e eram-no certamente. No entanto ninguém poderia afirmar por isso que Demócrito não sabia por exemplo mover-se competentemente no mundo que a sua física teorizava de forma inapropriada, ou que era incapaz de lidar com objectos. Ora, para ser capaz de tais habilidades, Demócrito necessitava de ter crenças na sua maioria verdadeiras acerca daquilo que o rodeava e das suas capacidades corporais.

(2) *A regra 1 segundo a qual "Os verdadeiros crentes acreditam sobretudo verdades" decorre do princípio fundamental da TSI segundo o qual devem ser atribuídas aos crentes as crenças que estes deveriam ter.* Este princípio e o princípio relativo aos desejos, segundo o qual ao sistema devem ser atribuídos os desejos que ele deveria ter, sustentam-se mutuamente. De entre os desejos, os desejos básicos (por exemplo, tratando-se de humanos, o desejo de sobrevivência) constituem o fundo, o limite, para o pedido de justificação da posse de crenças e desejos. Consequentemente, haverá uma circularidade na atribuição de crenças e desejos (aos sistemas devem ser atribuídos desejos daquelas coisas que os sistemas crêem serem boas, e desejos quanto aos meios que o sistema crê serem os meios para alcançar os fins).

(3) *A entrada em cena da linguagem natural especifica as crenças e desejos de uma forma absolutamente única no universo dos SI.* Nenhuma entidade sem uso de linguagem poderia, no exemplo de Dennett, ter um desejo de «uma omelete de dois ovos com cogumelos pão francês e manteiga e meia garrafa de Borgonha semi-frio»[266].

[265] DENNETT 1978: 18.
[266] DENNETT 1978g: 20.

(4) Além dos princípios acima mencionados, que regem a atribuição de crenças e desejos, *há princípios que regem a atribuição de uma racionalidade mínima a sistemas, princípios segundo os quais deve ser atribuída aos sistemas a crença nas implicações das crenças que têm e a ausência de crença em pares contraditórios de crenças.* Mas este é apenas um primeiro movimento, que terá inevitavelmente que ser corrigido. A irracionalidade do comportamento e a incoerência das crenças de sistemas levanta obviamente problemas de interpretação, como desde C&C e BS Dennett vem a afirmar.

Até aqui, nada de novo. No entanto considerando que a atribuição de mentalidade corresponde a um interesse pragmático do intérprete e que a utilidade das previsões da EI aumenta de forma directamente proporcional à complexidade do sistema, um problema novo é colocado em *True Believers.* Se um SI simples como um termostato é previsível como um crente a partir da EI, sendo também ao mesmo tempo totalmente compreensível no seu funcionamento mecânico, e portanto previsível, como não sendo um crente, levanta-se a hipótese de que um sistema dado (um humano, por exemplo) poderia ser para um sistema menos inteligente (um outro humano, por exemplo) um crente, enquanto que para outro sistema (mais inteligente, sumamente inteligente) seria previsível a partir da EF, não como um crente mas como um sistema físico. Para a exploração desta hipótese Dennett evoca um exemplo usado por Robert Nozick para criticar a ideia de EI. Robert Nozick imagina marcianos para quem nós, os humanos, somos mecanicamente claros como os artefactos mais simples podem sê-lo para nós, ou como um termostato pode ser para um engenheiro. Os marcianos são físicos laplacianos que podem adoptar a EF em relação a nós. Eles não têm necessidade da EI ou sequer da ED. Para os marcianos, os humanos não teriam que ser considerados como crentes pois seriam totalmente previsíveis como sistemas físicos. Ora Dennett afirma que estes marcianos, mesmo "sabendo tudo" acerca de humanos a partir da EF, *se não vissem os humanos como SI estariam a falhar padrões reais, apenas descritíveis a partir da EI.*

O grande poder da EI como instrumento de previsão assenta na sua neutralidade e abstracção relativamente aos detalhes de execução das acções previstas: essa é a razão pela qual os padrões reais apercebidos pelos terrestres através da EI escapariam aos marcianos imaginários. O poder de previsão da EI é neutro quanto aos detalhes da implementação no sentido em que, por exemplo, saber

que um agente económico vai comprar ou vender títulos não supõe saber através de que movimentos físicos dos dedos do corrector nas teclas do computador ou em que minuto e segundo essa acção vai ser cumprida. Assim sendo, os físicos laplacianos marcianos poderiam compreender a actividade em Wall Street ao nível microfísico sem no entanto saber dizer que acções ocorrem. Poderiam prever as vibrações da voz do corrector ao dar a ordem de compra mas não saberiam que essa mesma acção poderia ter ocorrido de inúmeras outras maneiras, pois há um «número indefinido de maneiras de colocar no mercado uma ordem sobre 500 acções da General Motors»[267]. Nesta situação seria portanto sempre mais iluminadora a previsão através da EI.

Em suma, a defesa de Dennett perante a objecção de Nozick consiste em manter que padrões que apenas aparecem a partir da EI são ainda assim reais: há alguma coisa aí que será ignorada se não fôr assumida a EI. Dennett pretende reforçar esta tese apoiando-se no Jogo da Vida de J. Horton Conway, que virá a ser o exemplo central de *Real Patterns*[268]. O Jogo da Vida não se joga para ganhar ou perder[269]. Dois estados (on e off) das células de uma grelha, determinados por um regra simples, constituem toda a física (determinista) do Mundo da Vida[270]. Essa física simples dá lugar a padrões reconhecíveis e previsíveis, mas que não são realidades do nível dos estados das células individuais. Assim, uma entidade que apenas reconhecesse a física do Mundo da Vida falharia os padrões. O propósito de Dennett com o exemplo do Mundo da Vida é mostrar que a EI fornece o ponto de vista para discernir a realidade dos padrões nele existentes. Transpondo a conclusão para os estados intencionais, também estes são padrões reais, ainda que a sua realidade "dependa" do reconhecimento. A expressão "padrões reais" é a nova expressão que Dennett adopta para referir "a ordem que está lá" de Anscombe e que ele persegue desde C&C. Anscombe também não diz que a ordem deve ser encontrada no mundo, no cérebro ou no comportamento – e essa é a dificuldade da decisão quanto à realidade dos padrões do Mundo da Vida. Se se olhar correctamente, os padrões "saltam aos olhos", se não, são invisíveis. O que se deverá então afirmar acerca da sua realidade? Quão realista acerca de padrões-que-dependem-de-reconhecimento pode alguém ser?

Retomando o exemplo anterior, o ponto de vista da EI é a razão pela qual, num concurso de previsão entre um marciano e um terrestre, este último disporia de uma faculdade de previsão que pareceria mágica ao marciano. O exemplo de Dennett é a observação por ambos (um humano e um marciano) de um telefonema de Mrs. Gardner, durante a qual esta diz «Olá querido. Vens cedo? A que horas? E trazes o patrão para jantar? Compra uma garrafa de vinho à vinda para casa e guia com cuidado»[271]. Com base nisto, o terrestre prevê a chegada de dois humanos num veículo determinado, dentro de uma hora, um dos humanos trazendo uma garrafa que contem um fluido alcoólico. O marciano produz a mesma previsão mas a partir do tratamento de uma quantidade de informação extraordinariamente maior. A capacidade de previsão do terrestre baseia-se no facto de os padrões reais terem a capacidade de "absorver"(i.e de tornar desprezáveis) variações na implementação física.

É claro que o que é inconcebível na ideia do concurso de previsões entre o marciano e o terrestre é o facto de a situação supor que o marciano trata o humano como um crente racional (e não deveria precisar) e que ele aposta num concurso (e não deveria precisar). Aliás, é possível objectar que a previsão do terrestre pelo marciano a partir da EF deveria incluir a capacidade de previsão do terrestre a partir da EI (não poderia senão incluí-la, sob pena de contradição). É certo que por outro lado se poderia objectar que nada foi dito acerca da seguinte questão: será que os marcianos, que poderiam prever tudo a partir da EF, se conceberiam a si próprios como seres inteligentes? No caso de o fazerem, isso não significaria uma irredutibilidade dos padrões concebidos a partir da EI? Uma escapatória possível para o defensor dos marcianos omniscientes seria evidentemente afirmar que os marcianos não necessitariam de se

[267] DENNETT 1978: 26. Dennett afirma em DENNETT 1994 (na resposta a M. Richard) que aperfeiçoou este exemplo dos marcianos com o exemplo das duas caixas negras que aparece em DENNETT 1995: 412-419 e que será mais à frente descrito. O ponto continua a ser a existência de generalizações inexplicáveis e misteriosas para "super-laplaceanos".

[268] DENNETT 1987:38, *Reflections: Real Patterns, Deeper Facts and Empty Questions.*

[269] Cf. uma introdução ao jogo em DENNETT 1978: 105. Dennett considera-o uma obrigação para filósofos, e uma fonte inesgotável de exemplos: «Todo o estudante de filosofia deveria ser responsável por um íntimo conhecimento do Jogo da Vida (...) um instrumento essencial no kit de experiências de pensamento, um gerador prodigioso de exemplos filosoficamente importantes» (DENNETT 1998: 105). Cf. a propósito www.cecl.pt/interact.

[270] A regra é a seguinte: cada célula tem 8 células vizinhas, e para determinar o estado no próximo instante do tempo (discreto) do Mundo "conta" quantas das células vizinhas estão ON. Se são 2 a célula permanece no estado em que estiver, se são 3, a célula ficará ON, em todas as outras condições a célula ficará OFF.

[271] DENNETT 1978:26.

conceber a si próprios como inteligentes uma vez que a sua capacidade de previsão a partir da EF seria perfeita.

No caso humano, a EI, e portanto o discernimento dos padrões reais da mentalidade e da racionalidade, é de acordo com Dennett, inevitável. A EI é assumida por cada um de nós em relação a si próprio, a seres da mesma espécie e a muitos outros sistemas. No entanto os padrões, embora reais e inevitavelmente discriminados, são também inevitavelmente incompletos e imperfeitos. Como já foi observado, não existe um pensador racional real perfeito: todo o pensamento fisicamente existente sofre constantes interferências em virtude das limitações da racionalidade realizada. A existência de crenças falsas, incoerentes e inconsistentes e de falhas da racionalidade é problemática para a TSI, que se baseia, ela própria, numa suposição de racionalidade. Relativamente às falhas da racionalidade, observar-se-á que nesses casos, não existe matéria de facto que permita decidir que crenças e desejos atribuir ao sistema. No entanto «isso não corresponde a subjectivismo ou relativismo porque a decisão acerca de quando e porquê está ausente matéria de facto é ela própria uma questão de facto objectiva»[272].

A ideia da inexistência de matéria de facto que permita decidir por uma interpretação intencional, de entre dois esquemas rivais e igualmente eficazes, decorre, como já se afirmou, da indeterminação quiniana da tradução. Mas o que importa acentuar aqui é o facto de Dennett pensar que interpretações radicalmente divergentes são «na prática negligenciáveis»[273] em relação a sistemas complexos. Assim apesar da afirmação central de *True Believers* segundo a qual ser um crente é ser uma entidade descritível e previsível a partir da EI e nada mais, pode-se considerar que Dennett apresenta uma versão forte do estatuto das crenças para SI interessantes e complexos como os humanos. Quanto mais um SI é complexo mais o critério de atribuição de crenças (aparentemente instrumental) envolve constrangimentos quanto à constituição interna de sistemas (quanto ao seu *design*, afinal). No exemplo de Dennett, quanto mais se enriquecer e complexificar as ligações ao mundo de um sistema físico (por exemplo o termostato de *True Believers*, que é à partida tanto um verdadeiro crente como um humano, na medida em que é previsível pela EI) "dando-lhe" orgãos sensoriais e conhecimento de factos de modo a ter versatilidade comportamental, mais a classe dos modelos indistinguivelmente satisfatórios desse sistema vai diminuindo até que uma interpretação única se impõe na prática. Um dispositivo que passe a ter crenças e desejos ricos e específicos torna-se muito difícil de "transpôr" de um mundo para outro, pois aquilo que o dispositivo é (é-para) está ligado ao (seu) mundo de múltiplas maneiras. Esta ideia é muito importante para compreender a posição geral de Dennett perante os casos da família da Terra Gémea de Putnam[274], a experiência de pensamento exemplar do externalismo. Em geral, Dennett não aceita as conclusões putnamianas no ponto em que elas ligam o externalismo com uma suposição quanto à determinação do conteúdo dos estados mentais. As deslocações ou transposições de sistemas físicos representacionais de modo a fazer variar os conteúdos dos seus estados mentais são essenciais para a exemplificação da teoria teleológica do conteúdo. Na obra de Dennett, estas transposições começam em *True Believers* com o caso do termostato e são levadas até ás últimas consequências com os vários exemplos de *Evolution, Error and Intentionality*[275]. Elas comprovam, segundo Dennett, que o "espelhamento" do meio pela organização de um sistema é um constrangimento central a considerar numa teoria materialista da representação.

É no quadro externalista da ligação apropriada entre o *design* interno do sistema e o ambiente, que a questão das representações internas deve ser pensada. Isto significa que segundo a TSI, os estados intencionais não são atribuídos porque se encontram representações internas em sistemas, mas porque, quando se encontra um sistema em relação ao qual a EI funciona dadas estas ligações complexas ao ambiente, se interpreta alguns dos seus estados internos como sendo representações. É também essa a razão para a inexistência de uma transição abrupta entre um verdadeiro crente como um termostato e um verdadeiro crente como um humano: há apenas diferenças de grau na organização interna dos sistemas intencionais, diferenças derivadas do embebimento (*embeddedness*) no ambiente da organização do sistema.

Dennett conclui *True Believers* enfrentando uma questão geral que se coloca à TSI: trata-se de saber porque é que a EI funciona

[272] DENNETT 1978: 29.
[273] DENNETT 1978: 29.

[274] Cf. PUTNAM 1975, *The Meaning of Meaning*, para a experiência mental da Terra Gémea, uma experiência mental que, como a já referida experiência mental do Quarto Chinês de John Searle, não mais saiu de cena na filosofia da mente.
[275] DENNETT 1987.

relativamente a um sistema. Ora, existem dois tipos de resposta a essa questão. A primeira espécie de resposta remete para o *design* apropriado do sistema: a EI funciona porque o sistema está feito-para-x, e bem feito. A segunda espécie de resposta faz apelo aos mecanismos e ao seu funcionamento. No caso dos humanos e outros sistemas cognitivos biológicos isto conduz a evocar a evolução por selecção natural, responsável pelo *design* pelo qual os humanos são crentes racionais. No caso do termostato como no caso dos humanos, a primeira resposta é muito pouco informativa. A questão verdadeiramente interessante diz respeito a mecanismos e *design*. Ora, segundo Dennett, não sabemos como responder a essa questão difícil, e é precisamente por isso que uma hipótese ousada como a Hipótese da Linguagem de Pensamento de Fodor é pertinente. No entanto, a Hipótese da Linguagem do Pensamento envolve algo que Dennett, como se sabe, não aceita: um isomorfismo entre os padrões reais atribuídos (a intencionalidade manifestada no comportamento do sistema) e padrões igualmente reais no cérebro ou no interior das criaturas intencionais e inteligentes. No entanto «duvidar da existência do(s) segundo(s) não é duvidar da existência do(s) primeiro(s)»[276].

2.1.1.7.2 The Intentional Stance *e a exploração dos problemas delimitados em* True Believers.

Os restantes artigos de *The Intentional Stance* expandem e exploram os problemas delimitados em *True Believers*. Em *Making Sense of Ourselves*[277], são explorados alguns problemas envolvidos na suposição de racionalidade, em *Fast Thinking* e *Evolution, Error and Intentionality* é criticada a ideia de intencionalidade intrínseca, em *Styles of Mental Representations* é desenvolvida uma distinção entre crenças-núcleo e crenças implícitas e são apresentadas razões para não seguir os Churchland na conclusão (o materialismo eliminativo) que estes retiram da inexistência de representações explícitas. A TSI continua portanto a ser apresentada como uma posição intermédia entre realismo intencional sentencial e eliminativismo.

[276] DENNETT 1978: 34.
[277] DENNETT 1987i.

O artigo *Making Sense of Ourselves* aborda, finalmente, a questão da natureza da racionalidade. As suas conclusões fugidias são ilustrativas da "aporia" fundadora da TSI: o facto de ter que supor a racionalidade e não poder defini-la ou explicá-la. O pretexto do artigo é a resposta a críticas de S. Stich segundo as quais seria impossível para a TSI, em virtude dos seus princípios, explicar, prever ou descrever a existência de falhas, insuficiências cognitivas ou erros nos SI. Ora, para Stich uma boa teoria da intencionalidade deve permitir decidir se é verdadeira ou falsa qualquer atribuição de crenças a um SI quando um erro é cometido. Dennett pensa que Stich é presa de uma ilusão análoga à que faria alguém pensar que, numa situação quineana de tradução radical e havendo dois manuais de tradução para uma linguagem, apenas um deles poderia por princípio ser o bom manual.

Os casos que Stich discute não são casos de irracionalidade "irredutível", nos quais haveria persistência no "erro" e defesa de uma prática irracional (como aconteceria por exemplo na loucura), mas casos de irracionalidade normal (Dennett defende aliás que não são casos de irracionalidade mas de erro). Tomar-se-á aqui como exemplo o caso do rapaz que vende limonada e que dá o troco errado. Cada copo de limonada custa 12 cêntimos. Uma pessoa dá 25 cêntimos ao rapaz, ele dá-lhe um copo, 10 cêntimos e mais 1 cêntimo de troco. O rapaz cometeu um erro. O problema é saber que crenças tem ele quando comete o erro. Dennett admite que casos semelhantes são imprevisíveis a partir da EI, mas que isso acontece porque o comportamento escapa à descrição intencional, na medida em que é função de um lapso no nível intencional. De qualquer modo, é a partir da EI que se faz sentido da situação. Uma suposição de racionalidade subjaz ao facto de se esperar que o rapaz se surpreenda e embarace e que dê mais 2 cêntimos, perfazendo o troco correcto quando o erro lhe é apontado. E se o rapaz se comporta como é esperado (corando e desculpando-se) ele é fundamentalmente racional. Na descrição e previsão do comportamento do rapaz as coisas passam-se deste modo porque lhe atribuímos a crença de que deu o troco certo, o desejo de não enganar os clientes, além da crença de que 25-12=13. Portanto, mesmo se o erro não é previsível a partir da EI, só a partir dela é possível fazer sentido, globalmente, da situação na qual o erro surge como erro.

Evidentemente, é difícil dizer qual foi exactamente o erro do rapaz. Procurando listar as crenças do rapaz encontra-se que ele acre-

dita: (1) que deu o troco certo, (2) que a pessoa lhe deu 25 cêntimos, (3) que a limonada custa 12 cêntimos, (4) que 25-13=12, (5) que 10+1=11, etc. De entre essas várias crenças, apenas a crença de que teria dado o troco certo é falsa. Parece pertinente perguntar a partir de qual crença teria ele inferido essa crença errónea. Ora Dennett defende que a *crença errónea não foi alcançada através de nenhuma inferência.* Inferências são passagens de pensamento com razões e Dennett pensa que não há razão para o erro do rapaz. A melhor maneira de fazer sentido do comportamento é considerá-lo como uma irrupção de mau funcionamento. Houve uma transição sem sentido, e nenhum relato em termos de crenças e desejos capturará a razão desta. Senão, o que se deveria pensar? Que o erro foi causado pelo facto de o rapaz ter pensado explicitamente para si "25-12=11"? Mas foi-lhe atribuída a crença de que 25-12=13, para explicar o espanto e embaraço. Poderia ele acreditar que "25-12=11" e "25-12=13", em irracionalidade flagrante? É improvável. Estas considerações mostram, segundo Dennett, que a explicação procurada não é de nível intencional. Não é possível decidir o que o rapaz realmente acreditava *quando cometeu o erro.* É preciso descer de nível para explicar o caso. Conclui-se que «Erros deste tipo são (...) desvios dos bons procedimentos, não manifestações de uma adesão a um mau procedimento ou princípio»[278]. Pelo contrário, Stich pensa que deve ser possível dizer o que o rapaz realmente acreditava. Considerar-se-ia então, por exemplo, que ele teria provisoriamente "perdido o endereço" de alguma coisa conhecida (nomeadamente as suas crenças aritméticas). Dennet crê que embora uma hipótese assim possa vir a ser confirmada futuramente, neste caso ela apenas é admitida por horror à contradição, não existindo evidência nesse sentido.

O caso do rapaz exemplifica ainda o seguinte problema geral acerca da vida mental: é frequente que uma crença seja possuída por um crente sem que no entanto compareça à sua vida mental explícita. No exemplo de Dennett[279], alguém é convidado para almoçar e diz que não pode, porque sabe que não pode, mas não faz a mínima ideia da razão por que não pode. Outro exemplo: alguém foi jogar ténis à hora do almoço sabendo perfeitamente (mas sem que isso lhe tenha vindo à ideia) que tinha um encontro importante à mesma hora. Como é que crenças tais podem "escapar-se

[278] DENNETT 1978: 87.
[279] DENNETT 1987: 89.

provisoriamente" continuando a ser conhecidas? A situação não é de todo rara nas vidas mentais humanas e aliás, como nota Dennett, alguma coisa pode escapar-se-nos mesmo quando estamos a pensar nela conscientemente e a nomeá-la linguísticamente (como quando uma pessoa que diz "Cuidado! Está quente – não ponhas a mão!" coloca ela própria a mão no objecto quente). Para Dennett a questão das crenças possuídas e que não comparecem explicitamente não é uma questão de memória e de esquecimento. O problema com a hipótese do esquecimento é que se trata de um "esquecimento" tal que nenhuma das crenças "fugidias" está esquecida definitivamente, ou é irrecuperável. O problema de Dennett é estabelecer que significado pode ter a atribuição global de racionalidade a uma vida mental indubitavelmente plena de tais situações, assegurada do ponto de vista sub-pessoal por armazenamentos em memória compartimentados, modulares.

Todas as crenças acima referidas, que não comparecem no momento próprio, participam na tentativa global em que cada humano está envolvido constantemente, a tentativa de fazer sentido de si próprio. Ora, o ponto de Dennett é que os processos e actividades mentais que nos auto-atribuímos para fazer sentido de nós próprios não se aproximam muito daquilo que uma psicologia científica sub-pessoal encontrará. *Uma das razões é o facto de postularmos a posteriori demasiados componentes explícitos da vida mental.* Isto acontece em parte devido à compulsão a formular-se verbalmente que as criaturas linguísticas experimentam (como Dennett repete sempre: *How can I tell what I think until I see what I say?*[280]).

Ao contrário de Dennett, Stich assume que embora nem a TSI nem a psicologia de senso comum tenham uma resposta para o problema apresentado, a psicologia científica encontrará os componentes explícitos da vida mental, estados funcionalmente salientes e causalmente eficazes aos quais pode ser atribuído um conteúdo determinado, permitindo decidir casos como o vendedor de limonada ou os esquecimentos. O argumento fundamental de Dennett contra esta possibilidade é o facto de considerar que assim como a interpretação de um fragmento de comunicação pública de crença (uma "opinião") depende da atribuição holista das crenças e desejos, também a inter-

[280] Esta formulação de E.M.Foster é um motivo constante de Dennett, e nomeia o facto de ser qualquer coisa como um auto-questionamento que precipita a individuação das crenças, produzindo opiniões e pensamento propriamente dito. Cf. por exemplo DENNETT 1991: 193.

pretação dos mecanismos cognitivos sub-pessoais depende das crenças e desejos atribuídas à totalidade do sistema, de acordo com idênticos princípios holistas. A própria identificação de estados e processos (tais como memória de p, esquecimento de q, crença em r, desejo de s ....etc) pela psicologia cognitiva sub-pessoal depende de etiquetas intencionais. Embora a psicologia cognitiva sub-pessoal seja uma investigação acerca de *design* e funções, o investigador só pode saber o que procurar servindo-se da EI: «os mecanismos da psicologia cognitiva sub-pessoal necessitam tanto de interpretação como o comportamento exterior»[281]. Assim, não existe forma de escapar à interpretação em teoria da cognição, não apenas na descrição das vidas mentais pessoais como também na psicologia cognitiva sub-pessoal. A teoria da cognição é inevitavelmente holista e conta a história mais racional possível. Mas como fica então a questão da natureza da racionalidade se, de acordo com a TSI, é uma atribuição de racionalidade que subjaz quer às práticas da psicologia de senso comum quer à psicologia cognitiva sub-pessoal? Antes de mais, essa racionalidade não é aquilo que de um ponto de vista teórico puro (por exemplo do ponto de vista da lógica ou da teoria da decisão) parece. Ela não consiste em fechamento dedutivo das crenças[282], nem em perfeita consistência lógica, nem numa perfeita capacidade inferencial do agente. Embora se verifique na observação empírica do comportamento de agentes racionais que a descoberta de contradições suficientemente explícitas na vida mental própria é ocasião para o disparar de um "alarme epistémico", as inferências psicologicamente reais não são exactamente implicações lógicas[283] e as escolhas racionais não exactamente escolhas maximizantes feitas a partir de todo o conhecimento. Estes processos seriam demasiado dispendiosos para vidas mentais reais. Mas o problema da ideia dennettiana de descer de nível para explicar falhas cognitivas tais como o erro do exemplo, bem como da intenção de fundamentar o *design* para a racionalidade na evolução, é que o teórico da mente deixa de poder fazer apelo a teorias formais da racionalidade[284]. Ora assim, a noção de racionalidade arrisca-se a não poder

[281] DENNETT 1987: 105.

[282] «An ideal agent's beliefs are deductively closed, or closed under logical implication, if and only if any proposition logically implied by some of those beliefs is itself also believed» (HARMAN 1995: 187).

[283] Cf. HARMAN 1995: 132

[284] Como faz por exemplo D. Davidson, quando considera que a teoria (normativa, formal) da decisão explicita aquilo que se entende por racionalidade e que está implícito na psicologia de senso comum. Cf. DAVIDSON 1980c, *Hempel on Explaining Action*.

dispôr de qualquer fundamentação. Quando Dennet, por outro lado e em paralelo com a suposição de racionalidade inerente à prática da interpretação, toma a racionalidade como um funcionamento de que a evolução dotou certos agentes ele apenas pode evocar uma vaga noção de "normal": num organismo resultante de selecção natural, a maior parte das crenças são verdadeiras e as estratégias de formação de crenças são na sua maior parte racionais. Mas se é a observações acerca do funcionamento de agentes que se vai buscar a noção de racionalidade, Dennett deve admitir (e admite) ao mesmo tempo que a aludida situação normal está constantemente a ser perturbada: não apenas o *design* para a racionalidade no sistema resulta de estratégias de *satisficing* por parte da evolução, como as estratégias dos agentes cognitivos são também frequentemente estratégias de *satisficing*, heurísticas e não procedimentos formalmente ideais. A situação deixa o intérprete que atribui globalmente racionalidade apoiado no *design* natural numa posição muito pouco confortável não apenas em relação à qualidade das suas previsões como também em relação aos seus próprios poderes.

Poder-se-ia ainda objectar que a selecção natural não teria sequer por que favorecer as crenças verdadeiras por contraposição às falsas: bastaria que as crenças falsas fossem vantajosas. Quando se trata aquilo que é "racional" fora dos manuais de lógica e dos modelos de teoria da decisão não é nada claro de que se está a falar. Como já se afirmou, é mais racional em muitas circunstâncias adoptar as estratégias heurísticas a que Herbert Simon chamou estratégias de *satisficing* – obtendo crenças de certo modo irracionais por terem sido obtidas mediante processos que não consideram tudo o que seria desejável considerar –nomeadamente se o custo do cálculo exaustivo for maior do que os benefícios que dele poderiam advir. A última palavra em *Making Sense of Ourselves* é que a TSI, embora sendo uma teoria da interpretação supondo a racionalidade, não tem que dar uma definição de racionalidade mas apenas que explorar a racionalidade como termo pré-teórico para excelência cognitiva. O problema desta posição é que de acordo com ela toda a teoria depende de algo que é deixado inespecificado por princípio. A posição não é no entanto um mero afastamento de problemas. Dennett pensa que *há boas razões para resisitir a fazer declarações acerca da natureza da racionalidade*. Em primeiro lugar, decididamente não há coincidência entre aquilo que os agentes reais racionais fazem e as teorias formais da racionalidade. Em segundo lugar, não

há como negar que por vezes é meta-racional não ser perfeitamente racional: frequentemente a racionalidade consiste em «bom senso acerca de quando se apoiar em quê»[285]. Por isso mesmo, para Dennett, «o conceito de racionalidade é sistematicamente pré-teórico»[286]. Não sendo razoável ligar a racionalidade a cânones que excluam práticas reais e intuições acerca do que deve ser acreditado, não há como evitar a necessidade de nos apoiarmos em intuições acerca do que deve ser acreditado e acerca de quais regras devem ser seguidas em ocasiões particulares. Assim, a única pedra de toque do "ser racional" será a resolução de problemas suficientemente eficaz (característica do comportamento global do agente cognitivo no ambiente), conduzindo à fixação de crenças. Mas é exactamente por essa razão que toda a caracterização da racionalidade é (deve ser) condicional e provisória, feita entre teorias formais e descrições dos métodos realmente utilizados por agentes para resolver problemas no mundo e intuições acerca do que deve ser acreditado em ocasiões particulares[287]. De resto, a interpretação supondo a racionalidade não é uma opção quando se trata de lidar com SI, não há como contorná-la na teoria da cognição ou na tentativa da psicologia de senso comum de fazer sentido de si ou de outrém. É certo que fica assim aberta uma possibilidade de certo modo perturbadora mas já esperada dada a oposição da TSI à TRM que nos tem orientado. A ideia acerca de racionalidade que corresponde à TSI por oposição à computação interna de representações proposta na TRM é a ideia de *adaptação* de sistemas cognitivos ao ambiente, eventualmente acompanhada da ideia segundo a qual a racionalidade não tem verdadeira natureza nenhuma[288], uma vez que esses agentes podem ser vários e diferentes. Ora se a racionalidade é a chave que abre a psicologia também esta não terá uma "verdadeira natureza".

[285] DENNETT 1987: 97.

[286] DENNETT 1987: 98.

[287] Note-se que esta situação é do tipo do equilíbrio reflexivo rawlsiano. Cf. RAWLS 1971 e BRONCANO 1995a. De acordo com Rawls, um juízo ponderado é obtido em equilíbrio reflectido quando existe uma calibração entre os princípios formais ou regras de procedimento e as intuições acerca de casos particulares.

[288] "Verdadeira natureza" no sentido em que se pode afirmar que por exemplo Davidson pensa que a teoria da decisão revela a verdadeira natureza da racionalidade implícita nas práticas comuns (cf. DAVIDSON 1980c). Note-se que Davidson restringe "a verdadeira racionalidade" aos humanos.

2.1.1.7.3 *Aquilo em que uma rã acredita:* Evolution, Error and Intentionality, *a biologia evolucionista e a indeterminação quineana.*

Stich acentua, com razão, que Dennett faz questões do tipo "Que crença tem um sistema S?" dependerem daquilo que uma rã acredita (numa referência ao muito citado caso de *What the Frog's Eye Tells the Frog"s Brain*)[289]. Dennett faz isso porque pensa que aquilo em que uma rã acredita não é mais susceptível de determinação do que aquilo em que o vendedor de limonada acredita quando comete o erro. A comparação entre as crenças de humanos e de rãs é aliás útil sobretudo na medida em que o mito das crenças individuadas guardadas na "caixa das crenças"[290] é menos forte no caso da rã. Uma rã percebe o ambiente e comporta-se e no entanto não lhe são imputadas crenças tão finamente especificadas como as humanas. Que crenças e desejos é que ela tem? Será que, no exemplo clássico, ela procura caçar "moscas-enquanto-moscas", "comida", ou apenas reage a pontos negros móveis? A resposta a estas questões é difícil. No entanto interpretar o comportamento da rã em termos de crenças e desejos é irresistível e indubitavelmente útil. Essa interpretação poderia, no entanto, ser considerada como sendo apenas uma extensão metafórica da noção de "crença", cujo âmbito de aplicação estrito seriam os verdadeiros crentes, os humanos. Mas Dennett pensa que a atribuição de crenças à rã não é uma extensão metafórica, precisamente devido à sua ideia deflacionária quanto ao que é ser um verdadeiro crente. Uma teoria neurofisiológica da rã não falará de crenças e desejos e não fará sentido do seu comportamento, e o que se passa com a neurofisiologia da rã passa-se com a neurofisiologia dos humanos (considere-se por exemplo que se compara a questão «em que é que a rã acredita?» com a questão «qual é exactamente o conteúdo da "crença" perceptual (de um humano) quando uma sombra a baixar no seu campo visual o faz pestanejar?»[291]). Num caso como no outro trata-se de interpretação e não é legítimo afirmar que, estritamente falando, as rãs, *ao contrário dos humanos*, não têm crenças e desejos. Não existe uma linha divisória entre falsos crentes (como a rã) e verdadeiros crentes (os humanos).

[289] Cf. LETTVIN, MATURANA, McCULLOCH & PITTS 1959. Cf. também DUMMETT 1988. Dummett considera tal aproximação absurda.

[290] A expressão foi introduzida por Stephen Schiffer para simplificar as discussões.

[291] DENNETT 1987: 110. Dennett considera que esses estados têm «froggy underspecificity» quando caracterizados a partir da EI.

Perante o realista que pensa que há questões de facto quanto às exactas crenças que causam o comportamento, Dennett convoca repetidamente o exemplo de Quine, o exemplo de Ralph e Ortcutt-o--espião-e-pilar-da-comunidade. Ralph viu várias vezes um homem de chapéu castanho e suspeita que esse homem é um espião. Também conhece um homem de cabelo cinzento, que é um pilar da comunidade e que Ralph só viu uma vez na praia (que ele saiba). Ralph não sabe, mas os "dois homens" são um só. No entanto, existe uma enorme diferença entre caracterizar Ralph como acreditando que o homem de chapéu castanho é um espião e caracterizar Ralph como acreditando que Ortcutt é espião: não se pode simplesmente "especificar" uma crença obtendo a outra, em virtude daquilo que seria a "realidade" (na qual Ortcutt-o-espião e Ortcutt-o-pilar-da-comunidade são o mesmo homem). Não há matéria de facto acerca do que seria a especificação da crença de Ralph. E o que se passa com Ralph passa-se com a rã e com o vendedor de limonada.

A convicta defesa da indeterminação quineana e da inexistência de factos mais profundos (*deeper facts*) acerca de estados internos de SI é o ponto de apoio para a desmontagem da ideia de intencionalidade intrínseca que Dennett pretende levar a cabo no artigo *Evolution, Error and Intentionality*, no qual retoma de novo o exemplo das crenças da rã. A rã não é capaz das distinções de grão fino possibilitadas pela linguagem. Os humanos são capazes dessas distinções. O banho de informação (em que os humanos e a rã estão submersos) é fundamentalmente diferente do banho de palavras. Mas isso não justifica a ideia segundo a qual a psicologia das atitudes proposicionais é a verdadeira psicologia humana, e apenas os humanos são verdadeiros crentes. Para Dennett, o erro do realista reporta-se a um passo pelo qual este diferencia por princípio as descrições intencionais de humanos das descrições intencionais de rãs, tomando como protótipo as crenças linguísticas humanas. Ora, não existem razões para essa distinção de princípio: a rã é tanto um verdadeiro crente como um humano. A atribuição de crenças a rãs é tão idealizante como a atribuição de crenças a humanos. Há tão poucos factos acerca do conteúdo exacto das crenças de Ralph como acerca das crenças da rã.

A exploração do problema das relações entre indeterminação e realismo continua no artigo *Evolution, Error and Intentionality,* onde é analisada a continuidade entre exemplos não biológicos, exemplos biológicos não humanos e exemplos humanos de intencionalidade.

Dennett pretende mostrar que a centração da teoria do conteúdo na maneira humana, linguística, de significar conduz a uma resistência a abordagens evolucionistas da mente e da significação e à distinção injustificada entre intencionalidade intrínseca e intencionalidade atribuída. Em *Evolution, Error and Intentionality* declara que J. Fodor, F. Dretske, T. Burge, S.Kripke, H. Putnam e J.Searle[292] partilham uma crença mais ou menos explícita na diferença de natureza entre intencionalidade genuína e intencionalidade atribuída e que essa crença interfere com o tratamento que fazem da questão do erro. Por isso, embora vários dos autores referidos sejam tentados pelo darwinismo eles rejeitam-no devido a razões imperceptíveis para os próprios.

Contra a distinção entre intencionalidade intrínseca e intencionalidade atribuída, Dennett propõe em *Evolution, Error and Intentionality* uma gradação de casos[293] que são numa primeira análise casos de *misrepresentation*[294] ou identificação errónea mas que podem ser considerados como casos de exaptação[295]. De acordo com Dennett, eles constituem prova da indeterminação pois são precisamente possibilitados pela indeterminação. Os casos, que devem ser pensados em continuidade de modo a desfazer a má intuição quanto à distinção entre intencionalidade intrínseca e intencionaldiade atribuída, são os seguintes:

(1) Uma máquina de venda de refrigerantes, que funciona com *quarters* americanos, mas também com moedas falsas ou de imitação. Uma vez transportada para o Panamá, i.e. "exaptada" para um novo contexto, a máquina virá também a funcionar com *quarters* panamianos, sem qualquer alteração dos estados internos. Dennett sugere que este primeiro caso é análogo ao muito discutido e já atrás mencionado caso nº2, relativo ao que o olho da rã "diz" ao cérebro da rã.

(2) A rã que caça moscas, i.e. o célebre caso de J. Y. Lettvin, H. Maturana, W. McCulloch e W. Pitts no artigo *What the Frog's Eye Tells the Frog's Brain*. O problema é saber o que se pode afirmar que o olho da rã transmite ao cérebro da rã e o que é interpretado erradamen-

[292] Dennett refere-se nomeadamente a DRETSKE 1986, SEARLE 1981, BURGE 1986, FODOR 1987 e KRIPKE 1982.

[293] E retoma em DENNETT 1995: 401, *The Evolution of Meanings.*

[294] O termo, extremamente útil, é introduzido por DRETSKE, cf. DRETSKE 1986.

[295] O termo *exaptação*, sugerido por S.J.Gould e E.Vrba, é utilizado pra nomear casos em que um item que na evolução não surgiu com uma determinada função é posteriormente recrutado com sucesso para cumprir essa função. O termo é por exemplo usado pelos seus autores para considerar coisas que os cérebros humanos fazem para favorecer a nossa sobrevivência (o problema é obviamente saber quais).

te (se é que alguma coisa é *interpretada*, e interpretada *erradamente*) no caso em a rã caça uma imitação de mosca e não uma mosca.

(3) A bactéria de F. Dretske, usada como exemplo em *Misrepresentation* (1986)[296]. Pode-se afirmar que de algum modo a bactéria está "enganada" quando, uma vez transposta para o hemisfério sul, ela se dirige à água tóxica, embora aparentemente não haja qualquer mau funcionamento do dispositivo sensorial sensível ao campo magnético que a orienta, cujo funcionamento no hemisfério norte fazia com que ela se afastasse da água tóxica. Dretske procura nesse artigo a pedra de toque da existência de "crença natural", e encontra-a na possibilidade de crer ou representar *erradamente (misrepresent)*. Dretske não quer inflacionar anti-naturalmente as funções do sistema, nem dotar de demasiado conteúdo ou de conteúdo demasiado específico a função do sistema. É certo que Dretske não pode, por outro lado, des-inflacionar demasiado, pois defende precisamente que, por contraste, meros indicadores ou signos naturais são incapazes de representar erroneamente (*misrepresent*).

(4) Os casos da família da Terra Gémea de Putnam[297]. O problema é saber o que, exactamente, se pode dizer que se passa quando um terrestre, um vez transposto para a Terra Gémea, identifica erroneamente por exemplo um *schmorse* da Terra Gémea como *"horse"*, ou XYZ da Terra Gémea como "água".

Nos quatro casos trata-se, segundo Dennett, da exaptação de um SI e da indeterminação da sua função ou do conteúdo dos seus estados. Só é possível afirmar que existe um erro ou um mau funcionamento determinados se se suposer que é possível olhar para a situação de um ponto de vista privilegiado, o ponto de vista do engenheiro criador do dispositivo ou de quem pratica a introspecção, presumindo saber "a partir de dentro" o que alguma coisa realmente tem como função ou realmente significa. Se se admitir que o estado interior realizado é o mesmo, o problema consiste em saber quando, exactamente, é possível afirmar que a função ou o significado dos estados mudaram. Dennett crê que nada de intrínseco, descritível de forma funcional em relação aos estados inter-

[296] Este texto de F. Dretske fez história. Nele, Dretske analisa sistemas biológicos cujos poderes de representação não derivam de outra fonte (que não eles próprios) e que podem errar. A "crença" é definida como a capacidade representacional em virtude de cujo exercício pode ser gerada uma representação errónea. O erro é assim colocado propositadamente no centro da discussão da teoria do conteúdo. A teoria do conteúdo elaborada por F. Dretske será discutida mais à frente neste capítulo.

[297] PUTNAM 1975.

nos dos sistemas, distingue os estados anterior e posterior à exaptação. Isto vale para os quatro casos. Não pode portanto por princípio existir uma resposta determinada à questão quanto ao momento da "mudança" de função ou significação.

No caso 1, como se trata de um artefacto mecânico e da sua função, e portanto parece tratar-se de um caso claro de intencionalidade derivada, é aparentemente natural afirmar que um tal sistema só poderia ter uma função atribuída pelo seu criador. O que Dennett faz ver é que novas funções não "intencionadas a priori" e bem sucedidas podem surgir, por meio da exaptação, mesmo para um artefacto desse género. Os filósofos interessam-se geralmente mais pela Terra Gémea e pelos seus visitantes do que por máquinas de vender refrigerantes. Dennett pensa que a conclusão a extrair de ambos os casos é a mesma: uma intuição externalista acerca de funções e significados. O que a análise de artefactos não biológicos (ou biológicos) mostra é que a indeterminação é uma pré-condição da possibilidade natural das funções. O significado é uma função da função e a indeterminação da função prolonga-se na indeterminação da possibilidade natural de significar. A perspectiva anti-essencialista e darwinista conduz a concluir que «assim como a função biológica, a significação não é qualquer coisa de determinado na sua origem»[298].

Só seria possivel sustentar a posição contrária em virtude de um injustificado essencialismo e de um desejo de *skyhooks* ("anzóis vindos de cima", na expressão que Dennett utiliza em *Darwin's Dangerous Idea*, e que alude a causas finais reais, à ideação da função e da significação previamente ao surgimento destas). É tão errado supor que alguma coisa fixa o que os termos significam no caso dos humanos na Terra Gémea como é errado supor que existe uma fixação apriorista da função e da significação nos casos 1, 2 e 3. Os significados e funções nos/dos humanos são tão derivados como os significados e funções nos/dos artefactos: «uma lista de compras na cabeça, no cérebro, não tem mais intencionalidade intrínseca do que a mesma lista de compras num pedaço de papel»[299]. A não ser que se presuma um acesso privilegiado a algo que seja "significar realmente", não se justifica uma distinção de princípio entre intencionalidade intrínseca e intencionalidade atribuída. No entanto, es-

[298] DENNETT 1995: 408.

[299] DENNETT 1987: 318.

sa suposição é a pedra de toque do Racionalismo do Significado (*meaning rationalism*)[300] associado ao essencialismo.

Dennett sugere mesmo, em *Evolution Error and Intentionality,* que se descobre a "Grande Divisão" entre posições teóricas quanto ao conteúdo descobrindo a atitude das pessoas quanto à interpretação de artefactos, procurando verificar se elas concordam com a separação entre intencionalidade intrínseca e intencionalidade atribuída (reservando obviamente esta última para os artefactos). E é isso que o próprio Dennett faz em relação a Dretske, Burge, Kripke, e Fodor. De modo a conduzir a uma redução ao absurdo a tese do campo adversário (i.e. a defesa da distinção entre intencionalidade intrínseca e intencionalidade atribuída) Dennett utiliza ainda como exemplo (em *The Intentional Stance* e em *Darwin's Dangerous Idea)* o caso uma máquina de sobrevivência análoga aos humanos, considerados enquanto máquinas de sobrevivência para genes, na expressão de Richard Dawkins[301]. Se um organismo humano é considerado uma máquina de sobrevivência para genes não se pode considerar que ele seja o seu próprio "intencionador" (a haver algum intencionador seriam os genes, cuja "intenção" a existência de organismos realiza). Ora, os humanos são, no entanto, o caso paradigmático de intencionalidade intrínseca. Assim, intencionalidade derivada – a intencionalidade humana é "derivada" da "intenção dos genes" – pode ter ou ser intencionalidade intrínseca.

Dennett sugere também a análise de um caso imaginário em que um humano decide experimentar a vida no século XXV e a única maneira de manter o corpo vivo é colocá-lo em hibernação durante um tempo determinado, durante o qual o corpo teria que ser guardado e protegido. Precisa-se de um sistema que assuma a função durante quatro séculos. Se se construísse um robô que soubesse procu-

[300] Esta é uma expressão de Ruth Millikan, que Dennett frequentemente retoma para nomear o erro por excelência no pensamento acerca da intencionalidade. O Racionalismo do Significado é uma posição apriorista acerca da natureza da significação. De acordo com o Racionalismo do Significado os produtores e intérpretes de Ícones Intencionais (o termo que Millikan utiliza para os veículos de significação, cf. MILLIKAN 1984) detêm uma autoridade especial quanto ao significado destes. Um sistema cognitivo deveria assim ser por princípio capaz de reconhecer a priori o que as suas representações significam, por exemplo saber se duas representações A e B que ocorrem em momentos diferentes representam o mesmo representado. Assim, a mesmidade de conteúdo de representações A e B seria necessariamente transparente ao pensador dos pensamentos A e B. Tanto Millikan como Dennett pensam que esta posição é profundamente equivocada, essencialmente porque, como Putnam viu (PUTNAM [1975]: 227) «*cut the pie any way you like, meanings just ain't in the head*».

[301] DAWKINS [1976].

rar, reconhecer, antecipar, evitar perigos, lidar com outros seres no ambiente, ele faria tudo isso "em função" do hibernador, que alberga dentro de si, mas o facto é que disporia de auto-controlo e de auto-monitorização, etc. Pode-se até imaginar que o robô se tornaria suicida, se se persuadisse da maior importância de algum projecto em relação àquele que lhe dera origem. A história não é senão uma versão da imagem de Dawkins das espécies biológicas como máquinas de sobrevivência para genes: aquilo que estamos dispostos a afirmar relativamente ao sistema é análogo àquilo que devemos dizer de nós próprios, de um ponto de vista biológico. De onde se conclui que a intencionalidade não é classificável em dois tipos estanques, intencionalidade intrínseca e intencionalidade atribuída. A forma correcta de colocar a questão é considerar que quando a atribuição de intencionalidade é justificada, ela captura padrões reais[302].

[302] Em *Darwin's Dangerous Idea* (Dennett 1995: 412) e no contexto do mesmo problema Dennett propõe uma experiência imaginária para exemplificar este ponto: nessa experiência há duas caixas, A e B, duas caixas negras ligadas por um fio de cobre. Há dois botões, a e b, na caixa A e três luzes, vermelha, verde e cor de âmbar na caixa B. As caixas são encontradas por cientistas e começam as tentativas de exploração: quando se pressiona o botão a na caixa A, acende-se a luz vermelha na caixa B, quando se pressiona o botão b na caixa A acende-se a luz verde na caixa B. A luz cor de âmbar parece nunca acender. A situação repete-se milhões de vezes. Parece ter que existir uma regularidade causal (algo que se passa no fio, pois se este for cortado, o efeito não se verifica), uma vez que *Todos os a's causam luz vermelha* e *Todos os b's causam luz verde*. No entanto os cientistas que examinam as caixas vêm a descobrir que no interior estão computadores e que aquilo que passa pelo fio de cobre são cadeias de bits, com uma determinada dimensão, diferentes de cada vez. Surpreendentemente, quando os cientistas resolvem interferir, alterando as cadeias que passam de A para B, a luz cor de âmbar passa a acender quase sempre. Alguns milhões de tentativa depois, constata-se que essa é a situação mais frequente. Nesse momento, os cientistas não sabem de todo como prever se uma cadeia proveniente de A vai fazer acender a luz vermelha, a luz verde ou a luz cor de âmbar. Aparecem então os autores das caixas, que explicam que tinham estado a trabalhar com bases de dados contendo proposições verdadeiras sobre tudo o que existe. Aborrecidos com o trabalho, resolveram fazer um brinquedo filosófico. Escolheram uma língua franca para traduzir entre os dois sistemas de representação (diferentes) em que trabalhavam, e ligaram os sistemas. Quando o botão a da caixa A é pressionado, é seleccionada aleatoriamente uma crença na caixa, traduzida para a língua franca e enviada para B, que a traduz para o seu sistema e a contrasta com a sua base de dados. Se B considera a frase verdadeira, acende-se a luz vermelha, se a considera falsa, acende-se a luz verde. Quando se interfere no processo, produzem-se cadeias mal formadas (que são em muito maior número que as cadeias bem formadas) dando lugar à luz âmbar em B. Segundo os autores das caixas, a propriedade emergente *vermelho* é "ser uma frase verdadeira na língua franca" e a propriedade *verde* é "ser uma frase falsa". Assim, o problema, que tinha confundido os cientistas durante anos, revela-se elementar. Para Dennett, a versão dos autores das caixas, quanto a propriedades emergentes, não é admissível: existem verdades na língua franca que excedem a dimensão da cadeia transmitida, há factos nas bases de dados que vieram a revelar-se falsos, em suma, existem números razões pelas quais a propriedade "vermelho" não é exactamente a propriedade "verdadeiro na língua franca". A versão dos autores é uma útil (e inabdicável) descrição *como-se* (e os filósofos na história de Dennett apressaram-se a comentar que a regularidade não é causal). Dennett considera este exemplo um aperfeiçoamento do exemplo dos marcianos de *True Believers* atrás analisado.

No caso da biologia evolucionista, a tese de Dennett acerca do estatuto da atribuição ou interpretação traduz-se na ideia segundo a qual não é possível descobrir regularidades sem apelo a razões, a padrões reconhecidos a partir de um ponto de vista. A identificação de razões não é, assim, neutra, e Dennett chega a considerar que o adaptacionismo biológio envolve uma "interpretação retrospectiva radical": tanto quando se trata de *design* como quando se trata de significação estão em jogo princípios interpretativos (logo idealizantes) incontornáveis. Quer se trate de mentalidade quer de funções, Dennett defende que, se se trata de encontrar padrões, simplesmente não existe substituto para a EI.

2.1.1.7.4 *Dennett e o Quarto Chinês ou quando outros filósofos atacam a separação entre conteúdo e consciência.*

Se o pomo da discórdia localizado em *Evolution, Error and Intentionality* é a distinção entre intencionalidade intrínseca e intencionalidade atribuída, a experiência mental do Quarto Chinês de John Searle é a situação em que tal distinção é exemplarmente posta em cena[303]. Searle pretende provar com a experiência mental do Quarto Chinês uma diferença de natureza entre intencionalidade genuína e intencionalidade atribuída. Se Searle é bem sucedido, toda a abordagem dennettiana fica invalidada. É por isso fundamental para Dennett mostrar que as intuições que Quarto Chinês convoca devem ser rejeitadas.

Dennett escreveu vários textos nos quais pretendeu desmontar a situação do Quarto Chinês. Atribui, além disso, a D. Hofstadter uma

[303] *A experiência mental do Quarto Chinês* consiste no seguinte. Alguém, que não fala chinês e que pode ser o próprio J. Searle, está fechado dentro de um quarto onde existem símbolos chineses em caixas. Essa pessoa tem um livro de instruções em inglês (a língua que fala) que explica como combinar os símbolos chineses e como enviar sequências de símbolos para fora do quarto quando são introduzidos no quarto outros símbolos chineses, através de uma pequena janela. A pessoa que está dentro do quarto não sabe nada disso mas as pessoas que estão fora do quarto chamam perguntas aos símbolos que introduzem e respostas aos símbolos que saem. Na perspectiva das pessoas que estão cá fora, o sistema fala, portanto, chinês. No entanto a pessoa que está lá dentro sabe que não percebe uma palavra de chinês. Segundo Searle a experiência mostra como um sistema pode ter intencionalidade atribuída sem ter intencionalidade genuína. Dennett considera no entanto que «Douglas Hofstadter and I first exposed the cute tricks that made the Chinese Room work» (DENNETT 1997: 117). Cf. também DENNETT 1991: 436 «The definitive refutation, still never adequately responded to by Searle, is Douglas Hofstadter's, in Hofstadter & Dennett 1981». Dennett refere-se a HOFSTDTER 1981a.

objecção definitiva à experiência mental de Searle. A objecção comum de Dennett e de Hofstadter é, basicamente, aquela que é conhecida como a resposta dos sistemas, a resposta que os teóricos da IA deram a Searle. De acordo com essa resposta, no caso apresentado por Searle o *entendimento só pode ser imputado à totalidade do sistema* – que inclui as instruções a serem levadas a cabo – e não ao "agente-Searle" no interior. A parábola de Searle só é eficaz devido a uma voluntária ocultação de questões de escala e ao apelo à identificação com o simulador animado. Searle passa por cima da diferença de complexidade entre níveis, e oculta voluntariamente o facto de a vida mental consciente dos humanos acontecer também a partir de actividades neuronais sem mente e sem entendimento.

D.Hofstadter situa a experiência imaginária de Searle como uma posição numa família de experiências de pensamento cujo propósito é explorar cenários de simulação de vida mental[304]. Cada experiência de pensamento faz variar as seguintes cinco características: (1) material físico da simulação, (2) nível de detalhe da simulação, (3) dimensão da simulação, (4) dimensão e natureza do demónio que leva a cabo a simulação e (5) velocidade a que o demónio trabalha. Apontados estes parâmetros, torna-se claro o parentesco da experiência de Searle com a experiência – imaginada por D.Hofstadter[305] – da conversa com o cérebro de Einstein, com a experiência da substituição dos neurónios de Z. Pylyshyn e com a experiência de J. Haugeland da intervenção de um demónio na condução de neurotransmissores no cérebro de uma mulher[306]. A *intuição* é o verdadeiro ponto de apoio do Quarto Chinês e o que se verifica, uma vez seguidas diferentes variações imaginárias, é que as intuições acerca da intencionalidade genuína e do nível a que ela se situa não são tão tenazes como Searle desejaria: experiências de pensamento "da mesma família" convocam intuições totalmente diferentes. Em suma, *intuições acerca de intencionalidade genuína pura e simplesmente não são confiáveis.*

Antes de criticar as premissas do argumento, Dennett nota que o Quarto Chinês de Searle não é exactamente um argumento mas antes uma parábola e um apelo à intuição. Searle admite esse facto[307], e posteriormente dá a forma de argumento à sua parábola. As

[304] HOFSTADTER 1981 a.

[305] HOFSTADTER 1981a: 430.

[306] Cf. as respostas a SEARLE 1980 no mesmo número da revista *Behavioral and Brain Sciences* em que o artigo foi publicado.

[307] DENNETT 1987. Cf. SEARLE 1992 e SEARLE 1997.

premissas seriam as seguintes: 1) Os programa são puramente formais/sintáticos; 2) A sintaxe não é equivalente, ou suficiente para, a semântica; 3) As mentes têm conteúdos semânticos. A conclusão seria que (1) implementar um programa não é suficiente para se ter uma mente e logo que (2) a IA Forte (a ideia de acordo com a qual qualquer sistema que implemente o programa correcto terá mente e consciência não sendo apenas uma simulação) é impossível.

É em torno destas premissas e conclusões que a crítica de Dennett se move. De acordo com Dennett, as próprias premissas são falsas se lhes fôr dada a interpretação necessária para o argumento funcionar. Quanto à primeira premissa, num certo sentido é claro que um programa é apenas sintaxe e por si não produz nada, se por tal se entender um programa não implementado, linhas de comandos escritas num papel, por exemplo. No entanto, apenas programas *implementados* merecem consideração nesta discussão: se Searle pretende afirmar alguma coisa relevante, só pode querer dizer que os poderes causais cruciais para a produção de mentalidade nada têm a ver com programas implementados. Ora, em relação a programas implementados não é correcto afirmar que eles são *apenas* sintáticos, pois eles têm efeitos causais.

Searle afirma que o núcleo do seu caso é que apenas cérebros humanos, e não programas, têm o poder de causar intencionalidade genuína e declara ser mal interpretado quando é lido como afirmando que *apenas cérebros orgânicos* podem causar intencionalidade. Dennett sugere em *Fast Thinking*[308] que se se pretender especificar os misteriosos poderes dos cérebros, uma questão pertinente é a velocidade, que é importante para a inteligência de agentes físicos. Nesse quadro, é razoável pretender que nada a não ser um sistema maciçamente paralelo poderia aproximar-se dos tempos da inteligência humana. Mas esta é uma questão de tecnologia e de progresso, e Dennett admite que poderia dar-se o caso de só se atingir a velocidade requerida com computação orgânica. Neste caso, não se trataria de facto de uma questão relativa ao programa mas ao *hardware*, e então seriam realmente os cérebros ou outro *hardware* e não os programas que teriam o poder de causar a intencionalidade. Mas não era esse o ponto de Searle, que sempre afirmou que o seu argumento diz respeito a princípios conceptuais e não a um estádio da evolução da tecnologia. O facto de os computadores digitais não

[308] DENNETT 1987 l.

serem suficientemente rápidos seria uma razão não misteriosa pela qual programas não teriam o poder de causar intencionalidade. Simplesmente, esse não seria já um argumento contra a IA forte (que pretende que um computador devidamente programado tem literalmente uma mente, *seja qual for o hardware*).

Finalmente, Dennett pensa que a conclusão do argumento de Searle – segundo a qual "Ter um programa não é suficiente para, nem equivalente a, ter uma mente" – é susceptível de duas interpretações e como tal ambígua. De acordo com uma interpretação, nenhum programa seria suficiente para produzir os fenómenos mentais que um cérebro orgânico com os seus poderes causais produz. De acordo com outra interpretação, não há maneira de programar um computador digital de modo a produzir o controlo da actividade intencional inteligente e rápida que os humanos exibem. Dennett defende que a primeira interpretação é incoerente e que a segunda, embora seja uma pretensão empírica plausível, não é o ponto que Searle quer fazer com o argumento.Tudo isto revela segundo Dennett que a verdadeira preocupação de Searle na sua experiência mental é a consciência subjectiva, o "sentir-se ser", e não a intencionalidade ou a semântica, embora seja essa a questão nomeada. Em textos posteriores de Searle isso tornar-se-á ainda mais claro. O problema de Searle não é o carácter formal dos programas: o que Searle não aceita é que semântica genuína possa existir sem consciência. Nos termos até aqui utilizados, Searle não aceita que a teoria do conteúdo possa ser independente da teoria da consciência. É a esta convicção, em apoio da qual Searle convoca a intuitiva identificação da pessoa que lê e compreende a experiência mental com o manipulador de símbolos dentro do quarto, que autores como Dennett e Fodor se opõem.

2.1.1.8 *Real Patterns.*

Terminar-se-á o percurso pela teoria do conteúdo com a análise de *Real Patterns*, um ensaio sobre ontologia que Dennett considera central no seu pensamento, e no qual continua a defender o realismo enfraquecido quanto à intencionalidade. A ideia básica continua a ser que a descrição e a previsão intencionais dependem da exploração de padrões no mundo, discerníveis no comportamento de entidades. Alguns dos exemplos de *Real Patterns*, que têm como finalidade mos-

trar a realidade dos estados intencionais são os exemplos (1) dos centros de gravidade, (2) dos padrões de pixels, mas sobretudo (3) dos padrões no Jogo da Vida (o exemplo é retomado de *True Believers*). Segundo Dennett, estados intencionais são igualmente padrões reconhecíveis no comportamento de agentes. J. Haugeland faz notar que Dennett oscila, em *Real Patterns*, entre a consideração dos estados intencionais como elementos de padrões e como padrões. Para Haugeland, apenas a primeira alternativa é pertinente (e apoiada aliás pela aproximação da EI à tradução radical de Quine e à interpretação radical de Davidson).

Um primeiro sentido da noção de padrão diz respeito a informação, medida em bits. A questão é a eficiência da transmissão dessa informação. A maneira mais ineficiente de transmitir a informação por exemplo relativa a uma disposição espacial (o exemplo apresentado em *Real Patterns* são padrões visuais constituídos por pontos, a que Dennett chama códigos de barra) é um mapa de bits, que descreve uma disposição ponto por ponto e que é portanto tão capaz de "dar" vários padrões como a total ausência de padrões. Uma série é aleatória (no sentido matemático, na definição de G. Chaitin evocada por Dennett em *Real Patterns*) se é incomprimível, só podendo ser transmitida por um mapa de bits. Ora algumas das disposições são susceptíveis de descrições mais curtas do que as fornecidas pelo mapa de bits e é precisamente de séries não aleatórias e descritíveis de modo mais eficaz do que por um mapa de bits que pode afirmar-se que têm padrões.

Um outro sentido de padrão envolve a noção de *reconhecimento*. Considerar-se-á aqui o caso dos padrões do Mundo da Vida[309]. No Mundo da Vida, aparentemente apenas uma variação cinética dos padrões de pixels, há padrões persistentes (tais como deslizadores (*gliders*), devoradores (*eaters*), etc) cujos constituintes mantêm relações entre si. Essa persistência pode ser vista *a partir de baixo*, da física do mundo, ou a partir daquilo que torna os padrões notáveis, nomeadamente um tipo de saliência que depende da previsibilidade do seu comportamento, i.e. *a partir de cima*. É evidentemente a possibilidade de *reconhecimento a partir de cima* que interessa Dennett.

[309] Não considerarei aqui um dos mais conhecidos exemplos do artigo, o exemplo dos centros de gravidade. Sigo J. Haugeland, que considera a aproximação enganadora pois os centros de gravidade não têm nenhum estatuto ontológico especial. Cf. HAUGELAND 1993: 55. Cf. RUDDER BAKER 1994 para uma opinião diferente.

Padrões assim são candidatos ao reconhecimento, estão ligados à existência de observadores e perspectivas. Este é portanto o problema: Dennett utiliza em *Real Patterns* duas noções de padrão, uma noção matemática, de acordo com a qual um padrão é algo que existe em séries não aleatórias e uma noção de padrão como candidato ao reconhecimento. A primeira definição – a primeira a ser evocada no artigo – pode ser considerada independente de observadores, a segunda evidencia a relatividade dos padrões às perspectivas: um padrão reconhecido por uma entidade pode não ser reconhecível por outra entidade (Haugeland nota que reconhecimento não é apenas resposta: reconhecimento é uma noção normativa, já que é possível reconhecer *erradamente*[310]). Mais: a ideia de reconhecimento não pressupõe nada acerca da natureza dos elementos do padrão. Esta ausência de pré-determinação dos elementos impede a própria questão da redução do padrão aos elementos, o que mesmo não acontece com a definição matemática, mais precisa acerca da natureza dos elementos.

Ora, padrões-como-candidatos-ao-reconhecimento adaptam-se melhor do que padrões-como-arranjos-ordenados aos fenómenos em causa na teoria da mente. Na sua interpretação de *Real Patterns*, Haugeland sugere que Dennett deveria juntar as duas noções de padrão, considerando que se aplicam a níveis diferentes dos padrões. Os elementos do arranjo, inabdicáveis na definição matemática, não têm que ser determinados na definição operacional de padrão como algo reconhecível. Neste último caso, basta que existam criaturas que reconheçam o padrão: ele não tem que ser, necessariamente, um padrão constituído por *elementos* determinados (no exemplo de Haugeland, os humanos não reconhecem sorrisos pelo facto de os sorrisos serem composições de elementos previamente determinados). A noção de padrão como "candidato ao reconhecimento" evita a tentação do reducionismo, enquanto a definição matematica é precisamente um convite ao reducionismo.

O que se joga na discussão das definições de "padrão" é a possibilidade de fundamentar objectos, que sirvam de critério para a correcção do reconhecimento. É exactamente para exemplificar a fundamentação de objectos por meio da noção de padrão real que Dennett recorre ao Mundo da Vida: são possíveis várias perspectivas

[310] HAUGELAND 1993: 57.

sobre o que neste acontece e no entanto, existe uma objectividade dos padrões. Haugeland estabelece uma analogia entre essa objectividade ligada à obediência a "regras" e a natureza de um jogo de xadrez e das respectivas peças. Em ambos os casos existe algo de especial descritível a partir de cima, que constitui um padrão. Considerando as peças de xadrez, de que tipo de realidade se está a falar se elas são definidas pela sua participação num jogo? Que tipo de realidade têm uma torre ou um bispo, que não são por exemplo necessariamente figuras tridimensionais de marfim ou de madeira, nem imagens num écran de computador? Como nota Haugeland, noções metafísicas de substância e propriedade não se aplicam a realidades tais, não reconhecíveis fora de contexto. A realidade de que se fala é uma realidade intermédia, dependente do contexto, que deve ser considerada no seio de "maneiras de ser" representadas pelas estratégias alternativas (a EF, a ED, a EI). Se é assim, o nível intencional da realidade é objectivo e não menos real do que o físico.

Dennett confessa que não pretendia dar à noção de reconhecimento a importância teórica que Haugeland lhe atribui: simplesmente a tomava no sentido comum na IA e na ciência cognitiva em geral. Mas é em última análise da ligação entre as noções de reconhecimento e de padrão que o realismo moderado dennetiano acerca da intencionalidade depende. Esta questão será retomada no Capítulo 6, quando se explicitar a ontologia subjacente à TSI. Na segunda parte deste capítulo procurar-se-á verificar a contraparte do realismo moderado quando se trata dos problemas da explicação psicológica e da individuação do conteúdo, tendo em mente que a noção de padrão serve sobretudo para evidenciar que tipos de materialismo como o realismo intencional de Fodor por um lado e o eliminativismo dos Churchland por outro são posições excessivamente fortes acerca da intencionalidade. No caso humano, os estados intencionais são remetidos à biologia, e portanto ao *design*, ou, em geral à organização funcional de sistemas cognitivos em cuja operação seja possível o reconhecimento interpretativo de padrões. O problema do *design* torna-se assim uma questão incontornável e Dennett sentiu inclusive necessidade de lhe dedicar toda uma obra, *Darwin's Dangerous Idea*.

O problema é saber se Dennett, remetendo ao *design* a teoria dos padrões e do reconhecimento pode dar-se ao luxo de continuar a declarar-se um realista *apenas moderado*. Dennett termina o artigo *Evolution, Error and Intentionality* afirmando que o tratamento da in-

tencionalidade conduz inevitavelmente a um acerto de contas com as perturbadoras consequências da destruição do Argumento do Desígnio levada a cabo por Darwin: não se pode ser realista acerca de significação sem realismo acerca de funções, e este último teria sido impossibilitado por Darwin. No entanto, o problema da posição de Dennett, por mais que ele reclame defender um tipo moderado de realismo, é que ela tende irresistivelmente para extremos opostos: por um lado para um eliminativismo total, um quadro de interpretação totalizante em que a intencionalidade finalmente não existe, a racionalidade finalmente não existe, toda a mentalidade é finalmente relacional, dependente de uma atribuição e não intrínseca, e no qual até mesmo a EF deveria ser considerada como uma questão de interpretação. Por outro lado, remeter o problema das representações mentais para o problema do *design* parece ser um apelo a uma teleologia real. É da acusação feita pelos críticos segundo a qual desta posição decorreriam implicações contraditórias, sendo portanto o realismo moderado instável, que Dennett tem que defender-se.

2.2 *A individuação do conteúdo e a explicação psicológica. Contra o instrumentalismo da TSI: R. Millikan, F. Dretske e J. Fodor.*

A posição de Dennett quanto à *natureza* de "entidades" representacionais – o realismo moderado – deve ser considerada em conjunto com uma posição quanto ao *conteúdo* das representações. Ora, a posição quanto ao conteúdo das representações que Dennett defende desde C&C é uma posição teleofuncionalista. Apenas a partir da conjugação do realismo moderado com o teleofuncionalismo se esclarece a forma como a TSI enfrenta as questões da explicação psicológica e da individuação das representações mentais.

2.2.1 *Funcionalismo e teleofuncionalismo: a reintrodução das funções biológicas na teoria da mente. Realismo acerca de representações. A TSI e o teleofuncionalismo. Problemas do externalismo.*

A noção de função suposta pelo teleofuncionalismo não se identifica com a noção de função do funcionalismo *tout court*. Esta última visa acentuar a autonomia da descrição de um dado nível de funcio-

namento de um sistema, considerando o sistema por si, enquanto a noção teleofuncionalista de função, identificável com a noção de biológica de função como *propósito ou finalidade*, diz respeito não apenas ao sistema considerado por si mas também à adaptação do comportamento deste ao ambiente. A convicção básica dos teleofuncionalistas é que apenas a reintrodução da noção biológica de função na teoria da mente permite dar conta do conteúdo das representações. Note-se que quando se aborda o problema do conteúdo das representações se trata de saber não o que faz de um conjunto de estados mentais *estados mentais*, mas o que os diferencia entre si (individua) enquanto estados mentais. Estado mentais acerca de A (gatos) e estados mentais acerca de B (estrelas) são individuados através desses conteúdos, mesmo se são definíveis identicamente enquanto estados mentais (por exemplo, como estados funcionais do cérebro). Eles diferem no entanto na medida em que um é acerca de gatos e o outro acerca de estrelas. Segundo o teleofuncionalismo, esta distinção faz-se através do apelo aos *propósitos* ou *finalidades* cumpridos pelos estados mentais do exemplo quando estes "apontam ou indicam" respectivamente gatos e estrelas. Os estados mentais a que aqui se alude são *veículos*. A partir deste momento convém ter constantemente presentes dois sentidos de "representação": por um lado, falar-se-á de representações enquanto *veículo*s (esta é uma noção "mecânica", utilizada na ciência cognitiva com o sentido de "estrutura do cérebro ou de outro *hardware* que de uma forma ou outra segue alguma coisa no mundo, nomeadamente co-variando com ela"), por outro lado falar-se-á de representações enquanto *aquilo que é representado*. Note-se que sendo o propósito dos estados mentais (enquanto veículos) indicar algo que está fora do sistema, aquilo que determina ou fixa o conteúdo ou significado (aquilo que é representado) não estará "dentro da cabeça" mas sim no mundo. A ideia básica dos teleofuncionalistas é que o *design* dos sistemas cognitivos é tal que estes produzem determinadas representações internas quando em presença de determinadas condições no mundo, representações essas que indicam as condições no mundo. Em geral o *design* dos sistemas é atribuído à selecção natural, que teria produzido soluções evolutivamente estáveis para o problema da coordenação do comportamento de sistemas com o ambiente[311]. Deste modo, de acordo com o teleofuncionalismo, a maneira humana de significar, nomeadamente a maneira linguística

[311] Ou também à aprendizagem como processo análogo à selecção natural.

de significar, não é de modo algum prototípica: significar pode perfeitamente ser um funcionamento pré-verbal ou a-verbal em sistemas cognitivos. Segundo Dennett, o facto de os teóricos da mente centrarem frequentemente a atenção na significação linguística, analisando-a através de meios lógicos, conduz a distorções e a resistências infundadas às abordagens naturalistas do significado. O teleofuncionalismo, pelo contrário, evocando funções biológicas e selecção natural é uma teoria totalmente naturalista do conteúdo das representações, que tem a grande vantagem de tratar o problema da normatividade da cognição (o facto de ser constitutivo de uma representação ela ser suposta ser acerca de x e não de y ou de z) remetendo-o para normas biológicas (i.e. para o desenho de dispositivos para cumprirem funções e para o normal funcionamento destes). De acordo com estes parâmetros, o grande problema para as teorias teleofuncionalistas é, como se verá, estabelecer a natureza do erro ou má representação (*misrepresentation*).

É preciso notar que nem todos os autores que defendem uma posição funcionalista quanto à natureza da representação defendem uma posição teleofuncionalista quanto ao conteúdo dos estados mentais. Por exemplo J. Fodor, que tem vindo a ser considerado o caso exemplar do teórico funcionalista, defende uma teoria nómico-informacional do conteúdo[312] e não uma teoria teleofuncionalista. O núcleo do teleofuncionalismo é, repita-se, a ideia segundo a qual a individuação do conteúdo de estados mentais se apoia no carácter apropriado (ao ambiente) do comportamento (guiado pela produção de representações) de sistemas físicos. A explicação psicológica dirá respeito a entidades internas, veículos seleccionados com a finalidade de representar. Ora, na medida em que Dennett defende que a adscrição do conteúdo é holista e interpretativa, quer a possibilidade de representações internas quer a possibilidade da explicação psicológica (que pressupõe que alguma coisa de natureza representacional no interior do sistema faz diferença no comportamento deste) se afiguram problemáticas. Noutras palavras, a TSI arrisca-se a apenas poder considerar o conteúdo de estados mentais como «epifenomenal» e «impotente»[313], não podendo nunca influenciar a, ou sobrepôr-se à, causação mecanicista local. Não é no entanto apenas a TSI que tem problemas com a impotência causal do conteúdo: a explicação psicológica e a in-

[312] Cf. FODOR 1990, FODOR 1994, FODOR 1998.

[313] DENNETT 1998b: 63. É nisto que se traduz, afinal, a impossibilidade de máquinas semânticas.

dividuação do conteúdo constituem problemas para o teleofuncionalismo em geral já que este considera como pedra de toque da questão do conteúdo das representações a apropriação do comportamento do sistema global, sendo sempre alguma coisa que está fora da cabeça (as condições no mundo) que é responsável pela fixação do conteúdo dos estados mentais que guiam esse comportamento.

As duas questões a partir de agora tratadas – a explicação psicológica e a individuação do conteúdo – estão, como se procurará mostrar, relacionadas. A grande opção que subjaz às teorias da explicação psicológica e da individuação do conteúdo desenvolvidas pelos diferentes autores convocados é uma opção quanto à realidade a atribuir às representações mentais ao nível do *design*. Ao relacionar a questão do conteúdo com a função biológica, o teleofuncionalismo, sendo tendencialmente realista acerca de funções, é tendencialmente realista acerca de representações (enquanto veículos). Dispositivos biológicos com funções (por exemplo orgãos como olhos, corações, etc) são dispositivos físicos reais, com funções que relacionam sistema e ambiente. Assim, a maioria dos autores que como Dennett desenvolvem teorias teleofuncionalistas e evolucionistas do contéudo, nomeadamente Ruth Millikan e Fred Dretske, são realistas acerca de representações internas e das suas funções. Millikan e Dretske consideram que é possível falar de veículos de conteúdo ou de informação que fazem diferença no sistema[314] e que a explicação psicológica se baseia nestes veículos. Assim, para F. Dretske a explicação psicológica terá a forma de uma explicação do comportamento através de *razões*, no caso de Millikan, a explicação psicológica envolverá normas e história[315]. Também J. Fodor, em-

[314] Sejam Ícones Intencionais (MILLIKAN 1984) alguns dos quais serão representações no pleno sentido da palavra (*full blown*), sejam indicadores internos (DRETSKE 1988). Millikan define os Ícones Intencionais como dispositivos com Funções Próprias (a noção de Função Própria será explicada a seguir) cuja produção e consumo servem para guiar o comportamento de sistemas. Dretske define indicadores internos como estruturas internas que transportam (*carry*) informação e que são selectivamente sensíveis ao que acontece no ambiente. Em ambos os casos de trata de veículos internos do conteúdo, seleccionados para representar algo de específico no exterior do sistema.

[315] De acordo com Millikan, uma explicação psicológica é uma *explicação histórica*, o que significa que as as funções psicológicas, como quaisquer funções biológicas são definidas pela história e pelas condições Normais. Defender que as funções são definidas pela história é negar que elas sejam definidas por disposições actuais. Esta é a raiz do diferendo entre R. Millikan e J.Fodor. Fodor admite perfeitamente que as correlações entre estados mentais e o ambiente, que explicam o significado, possam ser resultado de um processo darwinista mas não considera (ao contrário de Millikan) que uma explicação evolucionista dos mecanismos da representação seja explicação suficiente do seu funcionamento *presente*.

bora não sendo um teleofuncionalista, considera possível uma individuação das representações internas e procura conjugá-la com uma teoria da explicação psicológica. De acordo com Fodor, a teoria do conteúdo das representações deve ser externalista, informacional e lata. O esquema explicativo global de Fodor procura assim conjugar a intencionalidade das leis psicológicas, o carácter computacional dos pensamentos e a semântica informacional[316]. Neste âmbito, a ideia de conteúdo estrito a que atrás se fez alusão transformar-se-á na ideia de Modos de Apresentação (representações mentais com conteúdo aspectual que são, segundo Fodor, os veículos do pensamento)[317].

As razões a favor do realismo acerca de representações internas apresentadas por autores como Dretske, Fodor e Millikan são razões para contestar o normalmente chamado instrumentalismo (ou pelo menos interpretativismo) da TSI. Já se viu no entanto atrás que a posição defendida por Dennett quanto à natureza das representações não é exactamente instrumentalista: de acordo com uma posição estritamente instrumentalista, as descrições mentais teriam uma justificação meramente pragmática, sendo fechada a porta a questões ontológicas. Não é isso que se vê em *Real Patterns*. Não é verdade também que a posição de Dennett não considere aspectos ontológicos: em *Real Patterns* trata-se incipientemente de ontologia, da forma de existir dos padrões da mentalidade. Além do mais, é fundamentalmente diferente afirmar que se atribui estados mentais por razões pragmáticas relacionadas com a previsão do comportamento e afirmar que a existência do mental está ligada a um ponto de vista. Em *Real Patterns* fica claro que a TSI é uma opção pela segunda afirmação.

No entanto, se a TSI não é instrumentalista, o que é facto problemático é que ela não abdica da figura do intérprete para falar quer de veículos de representações quer do representado nas representações. Ora, se a própria existência e natureza de representações está

[316] A elaboração de uma metafísica informacional do significado é originalmente levada a cabo em DRETSKE 1981.

[317] Defender que os modos de apresentação (MAP) são os veículos do pensamento é considerar que ter um MAP é usá-lo para apresentar ao pensamento aquilo (seja o que fôr) de que o MAP é um modo de apresentação. Captar ou ter um MAP não é pensar sobre ele mas utilizá-lo dessa maneira. Um pensamento é assim uma *relação ternária* entre um pensador, um conteúdo (lato) e um MAP. Como diz Fodor, «This treatment of the individuation of thoughts is tailor made to mediate between a semantics that wants to reduce meaning to information and a psychology that wants to reduce thinking to computation» (FODOR 1994: 55).

dependente da interpretação do sistema cognitivo por um intérprete exterior, torna-se complicada a defesa de uma posição teleofuncionalista, que por princípio evoca as funções de veículos nos próprios sistemas para explicar a semântica das representações. Deixando por agora de lado os problemas que o princípio da dependência da mentalidade relativamente ao exterior pode envolver quando se considera a consciência, considerar-se-á aqui exclusivamente os problemas que essa dependência traz para a própria teoria do conteúdo. Assim, a posição de Dennett será contrastada apenas com posições de outros teóricos naturalistas da informação e do comportamento – nomeadamente Millikan, Dretske e Fodor – que admitem a pertinência da separação das teorias do conteúdo e da consciência quando se trata de desenvolver uma teoria do conteúdo.

Ainda que nem todos os autores considerados a partir de agora sejam teleofuncionalistas, todos defendem posições externalistas quanto ao conteúdo. Todos reconhecem no entanto que o puro externalismo (tal como é defendido por exemplo pelos behavioristas lógicos) não é uma solução suficiente para o problema do conteúdo mental. Até mesmo Dennett admite que o conteúdo mental deve poder fazer diferença no comportamento do sistema e é por essa razão que propõe o conceito de "mundo nocional". O mundo nocional é a resposta possível ao problema da individuação do conteúdo, permitindo supostamente caracterizar a diferença entre acreditar que p e acreditar que q sem incorrer em preconceitos sentencialistas ou linguísticos e sem conceber o conteúdo mental à imagem da linguagem humana. Segundo Dennett, são precisamente os preconceitos sentencialistas que arruinam por exemplo a teoria fodoriana da explicação psicológica. É portanto de novo face a Fodor que a Dennett define uma posição quanto à individuação do conteúdo, opondo a ideia de mundo nocional às representações mentais reais e sentenciais da TRM.

Individuar ou medir o conteúdo de modo a poder caracterizar a explicação psicológica implica antes do mais decidir se se está a falar quando se fala de acreditar que p ou q: de objectos abstractos como proposições, de estruturas cerebrais, da interpretação de comportamento de sistemas físicos, de formas sentenciais "dentro na cabeça" ou de qualquer outra coisa. A evidência que vem dos estudos empíricos leva apenas a crer que está a ser usada informação, sugere uma teoria computacional, mas nada diz de exacto acerca da forma dos conteúdos processados ou das suas unidades.

Quanto às unidades, o mínimo que se pode dizer é que teorias naturalistas do conteúdo não podem servir-se simplesmente e sem mais explicações de entidades abstractas como proposições. Perante esta situação, certas observações meta-teóricas e certas opções são inevitáveis. Dennett pensa que se a questão do conteúdo fôr tratada a partir dos estudos das atitudes proposicionais feitos na filosofia da linguagem e na lógica o resultado serão teorias sem interesse directo para as ciências empíricas da cognição. Ora, se há coisa que várias outras disciplinas procuram na filosofia é uma teoria das proposições, do conteúdo, da informação semântica, ou como se entenda chamar-lhe, já que as ciências da cognição não passam sem informação e transformações de informação.

Chamar-se-á a partir de agora, avançando com a terminologia de Dretske, "informação semântica" à informação enquanto conteúdo, por contraste com a informação tal como a teoria formal da informação a concebe[318] (i.e. como neutra quanto ao conteúdo, puramente quantitativa, sendo a quantidade de informação medida em bits (e bytes, megabytes, gigabytes,...)). O facto de a medida ser neutra significa que quantidades idênticas de informação podem corresponder a informação semântica totalmente distinta (por exemplo 1 MB de informação pode ser som, texto, imagem, etc). Ora a questão acerca da análise, individuação ou medição da informação semântica pode ser formulada perguntando que medida poderia ser apropriada para a informação semântica, que unidade de medida teria um papel análogo ao papel dos bits e bytes da teoria da informação. De entre os autores aqui analisados Dretske é aquele que visa exactamente, e essa foi, desde logo, a importância inaugural de *Knowledge and the Flow of Information,* desenvolver um conceito semântico de informação em continuidade com a teoria matemática da informação.

2.2.2 Beyond Belief*: mundos nocionais e proposições como medidas provisórias.*

Ao contrário de defensores do teleofuncionalismo como Dretske (para quem a teoria da informação mostra que a informação é qualquer coisa de objectivo, sendo essa revelação fundamental para a teoria semântica) e Millikan (que reporta a adequação do comportamen-

[318] Alude-se à teoria desenvolvida por C.Shannon. Cf. por exemplo DRETSKE 1981, 1ª Parte.

to de sistemas a Ícones Intencionais Internos), Dennett aborda a individuação do conteúdo através um conceito ainda interpretativo, o conceito de mundo nocional. O conceito é proposto no artigo *Beyond Belief*[319] e serve para especificar o que existe numa criatura que determina aquilo em que esta acredita. O problema de Dennett em *Beyond Belief* é duplo: ele pretende desconstruir a psicologia das atitudes proposicionais, i.e. o usual apelo a proposições como conteúdos de crenças, e encontrar um substituto para as proposições. Chegará ao fim com uma noção muito fraca de mundo nocional, através da qual os mecanismos psicológicos seriam indirectamente caracterizados. A noção pretende substituir as proposições sem recorrer ao subterfúgio da Hipótese da Linguagem do Pensamento, de acordo com a qual são símbolos em linguagem do pensamento e não proposições aquilo que no interior dos sistemas constitui as crenças. As várias unidades alternativas de conteúdo analisadas em *Beyond Belief* são portanto, respectivamente as *proposições* da "psicologia das atitudes proposicionais", as *frases em linguagem de pensamento* da "psicologia das atitudes sentenciais" e os *mundos nocionais* da "psicologia das atitudes nocionais" proposta pelo próprio Dennett.

Apesar da opção que o teleofuncionalista faz pela biologia como modelo para a psicologia, em detrimento de modelos lógicos, mesmo Dennett admite, em *Beyond Belief*, que determinadas questões acerca do conteúdo em seguida enumeradas, a que se chamará questões fregeanas[320], são mais facilmente tratáveis a partir de um ponto de vista lógico-linguístico sobre o mental. As questões fregeanas são questões relativas às condições para a existência de: (1) entidades dotadas de *valores de verdade*, (2) entidades intensionais que determinam "*extensões*", (3) capturabilidade (*graspability*) pela mente dessas mesmas entidades. Segundo Dennett, o facto de nada satisfazer simultaneamente todas as condições é problemático para o teórico da cognição[321]. É precisamente por não existirem tais entidades que doutrinas bastante distintas da natureza de proposições competem entre si.

Dennett defende em *Beyond Belief* que não existem unidades universais e reais de informação semântica (que seriam por exemplo

[319] DENNETT 1987j.

[320] Apesar da terminologia adoptada não se nega que seja possível considerar que «from Frege's standpoint Dennett's enquiry appears quite misconceived» (DUMMETT 1988: 513).

[321] Cf. DENNETT 1988d: 500. Dummett considera estas fomulações psicologistas e grosseiras (DUMMETT 1988).

"proposições"). A partir do momento em que ninguém sabe o que são proposições, a psicologia das atitudes proposicionais (i.e. a ideia segundo a qual os conteúdos de crenças são proposições) não tem fundamentos sólidos. É pelo facto de nem proposições nem frases numa linguagem do pensamento serem adequadas para traduzir o conteúdo que Dennett propõe a ideia de *mundo nocional*. A proposta corresponde à seguinte ideia: a informação semântica está de facto no organismo de uma forma que faz diferença, no entanto só é possível descrevê-la indirectamente. A descrição da informação semântica não é nem uma descrição do mundo exterior ao sistema nem uma descrição de alguma coisa de interior ao sistema: não se pode evocar nem "O mundo-tal-como-ele-é" nem frases em linguagem de pensamento para decidir que (retomando o exemplo de Quine) Ortcutt é o pilar da comunidade e é espião. A principal razão pela qual Dennett argumenta que a informação semântica deve ser medida indirectamente é o facto de o Racionalismo do Significado ser falso[322]. O Racionalismo do Significado está ligado ao princípio de Russell segundo o qual «não é possível fazer um juízo sobre um objecto sem saber sobre qual objecto se está a fazer o juízo»[323]. O que se passa é precisamente que é possível fazer juízos sem que se saiba sobre qual "objecto real" se ajuíza.

Recapitulando, as proposições seriam os primeiros candidatos a entidades dotadas de valores de verdade, as medidas do conteúdo, os bits da teoria da informação semântica. Existem várias sugestões em campo relativamente à natureza das proposições de acordo com as quais estas seriam nomeadamente (1) frases, símbolos com sintaxe, (2) conjuntos de mundos possíveis ou funções de mundos possíveis para valores de verdade, (3) colecções ordenadas de objectos e propriedades no mundo. Em qualquer caso, um objecto que fosse a medida de conteúdo ideal deveria satisfazer as três condições acima enumeradas: ser dotado de um valor de verdade, ser intensional e determinar extensões e ser capturável pela mente. Den-

[322] *O Racionalismo do Significado* consiste na atribuição de uma autoridade especial aos produtores e intérpretes de Ícones Intencionais quanto ao que estes significam, ou em suma, na ideia segundo a qual um sistema cognitivo deve por princípio ser capaz de reconhecer a priori aquilo que as suas representações representam, e se duas representações diferentes em momentos diferentes do tempo representam ou não representam o mesmo "representado". De acordo com o Racionalismo do Significado a mesmidade de conteúdo de A e B, duas representações mentais, seria transparente ao pensador dos pensamentos A e B, que seria imune ao erro nos actos de identificação do mesmo como mesmo. Para Millikan, o erro da posição resulta de uma ocultação da mecânica dos veículos do pensamento .

[323] DENNETT 1987: 204. O princípio foi assim chamado pelo filósofo inglês Gareth Evans.

nett dá como adquirido que vários argumentos (de H. Putnam, S. Stich, D. Kaplan, etc) mostram que nada satisfaz todas as condições simultaneamente[324]. É por isso que a Hipótese da Linguagem do Pensamento surge como solução alternativa: admitindo que a questão da natureza das proposições é irresolúvel, supõe-se que algo de concreto – símbolos na linguagem da mente – é fisicamente instanciado quando um "pensamento" é captado. As crenças acerca de X ou Y têm que ser compostas por alguma coisa (não por X ou Y mas por símbolos, estruturas sentenciais que dão conta da característica opacidade referencial dos estados intencionais). Feita a distinção de D. Kaplan entre conteúdo e carácter[325], obtém-se que as frases na cabeça são entidades com carácter. Assim, o facto de duas entidades acreditarem a mesma coisa já não remete para objectos abstractos (proposições, cuja natureza é problemática) mas para uma forma sentencial no interior de um sistema, uma frase em linguagem de pensamento.

Seria de esperar que Dennett não se convertesse à Hipótese da Linguagem do Pensamento. De facto, embora admita que a proposta de Fodor não se compromete com proposições, ele faz-lhe várias críticas. Antes de mais, a Hipótese da Linguagem do Pensamento distingue os estados psicológicos de forma demasiado fina, obrigando a admitir que as contribuições semânticas para um sistema podem sempre ser colocadas em forma de linguagem. Além disso, não é de todo clara a origem do carácter sintáctico da linguagem do pensamento nem são claros os princípios que permitem distinguir o que conta como sintáctico no cérebro. Por outro lado, a neutralidade em relação à linguagem, a vantagem dos tratamentos teóricos do conteúdo em termos de proposições, desaparece com a Hipótese da Linguagem do Pensamento. No entanto, nos termos da TSI, o núcleo da crítica à Hipótese da Linguagem do Pensamento é o seguinte: *os princípios do psicologia das atitudes sentenciais são impraticáveis pois só a EI pode guiar a descoberta do design*. Aquilo que a psicologia das atitudes sentenciais propõe é uma separação entre psicologia sintáctica e psicologia semântica que parece a Dennett impossível: «ter o texto independentemente da interpretação deste não é uma verdadeira possibilidade da psicologia»[326]. Noutras palavras, ou se supõe, na linha da TSI que nada é intrinsecamente representacional, ou ter-se-á que admitir, como Fodor, representações que não representam o que quer que seja[327]. Dennett admite que alguma distinção entre semântica e sintaxe será em última análise retrospectivamente realizada numa psicologia madura. Mas aquilo a que Dennett objecta na psicologia das atitudes sentenciais é precisamente a distinção entre sintaxe e semântica, i.e. ao facto de esta conceber o interior representativo dos sistemas à imagem de linguagens formais exteriores, comprometendo-se com um sentencialismo gratuitamente forte.

Afastada a Hipótese da Linguagem do Pensamento, o que é essencial a uma teoria do conteúdo é uma especificação intermédia dos traços psicológicos, suficientemente independente de factos relativos ao ambiente externo do organismo de modo a capturar a "contribuição do organismo". Essa é a função do mundo nocional, uma noção minimamente internalista[328], que corresponde a uma caracterização indirecta dos mecanismos psicológicos através de proposições (a caracterização é indirecta pois é uma caracterização do "mundo" (nocional) para lidar com o qual os mecanismos foram desenhados). As proposições são portanto convocadas, mas Dennett declara adoptar a posição neutra de Paul Churchland quanto ao uso das proposições em teoria da cognição (Churchland compara o papel das proposições na teoria da cognição com o papel dos números na física)[329]. Dennett acrescenta que as proposições são mais como dólares do que como números,

[324] DENNETT 1987j: 121-130. Por exemplo, de acordo com Dennett a experiência mental da Terra Gémea de Putnam sugere ou que o significado não está na cabeça, não é capturável pela mente ou que o significado não determina a extensão.

[325] Cf. DENNETT 1987j: 131-133. Recorde-se que é a distinção de D. Kaplan entre *carácter* e conteúdo (para caracterizar o significado) que inspira a distinção de Fodor entre conteúdo lato e conteúdo estrito. A ideia é que algo com o mesmo carácter pode ter conteúdos diferentes. Por exemplo "eu" tem um carácter ou significado constante mas um conteúdo diferente (dependente do uso por diferentes locutores). O carácter é uma função do contexto para o conteúdo. Fodor transpôe esta distinção feita por Kaplan relativamente a indexicais para o conteúdo psicológico: o conteúdo estrito de uma crença e o contexto determinam em conjunto o conteúdo lato da crença. A terminologia é no entanto abandonada na actual teoria fodoriana do conteúdo (cf. FODOR 1998).

[326] DENNETT 1987j: 146.

[327] MILLIKAN 1993b: 84.

[328] Esta noção, inspirada nos mundos possíveis e na relatividade destes ao agente, continuará a ser explorada no Capítulo 3 no contexto do Modelo dos Esboços Múltiplos sob o título de "mundo heterofenomenológico". A heterofenomenologia contrasta com a auto-fenomenologia que seria supostamente praticada por fenomenólogos husserlianos (cf. DENNETT 1987j: 153). Nas palavras de Lynn Rudder Baker (RUDDER BAKER), entramos no *território em que o conteúdo se encontra com a consciência e em que as respectivas teoria não podem evitar cruzar-se*. Segundo L. Rudder Baker o conteúdo encontra a consciência em dois pontos: 1. Quando se trata de relatos de crenças conscientes 2. Quando se fala de fixações de conteúdo no cérebro. Esta última questão será tratada mais pormenorizadamente no Capítulo 3.

[329] Cf. DENNETT 1987: 123 e DENNETT 1998: 114.

i.e. são medidas eficazes, que possibilitam uma uniformização útil, mas não valores absolutos ou entidades reais. As proposições servem para "medir" mundos nocionais, objectos e eventos em que o sujeito acredita, caracterizados do ponto de vista da terceira pessoa. Esta caracterização é no entanto um atalho heurístico e não poderá nunca ser precisa pois não corresponde à captura pela mente de entidades abstractas reais e determinadas. O que a utilização das proposições como medidas provisórias permite é uma aproximação ao mundo nocional do sujeito, definido como «o ambiente ao qual o organismo, tal como é actualmente constituído, está idealmente adaptado»[330]. A relatividade ao agente é, de resto, de acordo com Dennett a grande vantagem do tratamento das proposições como conjuntos de mundos possíveis, uma vantagem que Dennett pretende transferir para a ideia de mundo nocional. De acordo com Dennett, constituem versões da ideia de mundo nocional a fenomenologia husserliana, a IA, a doutrina quineana da indeterminação radical e a interpretação da ficção por teóricos literários. É importante sublinhar de novo que o mundo nocional «não consiste em representações mas em representados»[331]: o mundo nocional é o mundo-do-sujeito, não o mundo de representações sintácticas vazias dentro da cabeça do indivíduo que o solipsismo metodológico da Hipótese da Linguagem do Pensamento sugere.

A objecção maior à ideia de mundo nocional é, como seria de esperar, um argumento do tipo Terra Gémea e em *Beyond Belief* Dennett procura enfrentá-lo, admitindo que é possível que o Oscar 1 da Terra e o Oscar 2 da Terra Gémea partilhem o mesmo mundo nocional encontrando-se em "mundos reais" distintos, o que explicaria a semelhança dos seus comportamentos e a diferença do conteúdo das suas crenças[332]. O ponto é ilustrado com a história de um local chamado Shakey's Pizza Parlor[333]. Uma pessoa adormecida (Tom) é transposta de um estabelecimento numa localidade (Shakey's

[330] DENNETT 1988: 499.

[331] DENNETT 1987j: 154.

[332] Apesar desta "solução" Dennett revela usualmente ter objecções de princípio ao uso das experiências mentais como a Terra Gémea. De um ponto de vista menos caritativo, dir-se-ia que não sabe utilizá-las ou que não compreende o uso da estipulação que elas envolvem, uso que se destina a verificar qual teoria se aplica, preferindo antes *exemplos*, que são aplicações de teoria e podem ser vistos de várias maneiras. Nesse sentido o uso de exemplos, característico de Dennett, seria menos sofisticado e menos controlado do que o uso, que ele critica, de experiências de pensamento.

[333] DENNETT 1987: 167-168. Tom, a personagem da história é transposto do Shakey's Pizza Parlor de Costa Mesa para o Shakey's Pizza Parlor de Westwood Village, cuja decoração é idêntica até ao último pormenor.

Pizza Parlor, em Costa Mesa) para um outro exactamente idêntico noutra localidade (Shakey's Pizza Parlor, em Westwood Village). Acordado, Tom, aposta que as suas iniciais estão gravadas numa determinada porta. Vai verificar e perde a aposta. O problema que se coloca –caracterizar o estado psicológico da pessoa quando pensa na porta – é análogo aos famosos problemas do Pierre de S. Kripke[334], que pensa que Londres é uma bela cidade e que *esta* cidade (que é de facto Londres, sem que Pierre o saiba) é uma cidade horrível. Regressando ao exemplo de *Beyond Belief*, não é, segundo Dennett, possível distinguir nas crenças de Tom aquelas que são acerca do Shakey's Pizza Parlor de Costa Mesa daquelas que são acerca do Shakey's Pizza Parlor de Westwood Village.

É verdade que, tal como se acentua nas discussões acerca de crenças *de re* e crenças *de dicto*, quando o objectivo é distinguir a especificidade das crenças *de re*, a existência de crenças *de re* não ocorrerá normalmente na ausência das relações causais correctas. No entanto (e esse é o ponto de Dennett) não existe uma maneira legítima de fazer uma estrita distinção entre crenças *de re* e crenças *de dicto*, pelas simples razão de que é perfeitamente possível fazer juízos sobre objectos sem se saber sobre quais objectos se está a ajuízar (é o que acontece com Tom no Shakey's Pizza Parlor, com Oscar 2 na Terra Gémea de Putnam, com o Pierre de Kripke e com o Ralph de Quine). Assim, o Princípio de Russell, o expoente do Racionalismo do Significado, de acordo com o qual é impossível formular um juízo sobre um objecto sem se saber sobre qual objecto se ajuíza, é um princípio falso, como é falsa a ideia de uma especial intimidade ou contacto com o mundo das crenças que seriam *de re*.

A ideia de mundo nocional resume todas estas críticas: críticas ao Racionalismo do Significado, ao sentencialismo fodoriano, à possibilidade de medidas proposicionais reais e determinadas de conteúdo, à distinção entre crenças *de re* e *de dicto*. Como se verá no próximo capítulo, a ideia de mundo nocional, que é a última palavra de Dennett quanto ao conteúdo de estados mentais, encontra vários problemas, sobretudo quando se procura conjugá-la com o tratamento da consciência através do Modelo dos Esboços Múltiplos. De facto, embora a categoria de mundo nocional não seja uma categoria causal (quando correctamente concebida ela envolve precisamente a suspensão da referência a causas da experiência mental do sistema), no âmbito

[334] KRIPKE 1979.

do MEM Dennett virá a admitir algo como uma verificação ou contraste do objectos do mundo nocional relativamente ao mundo "real", uma verificação que seria feita, no caso humano, a partir do que se encontra no cérebro do indivíduo. Apesar de todas as suas objecções à escrita cerebral fodoriana, Dennett sugere por vezes que é possível determinar precisamente através de eventos internos se os relatos sinceros dos sujeitos acerca daquilo em que acreditam são verdadeiros, mais especificamente através de um teste de mapeamento cerebral (*brain-mapping*)[335]. Como L. Rudder Baker faz notar[336] não se trata aí apenas de admitir que as pessoas cometem erros: trata-se de defender que descobertas acerca de eventos cerebrais podem "desmentir" a auto-atribuição de crenças conscientes, i.e. podem mostrar que um sujeito pensa falsamente que pensa aquilo que pensa, e em última análise aniquilar o pensamento e o sujeito. Argumentar-se-á à frente que esta sugestão não é coerente. Nas formulações do próprio Dennett o mundo nocional ou heterofenomenológico aparece rodeado de lapsos psicologistas e "representacionistas". Por exemplo a ideia segundo a qual a natureza do mundo nocional poderia ser subvertida por inspecção cerebral é uma noção psicologista, i.e. pretende responder a questões acerca do representado na representação com observações empíricas acerca de veículos de representação (é como se se pensasse, numa situação em que se agarra por exemplo um cubo com a mão, que a análise da mão revela aquilo que ela está a agarrar). No entanto, a operacionalidade da ideia de mundo nocional supera os lapsos da formulação de Dennett, bastando que (ao contrário do que Dennett por vezes faz) não se esqueça os seguintes pontos: a descrição do mundo nocional não é introspecção, o mundo nocional não é um "interior do espírito", a descrição do mundo nocional não é uma descrição de veículos intermediários (i.e. de representações no sentido da ciência cognitiva). Em suma, o que o mundo nocional pretende fazer ver em teoria da cognição é que não é do processo pelo qual um sistema apercebe objec-

[335] Cf DENNETT 1991: 85 «My suggestion, then, is that if we were to find real goings on in people's brains that had enough of the "defining" properties of the items that populate their heterophenomenological worlds, we could reasonably propose that we had discovered what they were really talking about – even if they initially resisted the identifications. And if we discovered that the real goings-on bore only a minor resemblance to the heterophenomenological items, we could reasonably declare that people were just mistaken in the beliefs they expressed, in spite of their sincerity». Para a ideia de *brain-mapping* e a crítica ao que esta representa (nomeadamente uma indiferenciação da TSI como teoria do conteúdo relativamente à TRM de Fodor) cf. RUDDER BAKER 1994: 5.

[336] RUDDER BAKER 1994: 5.

tos no seu mundo nocional que o sistema se apercebe mas sim dos próprios objectos apercebidos. Este é o âmbito das questões atrás chamadas fregeanas, que são relativas à teoria do pensamento em geral. Há boas razões para pretender que esta deve ser independente da psicologia (entendida como uma disciplina que teria entre os seus objectos por exemplo questões relativas àquilo que acontece nos canais de sódio e cálcio dos neurónios de sujeitos humanos – questão que pode ser considerada irrelevante para uma teoria do sentido, da verdade e da referência).

Se Dennett critica a Hipótese da Linguagem do Pensamento por considerar o conteúdo como um objecto independentemente existente, o uso que por vezes faz da ideia de mundo nocional é ele próprio susceptível da mesma crítica. Um dos problema com o mundo nocional e com o facto de Dennett não rejeitar a associação deste a um teste de *brain-mapping*, é que isso além do mais contraria os princípios gerais (externalistas) do teleofuncionalismo. Como L. Rudder Baker pergunta, como poderia o conteúdo de uma crença como "Os invernos são frios em Vermont" ser verificado por um teste de *brain-mapping*?[337]. De acordo com o externalismo característico de posições teleofuncionalistas o que fixa o conteúdo dessa crença tem a ver com os invernos em Vermont e não com o cérebro do indivíduo. O teste de *brain-mapping* poderá ser localmente pertinente para comparar hipóteses em casos experimentais, nomeadamente quando se trata de relatos introspectivos que acompanham testes de percepção como os que serão discutidos no próximo capítulo. Fora disso torna-se bizarro. Além do mais, o facto de uma vida mental ser a vida mental própria de um indivíduo não faz dela uma introspecção ou auto-relato descritivo contínuo, e essa ideia subjaz tacitamente ao tema do mundo nocional, mesmo se vai contra os pressupostos de Dennett segundo os quais a representação linguística não é de todo exemplar.

Com a especificação relativa a mundos nocionais fica completa a abordagem da teoria do conteúdo na TSI. Começou-se por caracterizar esta como um realismo moderado quanto à natureza de representações apoiado na existência de padrões reais reconhecidos pelo intérprete. Relativamente ao conteúdo das representações, Dennett defende uma posição teleofuncionalista, que é como qualquer posição teleofuncionalista, externalista e no caso de Dennett também interpretativista e explicitamente anti-sentencialista. É certo que o mundo

[337] RUDDER BAKER 1994.

nocional é uma contribuição "individualista": algo no interior do sistema deve cumprir uma função de representação. No entanto não é legítimo pretender que é possível a individuação desse conteúdo a não ser provisoriamente utilizando as proposições como medida de modo a configurar o mundo nocional. Sendo dependente de uma interpretação e dizendo respeito ao representado na representação e não aos veículos da representação, o mundo nocional não é uma categoria causal mas mais propriamente fenomenológica. O "mundo nocional" captura o conteúdo da experiência do sistema cognitivo: sendo ou não de alguma forma ilusória, essa experiência não é nunca acerca do *hardware* da cognição. Nestas condições, o mundo nocional não é suficiente para dar lugar a *explicações* psicológicas. Apenas são possíveis *descrições* e *previsões* intencionais.

2.2.3 *Explicação psicológica e individuação do conteúdo de acordo com outras posições externalistas.*

2.2.3.1 *A biopsicologia de R. Millikan.*

Dennett refere-se frequentemente à filosofia da psicologia desenvolvida por Ruth Millikan como um prolongamento e um aprofundamento das suas próprias posições. Para Dennett, Millikan é «a teórica naturalista completa»[338]. Quer Millikan quer Dennett defendem que uma abordagem naturalista do significado conduz a concluir que mesmo quando se trata de significação humana sofisticada e linguística não podemos proclamar que somos nós que sabemos o que "queremos-dizer" (*mean*), i.e. não podemos afirmar que sabemos exactamente o que os nossos usos significativos de símbolos significam[339]. Apenas uma teoria da significação erradamente apriorista permitiria

[338] Cf. DENNETT 1984d, Prefácio a MILLIKAN 1984.

[339] Dennett compara esta convicção com a falácia intencional em crítica literária, i.e. com a convicção de que se deve perguntar ao autor do texto pelo verdadeiro sentido do texto, uma vez que ele é o único que pode conhecê-lo. De acordo com a perspectiva teleofuncionalista de Millikan e Dennett, o Racionalismo do Significado representa uma psicologização do aspecto normativo do significado. Millikan (MILLIKAN 1984) considera que até mesmo as teorias do significado de autores tipicamente anti-interioristas como Quine e Wittgenstein estão infectadas por este Racionalismo do Significado. Millikan acusa especificamente estes dois autores de não terem podido desenvolver uma teoria dos actos de identificação por terem confundido o anti-cartesianismo como ataque à certeza atingida por auto-acesso com uma proibição de referir o que quer que seja de interior (à cabeça, ao cérebro) na explicação da significação (cf. MILLIKAN 1984, Part IV, Theory of Identity).

pensar assim, mas esta não passa de um Mito do Dado quanto à significação, desmontável pelo naturalismo correctamente concebido.

Em *Language, Thought and Other Biological Categories*[340] Millikan apresenta pela primeira vez a sua posição naturalista e realista em filosofia da psicologia, no âmbito da qual formula uma teoria da "explicação" psicológica. De acordo com Millikan, a biopsicologia não lida com leis, como aliás é o caso das ciências biológicas em geral. Estas descrevem mecanismos que contribuem para o desenvolvimento e sobrevivência dos seres vivos, mecanismos que têm condições normais de operação mas que são imperfeitos e muitas vezes falham. As funções biológicas não são em geral muito fiáveis, nomeadamente porque o seu sucesso depende do ambiente exterior ao organismo. Assim, em geral, a "explicação" nas ciências biológicas, incluindo a psicologia, é uma descrição e um estudo de normas e de funções próprias, e não uma descoberta de leis: «não é certamente uma lei psicológica, por exemplo, que as nossas crenças são verdadeiras, embora seja uma (teleo)função dos nossos sistemas de fixação de crenças fixar crenças verdadeiras»[341]. A biopsicologia é, assim, vista por Millikan como o estudo de um processo, o comportamento, cujas fronteiras não se identificam com os limites do organismo individual.

Na medida em que a psicologia é parte da biologia, as categorias psicológicas (como crenças, desejos, etc) são, exactamente como as categorias biológicas, categorias de Funções Próprias. A teoria das Funções Próprias é o fundamento da teoria do comportamento. Millikan define a Função Própria de um dispositivo[342] como aquilo para que este foi desenhado (*designed*). A sua intenção com a noção de Função Própria é estabelecer o contraste com as funções acidentais ou impostas. É a história de um dispositivo (*device*), as funções que os seus antecessores cumpriram e que fizeram com que estes fossem selecionados, que define a sua Função Própria. Sem especulação acerca de Funções Próprias não existiria diferença entre a física e a biologia (incluindo a psicologia) de um sistema. Pensando num animal, as actividades deste podem ser descritas de inúmeras maneiras, no entanto essas descrições não serão todas

[340] MILLIKAN 1984.

[341] MILLIKAN 1993d: 177. A biopsicologia estuda o que é Normal, mas esse normal não é uma média estatística. Millikan maiuscula o termo Normal para marcar o seu uso normativo.

[342] Millikan utiliza o termo *device* para referir artefactos construídos, orgãos, organismos, i.e. em geral objectos com *design*, que têm Funções Próprias.

elas descrições de comportamento: o comportamento é apenas a *forma funcional da actividade do organismo*, embora não seja distinguível das outras actividades por qualquer traço aparente. Millikan defende que «a tarefa do cientista comportamental não é apenas estudar as propriedades de um pedaço de matéria viva (...) mas as propriedades de um sistema biológico, as propriedades que foram responsáveis pela proliferação e sobrevivência dos ancestrais da criatura. Estas propriedades figuram na explicação de como aconteceu que uma proporção crítica de incorporações históricas do sistema (...) tenha evitado a destruição e se tenha reproduzido. O que um sistema biológico faz enquanto sistema biológico e não enquanto conjunto de átomos é o que os seus ancestrais historicamente fizeram e que lhes permitiu sobreviverem e reproduzirem-se»[343]. Millikan insiste que a função biológica não sobrevem à física do sistema: que aquilo que um paleontólogo encontra seja um dente ou um corno não é uma propriedade superveniente à forma física do objecto encontrado antes depende da história. O mesmo se deve afirmar quanto ao conteúdo ou significado, enquanto função da função. Dada esta dependência da história, os estados mentais de uma réplica física instantânea e exacta de cada um de nós não teriam, de acordo com Millikan, conteúdo semântico[344].

A partir do momento em que Millikan defende que as categorias psicológicas são categorias biológicas, e como todas as categorias biológicas, categorias de Funções Próprias, está a afirmar que estados mentais são os estados mentais que são por estarem projectados (*designed*) para uma determinada função. Tratando-se de funções de dispositivos naturais estas funções são aquelas das quais a reprodução continuada desses dispositivos depende, não se identificando com as suas disposições actuais nem com a sua constituição física. As Funções Próprias apenas podem ser identificadas olhando para a história, e para as intenções nesta, seja o autor do dispositivo um humano ou a selecção natural. É a partir da noção de Função Própria que Millikan obtem o conceito de Normal. O conceito aplica-se às condições para o cumprimento das Funções Próprias historicamente fixadas do dispositivo. A ideia de Normal acentua que um dispositivo pode sempre falhar no cumprimento da sua Função Própria. Esta pode até nem ser estatisticamente dominante no conjunto das performances do dispositivo (o exemplo clássico de Millikan são os espermatozóides, que pouquíssimas vezes cumprem a função para que foram desenhados[345]). É a importância da noção de Normal nas categorias biológicas em geral que permite conceber os impedimentos à existência de leis psicológicas: os dispositivos com Funções Próprias podem sempre falhar, não obter. Não é assim pertinente descrever os seus funcionamentos em termos de leis.

Apesar das salvaguardas quanto à forma da explicação psicológica, Millikan admite a realidade de representações, fazendo diferença no sistema. A teoria das Funções Próprias prevê a existência de Ícones Intencionais reais: «o Ícone Intencional (é uma coisa tal que) mesmo que não exista nenhum traço ambiente ao qual ele se aplique apropriadamente, ainda assim deveria existir esse traço»[346]. Um exemplo elucidativo de Ícone Intencional utilizado por Millikan é o exemplo das danças de abelhas que indicam onde está o néctar. O exemplo é importante pois evidencia vários traços da intencionalidade natural: mostra que pode existir intencionalidade sem pensamento consciente nem identificação de referentes, mostra que a produção e o consumo[347] de Ícones Intencionais podem ser devidos a dispositivos diferentes, que foram desenhados para cooperar entre si (neste caso a dança executada por certas abelhas servirá para guiar outras abelhas) e mostra que o conteúdo ou significado dos Ícones Intencionais é descrito nomeando características que devem existir no ambiente dos sistemas.

É através das noções de Funções Próprias Normais e de Ícones Intencionais (entre os quais Millikan inclui representações, caracte-

[343] MILLIKAN 1993c: 149.

[344] Esta tese – «such a double has no proper functions because its history is not right» (MILLIKAN 1993f: 18) – constitui o ponto exemplar da teoria das funções psicológicas como Funções Próprias. A situação é evidentemente análoga ao clássico Swampman de Davidson (cf. DAVIDSON 1987). Fodor considera a conclusão absurda (Fodor 1994: 115) e evidenciadora daquilo que está errado com o apelo teleofuncionalista à história. Para Fodor não é a história mas sim as relações nómicas que ligam os estados mentais a matérias no mundo que explicam o conteúdo semântico de estados mentais. Essas relações podem acontecer sem a história correcta e é por isso mesmo que devem ser abordadas através de contrafactuais, mesmo que os dispositivos em causa resultem de processos darwinianos. Já Dretske (DRETSKE 1990: 13-14) defende uma ideia muito semelhante à de Millikan. Os movimentos do duplo instantâneo não seriam acções com significados por trás: «I move my arm in this way in order to frighten away a pesky fly. With such a purpose I am, let's say, shooing away a fly. That is my action. My biological twin, though he moves his arm in the same way (with the same result) does not shoo away a fly. He doesn't have wants or beliefs, the kinds of purposes I have in moving my arm. He isn't therefore, performing the same action».

[345] Cf. MILLIKAN 1984. Cumprem-na, é claro, vezes suficientes.

[346] MILLIKAN 1993b: 101.

[347] É este o termo que Millikan utiliza para dizer que os Ícones Intencionais são utilizados para guiar o comportamento.

rizadas pelo facto de estarem envolvidas em inferências[348], as quais pressupõem termos médios explícitos e a identificação destes[349]) que Millikan constrói uma teoria naturalista da intencionalidade, na qual, poder-se-ia dizer, a diferença entre a EI e a ED desaparece. Espécimes ou acontecimentos de pensamento ou de linguagem caem sob categorias de Função Própria, categorias de *design*. É a história natural causal dos Ícones Intencionais que explica a sua intencionalidade. Assim, nada significa ou representa intrínsecamente: apenas coisas no mundo que são especiais em virtude de uma história natural causal significam. Dada a dependência da normatividade relativamente a uma história de selecção natural de um *design*, de acordo com Millikan os produtores e intérpretes de Ícones Intencionais – i.e. os sistemas que os utilizam para significar – não acedem conscientementemente ao significado destes nem controlam a sua normatividade.

Recapitulando, Ícones Intencionais são eventos físicos com *design* que têm como Função Própria Normal significar. Por definição, um Ícone Intencional adapta o seu utilizador (produtor ou intérprete) a uma característica (do ambiente) mapeada. Esta característica é o seu conteúdo (o exemplo clássico são as danças das abelhas que mapeiam o lugar onde existe o néctar). A intencionalidade é assim capturável através da ideia de utilização (produção e consumo) Normal de Ícones Intencionais para significar. A falsidade será precisamente uma falha do Normal, um defeito. É claro que um tal defeito apenas é discernível de um ponto de vista normativo: nada é defeituoso em si próprio ou devido a disposições actuais mas apenas em relação ao Normal. Esta é a base da existência do falso por entre as representações. Nestas circunstâncias, não são os utilizadores dos Ícones Intencionais que sabem o que estes significam: o Racionalismo do Significado é falso. A oposição ao Racionalismo do Significado é o principal ponto de acordo entre Dennett e Millikan, enquanto teleofuncionalistas. O realismo de Millikan quanto a representações internas condu-la por outro lado à defesa de posições acerca da natureza da racionalidade e dos conceitos que diferem significativamente das posições de Dennett. Por exemplo e relativamente à racionalidade, de acordo com Millikan é possível falar de racionalidade propriamente dita e não apenas de

[348] MILLIKAN 1993: 98-101.

[349] Millikan considera que esta (re) identificação de uma representação interna como a *mesma* não é feita de forma apriorista.

atribuição de racionalidade, nomeadamente quando existem Ícones Intencionais internos de um tipo específico, Ícones Intencionais internos que são representações, que participam em inferências mediadas. Aliás, de acordo com Millikan o facto de a adscrição do conteúdo na TSI ser feita globalmente, livre de referência a representações internas, é em geral incompatível com factos que evidenciam a realidade e as relações entre si de Ícones intencionais específicos usados pelos sistemas. Por exemplo, existe uma certa ligação do conteúdo de representações mais específicas com o conteúdo de representações menos específicas. Considere-se a dança das abelhas e a frase em língua natural que a descreve "Existe néctar a sudoeste a uma distância d". Os Ícones Intencionais em causa têm essa ligação entre si porque um mesmo conteúdo é mapeado por ambos, ambos obtendo em idênticas condições Normais. Esses Ícones Intencionais são individualizáveis e seu conteúdo não deve ser concebido como uma atribuição global ao sistema. Millikan discorda portanto do holismo da TSI.

As incompatibilidades entre a TSI e a biossemântica de Millikan não ficam por aqui. As questões que conduzem Millikan ao aprofundamento do realismo ligado à biossemântica e que a aproximam de uma teoria a que se poderia chamar uma teoria fodoriana dos conceitos são questões acerca de identidade e mesmidade de Substâncias e Propriedades no mundo. De acordo com Millikan, dispositivos intencionais como organismos foram desenhados para reconhecer a identidade daquilo que no ambiente importa para a sua sobrevivência, nomeadamente a identidade de Substâncias e Propriedades. Assim, uma posição realista acerca de Substâncias e Propriedades no mundo é perfeitamente defensável e totalmente independente do (mau) Racionalismo do Significado. As considerações realistas de Millikan decorrem de uma teoria dos actos de identificação – precisamente um fenómeno que Dennett considera biopsicologicamente importante e normalmente ausente de teorizações da significação inspiradas na lógica. Um acto de identificação é o reconhecimento por um sistema cognitivo de algo no mundo como o mesmo em momentos diferentes (por exemplo e parafraseando um título de Millikan, o reconhecimento por um sistema cognitivo de *more mama, more mouse more milk*[350]). Este acto, que permite o reconhecimento da uma mesmidade ao longo do tempo por um siste-

[350] MILLIKAN 1998.

ma cognitivo, cumpre uma função psicológica básica em organismos (que possibilita por exemplo a coordenação multimodal daquilo que é percebido pelo organismo, bem como actos de inferência), e é uma habilidade prévia à existência de juízos explícitos, uma habilidade que suporta por exemplo uma incipiente "racionalidade" animal. Esta poderia ser definida como uma inferência mediada em que dois veículos de conteúdo são combinados (utilizando um termo médio) de modo a produzir um terceiro veículo contendo nova informação. Num exemplo de Millikan[351], imagine-se uma criatura que tem na cabeça Ícones Intencionais que mapeiam os locais onde esteve. A criatura tem nomeadamente um mapa do local onde viu água pela última vez e um outro do local onde viu leões pela última vez. "Sobrepondo" estes mapas e os seus ocupantes ela obtem um terceiro mapa que mostra a proximidade dos leões em relação à água e age de acordo com a nova informação (por exemplo afastando-se). Esta seria uma inferência mediada, envolvendo o reconhecimento de que o conteúdo de dois veículos de pensamento se "sobrepõe". É importante para Millikan fazer notar que para um reconhecimento semelhante ocorrer não é suficiente a ocorrência na mente dos dois veículos de pensamento. A mesmidade do conteúdo tem que ser efectuada, não é manifesta: nada identifica entre si veículos de conteúdo enquanto veículos de um *mesmo* conteúdo. É por isso que, como Millikan gosta de sublinhar, não existe nunca imunidade, na vida mental, à introdução do erro através de actos de identificação[352].

Ideias acerca de Ícones Internos e actos de identificação conduzem Millikan a uma teoria dos conceitos cujo primeiro princípio é o seguinte: *o problema para o organismo – identificar o mesmo como o mesmo ao longo do tempo – não é um problema relativo à divisão do mundo nas categorias certas*. Como diz Millikan, quando se trata de ter o conceito "verde" ou o conceito "amarelo" não se trata de dividir o mundo estabelecendo condições para quando algo deixa de ser verde e

[351] MILLIKAN 1993e: 114.

[352] Millikan considera a identificação essencial para a "mente racional intacta". É dela que está suspensa toda a teoria da racionalidade, uma vez que o mesmo tem que continuar a ser pensado como o mesmo durante o acto de raciocínio para que este não se extravie. Ora Millikan nota que na medida em que a mesmidade depende a todo o momento de actos de identificação, nada disto está garantido a priori. Por alguma razão a inspiração para o título de MILLIKAN 1993, *White Queen Psychology*, é a Rainha Branca da história de *Alice in Wonderland*, que devota meia hora por dia a acreditar coisas impossíveis. Do mesmo modo que a identidade, a possibilidade e a impossibilidade não são, de acordo com Millikan, acessíveis a priori ao pensamento.

passa a ser amarelo: o que interessa é poder identificar amarelo-de-novo, verde-de-novo, ou esta-mãe-de-novo, predador-de-novo, etc. De acordo com Millikan, tais actos de identificação são possíveis porque encontram suporte no mundo, na identidade de Substâncias e Propriedades. Ora, actos de identificação podem estar certos ou errados, dependendo da estrutura do mundo e não do que está na cabeça. É nesse sentido que a identidade ("reconhecida" nos actos de identificação) é um fenómeno natural. Aliás, de acordo com Millikan, os organismos encontram (não produzem) os seguintes factos no mundo: Propriedades que se recusam a admitir propriedades contrárias na substância que qualificam, Substâncias que insistem em ter propriedades com exclusão das contrárias. De acordo com Millikan, a posse de conceitos depende do aproveitamento destes factos em *testes de conceitos*.

Antes de mais, segundo Millikan, é porque a mesmidade de Substâncias e Propriedades é uma necessidade natural no mundo que os conceitos básicos em organismos são conceitos de Substâncias e Propriedades. Ter conceitos adequados é ser capaz de produzir actos de identificação competentes. Para isso, os conceitos são testados em pequenos grupos através do *teste da não-contradição*. Identificar x não é emitir uma resposta do mesmo tipo (como diz Millikan, o que um cão precisamente não faz quando nos identifica é ladrar-nos como da primeira vez que nos viu). Identificar é um acto relacional: identificar é identificar *com*. Para cada acto de *identificação* são necessários dois Ícones Intencionais com conteúdos (a característica do mundo que mapeiam) semelhantes. Chegar a ter um conceito é chegar a poder produzir tais actos de identificação. Se os termos internos não são inatos (se não temos desde logo, por exemplo, o conceito "computador") deve ser possível forjar novos termos e testá-los em tempo real. Se a rotina de gerar e testar conceitos se aplicasse à totalidade da teia das crenças, se se tivesse que esperar por toda a evidência disponível, nunca poderiam ser desenvolvidos novos conceitos. A alternativa consiste em testar os conceitos em pequenos grupos. É precisamente para decidir quão pequenos devem ser esses grupos que Millikan recorre às Substâncias, às Propriedades e à relação de contrariedade. O teste para conceitos (que são afinal habilidades de reconhecer identidade objectiva no mundo) é o teste da não-contradição: alguma coisa no mundo (a estruturação deste em Substâncias e Propriedades, identidade e contrariedade, segundo Millikan) determina se e quando os concei-

tos se aplicam, se os actos de identificação estão certos ou errados. O teste não é infalível, mas é certamente suficiente para mostrar quando é que dispositivos cognitivos estão a produzir conceitos que se aplicam suficientemente bem a particulares no mundo. Assim, a não-contradição pode ser um teste da "qualidade" de conceitos sem qualquer evocação do holismo das crenças: a coerência no pensamento simplesmente corresponde a um "traço abstracto" do mundo.

A posse de conceitos por sistemas, que permite a identificação de características importantes no mundo, é a maneira básica de significar em dispositivos, e a teoria dos conceitos é outra das razões que levam Millikan a não aceitar o holismo da TSI. Note-se que esta teoria dos conceitos é, como será o caso da teoria defendida por Fodor, atomista e anti-descritivista[353]. Saber reconhecer o mesmo-como-o-mesmo no mundo não é ter conhecimento explícito sobre o que a coisa reconhecida é, nem ter conhecimento sobre outros conceitos com os quais esse conceito tem uma relação semântica (possibilitando inferências). Esta última imagem do funcionamento da significação, o holismo da significação, de acordo com o qual qualquer significação se relacionaria com a totalidade das crenças da teia ter-se-ia, segundo Millikan, difundido entre os teóricos da significação por um excesso intelectualista. Se em *Two Dogmas*[354] existe uma plausível teoria das crenças científicas e do modo como elas são revistas, não existe no entanto nada que se assemelhe a uma teoria dos conceitos como habilidades psicológicas imprescindíveis. Não é para fazer ciência mas para algo de muito mais básico[355] que os organismos necessitam de poder aplicar fiavelmente conceitos.

A teoria dos conceitos de Millikan tem muitos pontos em comum com a de Fodor, a começar pelo facto de negar que a posse de um conceito consista num saber de características. É curioso,

[353] Uma teoria dos conceitos é *não-descritivista* quando postula que ter um conceito não é idêntico a possuir conhecimento acerca daquilo que é identificado por meio do conceito. Uma origem de tais posições não descritivistas é PUTNAM 1975.

[354] QUINE 1953.

[355] E não comunitário, acrescentaria Millikan. Esse algo de mais básico (cf. MILLIKAN 1998) é a identificação de matérias (*stuffs*) como leite, ouro, tipos reais (*real kinds*) como gatos, cadeiras, e indivíduos (como mamã, Bill Clinton ou o Empire State Buiding). Todos esses conceitos são para Millikan conceitos de substâncias e possuí-los não consiste em saber analisar as suas propriedades mas em conseguir identificar coisas no mundo como as mesmas em circunstâncias variadas e momentos diferentes, podendo assim segui-las (*track them*).

dada a "adopção" da filosofia da psicologia de R. Millikan por Dennett, que no pensamento de Millikan existam significativos pontos de convergência com Fodor, nomeadamente na teoria dos conceitos. Embora Millikan pense que a teoria do conteúdo de Fodor parte de uma má concepção de comportamento[356], existe um óbvia convergência entre ambos no que respeita ao *atomismo, não-descritivismo e rejeição do holismo*. No entanto, Millikan considera que o "comportamento" visado pela psicologia não pode excluir o meio e sendo o objectivo da explicação psicológica a explicação do comportamento leis intencionais no sentido de Fodor são impossíveis. De resto, uma psicologia feita à imagem da biologia não teria por objectivo leis nem previsões. De qualquer modo, uma tal biopsicologia envolve ainda uma concepção das representações mais realista do que aquilo que Dennett concede, de tal modo individuadora das representações que permite mesmo uma teoria atomista dos conceitos. O holismo da TSI impede o desenvolvimento de uma teoria atomista dos conceitos. A TSI envolve-se naturalmente com uma semântica do papel inferencial[357] contra a qual Fodor e Millikan apresentam evidência. Mas sobretudo e ao contrário de Dennett, a investigação biopsicológica acerca do conteúdo semântico conduz Millkan a concluir que embora a escrita cerebral (*brain-writing*) possa vir a ser ilegível para o observador exterior o padrão interno está seguramente lá: a capacidade de significar existente em sistemas cognitivos depende de veículos no interior destes que não dependem de qualquer interpretação.

2.2.3.2 *Informação e explicação do comportamento segundo F. Dretske.*

Embora Dretske partilhe com Dennett o princípio básico do teleofuncionalismo segundo o qual o *significado dos estados mentais é uma função da função biológica*, ele admite, ao contrário de Dennett, explorar o *papel causal do significado*. Evidentemente, «explorando a possibilidade de o significado ter um papel causal, está-se a explorar

[356] Millikan acusa Fodor de uma má concepção do comportamento pelo facto de não subscrever as ideias "ecológicas" da biopsicologia.

[357] A semântica do papel inferencial remete para o holismo da significação, a ideia segundo a qual o significado de algum termo consiste nas/envolve as relações deste termo com outros termos, e segundo a qual por exemplo o facto de a palavra "cão" significar a relaciona com todas as outras palavras (ou com muitas) do português.

não a possibilidade de o significado ele próprio ser uma causa (seria como ver propriedades universais, como "triangularidade" causando alguma coisa) mas a possibilidade de alguma coisa tendo significado (*a thing's having meaning*) ser uma causa»[358]. Assim, ao contrário de Dennett e do mesmo modo que Fodor, Dretske não exclui que as explicações psicológicas possam ser de algum modo verdadeiras explicações, reportadas à causalidade.

Como já se referiu, a enorme importância da obra *Knowledge and the Flow of Information* de Dretske deveu-se ao facto de ser aí desenvolvido um conceito semântico de informação em continuidade com o conceito de informação da teoria matemática da informação. De acordo com Dretske, embora os aspectos semânticos da informação sejam irrelevantes para o problema do armazenamento e transmissão de informação enquanto problema de engenharia, os aspectos de engenharia não são de todo irrelevantes para o problema semântico e mostram antes de mais que a informação é qualquer coisa de objectivo (*an objective commodity*). De acordo com a teoria da informação semântica iniciada com *Knowledge and the Flow of Information*[359], a informação consiste em indicação ou correlação[360]. Aquilo que dá conteúdo a representações mentais são relações causais-nomológicas com aquilo que elas indicam. O conteúdo semântico relaciona-se portanto com o facto de determinados sistemas estarem ligados da maneira certa ao mundo e não com relações epistémicas de justificação. A ligação pode aliás ser totalmente desconhecida para o sistema cognitivo, por exemplo para um sistema cognitivo que tem crenças. "Estar ligado a" é uma relação externa ou extrínseca.

Mas o que interessa para o problema da individuação do conteúdo e da explicação psicológica, é que Dretske prolonga a sua teoria da informação semântica com uma *teoria da explicação do comportamento*. Segundo Dretske, se a informação não explicasse e não causasse de algum modo, não seria necessário evocá-la para explicar

[358] DRETSKE 1988: 80. Dretske dá como exemplo o canto de um soprano, uma súplica lancinante que chega a estilhaçar um vidro sem que o significado seja de todo relevante para tal efeito. Ora «What is true of the soprano's output is true of reasons» (DRETSKE 1988:79).

[359] O objectivo da obra era construir uma teoria externalista do conhecimento assente sobre uma teoria da informação semântica.

[360] De acordo com Dretske, objectos de um determinado tipo O transportam (*carry*) informação acerca de estados de coisas de tipo S se estados de coisas de tipo S são responsáveis por instanciações de O. De acordo com esta definição não poderia haver má informação, por isso parte do problema da teoria da Dretske será especificar quando passa a poder existir "*misrepresentation*".

comportamento. Noutras palavras, embora a teoria do conteúdo de Dretske seja externalista ela é formulada de modo a possibilitar *explicações* do comportamento, uma via que aparentemente não está aberta à TSI. De facto e utilizando uma expressão do próprio Dretske, é possível pôr o conceito de informação a trabalhar na explicação do comportamento, já que o conceito de informação semântica é um conceito aplicável a eventos físicos que têm, enquanto tal, efeitos causais.

Como se afirmou, a informação semântica é uma forma de uma mente estar ligada ao exterior. Uma primeira maneira de estar assim ligado é a *percepção*, outra é a *crença*. Explicar o que faz de alguma coisa percepção e crença envolve antes de mais a explicação do que faz de estados físicos *estados intencionais*, e o que faz de estados físicos estados intencionais é precisamente o facto de transportarem *informação* (o facto de providenciarem informação no caso da percepção e o facto utilizarem informação no caso das crenças). Dretske defende que eventos físicos particulares que transportam informação têm efeitos causais: as "crenças", nomeadamente, servem para organizar a informação necessária para controlar comportamentos e para os coordenar com as condições internas e externas nas quais obterão sucesso. Elas são, na imagem favorita de Dretske, mapas por meio dos quais navegamos, e que têm ainda um papel executivo no controlo do comportamento. Segundo Dretske, os organismos passam a ter crenças pelo mesmo processo (a aprendizagem) pelo qual se desenvolve a base neuronal dos comportamentos para cuja explicação as crenças são necessárias. O conteúdo das crenças e desejos pode então aparecer na explicação do comportamento sob a forma de razões.

Para Dretske, como para Millikan, especificar aquilo que constitui "comportamento" é essencial numa teoria da explicação psicológica. Também Dretske oferece uma definição de comportamento, sublinhando que este não é identificável com movimento corporal ou com os estados físicos nos quais tipicamente culmina. O comportamento é um processo causal que tem o movimento corporal como parte, é processo e não produto. Se o comportamento pode ser explicado intencionalmente é porque ele consiste no causar (*causing*) de movimento físico por razões.

A aplicação da teoria do conteúdo informacional à explicação do comportamento envolve, como se disse, uma insistência no papel da aprendizagem e na diferença que a existência de aprendi-

zagem faz na forma que a explicação de movimentos de sistemas assume. Só quando existe *aprendizagem* é que o *comportamento* pode ser explicado através de razões. Estruturas físicas internas adquirem no processo de aprendizagem uma função específica de transporte de informação. Após a aprendizagem, espécimes de um determinado tipo de estruturas físicas quando causados por estimulação "dizem que" – i.e. significam – aquilo que é sua função significarem. A aprendizagem converte estados de *hardware*, por exemplo estados neuronais, que transportam informação em estados com funções naturais, funções naturais próprias e não atribuídas. Isto quer dizer que para Dretske existe uma linha divisória entre *intencionalidade intrínseca* (a intencionalidade dos estados internos que têm a função natural de transportar informação) e *intencionalidade atribuída*. Esta distinção que, como já se viu, a TSI não admite, é justificada por Dretske através do estatuto da aprendizagem. É certo que Dretske admite que a intencionalidade intrínseca pode ser *alcançada*, por exemplo por uma máquina. Mas ainda assim, existe uma diferença importante entre a significatividade (para o próprio sistema e para o controlo do seu comportamento) de estruturas físicas internas ao sistema que têm como função ser suporte de informação, e a significatividade meramente atribuída a partir de fora do sistema. Por isso mesmo, para Dretske, estados mentais com conteúdo conceptual apenas têm um papel na explicação do comportamento de sistemas capazes de aprendizagem. Não é necessário ou apropriado evocar crenças para explicar reflexos ou outros comportamentos geneticamente determinados de sistemas (muito menos aquilo que se passa num termostato) mesmo que todos estes acontecimentos sejam (a partir de fora) descritíveis intencionalmente. É certo que Dretske se serve quer de comportamentos instintivos quer do funcionamento de máquinas para esclarecer a sua teoria das funções. No entanto, embora os conteúdos semânticos sejam definidos a partir da função, de um ponto de vista filogenético as funções começam a existir muito antes da crenças e desejos e à medida que se desce a escala filogenética o mental tem cada vez menos utilidade. O papel do mental na explicação do comportamento só chega com sistemas capazes de aprendizagem. Antes disso movimentos ou comportamentos não têm razões ou intenções embora tenham funções. Com esta teoria Dretske impede o espalhamento generalizado da mentalidade característico da TSI. A aprendizagem é a fonte do conteúdo mental genuíno, mesmo se aquilo que é indicado é um factor extrínseco ao cérebro.

O grande problema desta teoria, apontado por exemplo por Fodor é o *problema da disjunção*[361]. Este consiste no seguinte: se os conteúdos dos crenças são as condições externas que causam as crenças, como se pode saber que que a crença C foi causada por uma específica causa e não por todas as causas posssíveis tendo portanto condições de verdade disjuntivas? Que razões existem para dizer que o conteúdo de um estado mental é A e não A∨B∨C? Apenas evocando uma restrição artificial para separar situações de aprendizagem de outros enfrentamentos do ambiente pelo sistema é possível a Dretske declarar que só em situação de aprendizagem estruturas adquirem uma função determinada, própria, tornando possíveis as falsas crenças. Foi precisamente para responder a esta objecção que Dretske se propôs explicitar de forma mais fundamentada o papel causal ou explicativo da informação[362].

Evidentemente, uma vez definida a informação semântica como correlação entre eventos tornava-se difícil considerar que essa correlação tinha algum poder causal, por exemplo alguma capacidade de despoletar novas crenças. A investigação inicial de Dretske dizia respeito àquilo que constitui, para eventos físicos, transportar informação. Uma conclusão a que chega é que não pode ser a informação semântica a causar o que quer que seja, por exemplo crenças, mas apenas os eventos que transportam informação. Ora, de acordo com Dretske a aprendizagem é o lugar certo para o efeito que (aparentemente) consiste em *causação envolvendo crenças*. Tendo estipulado que o comportamento não é movimento corporal nem o produto final do movimento mas um processo causal que tem o movimento como parte, a explicação psicológica poderá apoiar-se no causar (*causing*) do movimento corporal por razões: «as razões explicam o comportamento não causando-o, mas antes porque o seu conteúdo – aquilo em que se crê e que se deseja – explica o causar»[363]. A aprendizagem é o processo no qual as correlações informacionais relevantes assumem o papel de causa estruturante (*structuring cause*) do comportamento. Reconstituindo os circuitos de controlo do comportamento, a aprendizagem contribui para qualquer futuro comportamento

[361] FODOR 1990: 31, *Semantics Wisconsin Style*. O problema da disjunção é a razão principal pela qual Fodor rejeita o teleofuncionalismo (cf MILLIKAN 1993).

[362] É este o objectivo de DRETSKE 1988.

[363] DRETSKE 1994: 264.

que deles dependa. Em suma, o conteúdo representacional constitui-se ao mesmo tempo que se torna o guia do comportamento dos sistemas. Assim, uma posição como a de Dretske só pode ser realista acerca de estruturas internas que transportam informação e cujo conteúdo semântico pode legitimamente ser evocado para explicar o comportamento através de razões, relegando o instrumentalismo para as margens da verdadeira teoria do conteúdo.

2.2.3.3 *J. Fodor: a teoria nómico-informacional e o atomismo*[364].

A teoria do conteúdo de Fodor é, na sua forma actual, uma teoria informacional-causal, apoiada numa teoria atomista dos conceitos. O atomismo informacional enquanto teoria da natureza dos conceitos consiste na defesa de que os conceitos – por exemplo no caso humano quase todos os conceitos lexicais[365] – são símbolos não estruturados cuja identidade é determinada pela informação que transportam[366] acerca do ambiente. Ao contrário de Dennett, Millikan e Dretske, Fodor rejeita a abordagem teleofuncionalista do problema do conteúdo[367], embora admita que esta produz uma interessante solução para o problema da normatividade. Considera, no entanto, o teleofuncionalismo é incapaz de lidar com a indeterminação trazida pelo problema da disjunção. O apelo à evolução por selecção natural não é, de acordo com Fodor, explicação suficiente dos mecanismos que efectuam a ligação entre o interior de um sistema e o seu exterior, ligação que para Fodor constitui o conteúdo. Embora o darwinismo possa explicar como é que os mecanismos produtores de representações vieram a *estar lá*, foram seleccionados, não explica como é que eles funcionam *agora*. Daí que por

[364] Considerar-se-á sobretudo FODOR 1987, FODOR 1990, FODOR 1994 e FODOR 1998. Na convenção adoptada por Fodor a maiusculação marca um conceito como particular mental e o itálico marca propriedades no mundo (ou significados). As aspas marcam as palavras que exprimem os conceitos que são para Fodor, repita-se, particulares mentais. A doutrina fodoriana dos conceitos é formulada em conjunto com uma metafísica das propriedades. Não se entrará aqui no entanto, na discussão dos aspectos ontológicos subtis relativos a propriedades da teoria dos conceitos de Fodor nem na questão da composicionalidade das Representações Mentais.

[365] Um conceito lexical é expresso por uma palavra única numa língua natural (mais à frente o exemplo utilizado será DOORKNOB, puxador de porta, um conceito lexical em inglês). Não ter estrutura significa não ser constituído por partes semanticamente interpretáveis.

[366] Fodor adopta as linhas gerais da teoria da informação de Dretske.

[367] Cf. por exemplo FODOR 1994. Em FODOR 1987 Fodor explora o que seria uma teoria teleofuncionalista mas acaba por afastá-la.

exemplo a ideia segundo a qual um duplo instantâneo (sem história) de cada um de nós teria uma mente des-conteúdada seja objecto do escárnio de Fodor. Comentando o Swampman de Davidson[368], Fodor afirma: «Talvez seja verdadeiro por definição que as crenças, desejos, ânsias, etc, são constituídos pelas suas histórias; nesse caso, é claro, o Swampman não os tem. Mas, e depois? É intuitivamente plausível que ele tem estados que são as exactas contrapartes a-históricas desses estados e que os estados que ele tem são intencionais»[369]. Para Fodor, no caso do Swampman, todos os mecanismos da representação estariam a funcionar. A teoria do conteúdo defendida por Fodor não é portanto, ao contrário das teorias teleofuncionalistas, uma teoria *histórica*. Para explicar como é que representações internas têm o conteúdo que têm o que é preciso explicar é como elas funcionam no presente. *São por isso os contrafactuais que contam na teoria do conteúdo, não a história.* Basicamente «o que dá conteúdo a representações mentais é alguma coisa acerca de relações causais-nomológicas com aquilo que cai sob elas; por exemplo o que dá a uma representação mental o conteúdo *cão* é alguma coisa acerca das suas instanciações ou espécimes (*tokenings*) serem causados por cães»[370]. Fodor, como Dretske[371], considera que são representações mentais internas deste género que têm intencionalidade original.

Além de "presentista" a teoria informacional-causal de Fodor é *atomista*. A ideia básica do atomismo é que as Representações Mentais (RM) são átomos sem estrutura. O que conduz Fodor ao atomismo é a oposição ao holismo semântico de acordo com o qual o que determina o conteúdo semântico de alguma coisa é o papel inferencial. Segundo a "semântica do papel inferencial" (*inferential role semantics*), se é possível por exemplo «ter o conceito ÁGUA e não ter o conceito H_2O é porque é constitutivo de ter o último mas não de ter o primeiro que se aceite inferências tais como *contem* $H_2O \rightarrow$ *contem H*»[372]. Pelo contrário, de acordo com o atomismo informacional, o conteúdo dos estados mentais é determinado directamente por relações causais entre a mente e o mundo e não por relações inferenciais com outras coisas significativas "dentro da

[368] DAVIDSON 1987.

[369] FODOR 1994: 117.

[370] FODOR 1998: 12.

[371] E contra "vagos rumores vagamente vienenses" (cf. FODOR 1998: 9) de acordo com os qual a intencionalidade original residiria não em representações mentais nem em pensamentos mas nas frases de linguagem pública.

[372] FODOR 1998: 13.

cabeça". O atomismo quanto ao significado vem juntar-se, na teoria da mente desenvolvida por Fodor, ao seu (menos controverso) atomismo quanto à percepção, i.e. à ideia de *modularidade*[373]. Em geral, Fodor declara constantemente o quanto detesta as posições holistas e relativistas na teoria da mente, onde se incluem, como é óbvio, posições "relativistas dogmáticas" como as de Dennett[374].

Admitindo que as questões fregeanas referidas a propósito de *Beyond Belief* são boas questões e que Frege nunca considerou a possibilidade de as tratar de forma naturalista, a tentativa que Fodor leva a cabo, nomeadamente em *Concepts,* com a sua teoria do conteúdo, e que é precisamente uma tentativa de encontrar uma resposta naturalista para essas questões, assume uma especial relevância. Na sua teoria nómico-informacional do conteúdo, Fodor parte dos seguintes princípios gerais, já anteriormente expostos neste trabalho. O comportamento inteligente de sistemas físicos é o objecto da explicação psicológica. Esta é nomológica e intencional, i.e. as leis alcançadas expressam relações causais entre estados especificados sob descrições intencionais. Não pode existir explicação psicológica sem apelo a estados mentais intencionais e a processos mentais racionais. É por isso preciso começar por explicar como são possíveis os estados mentais intencionais e os processos mentais racionais. A sugestão é que os estados mentais intencionais são relações com Representações Mentais. As Representações Mentais (sentenciais) são os sustentáculos primitivos do conteúdo intencional, o lugar da intencionalidade originária rejeitada por Dennett. Por sua vez, o pensamento é computação. A teoria do pensamento como computação é uma teoria sintáctica, e uma teoria sintáctica não é só por si uma teoria do significado. O grande argumento a favor da ideia de pensamento como computação é a preservação de propriedades semânticas (nomeadamente a verdade) nos processos mentais. Aliás, Fodor faz esta afirmação de um modo mais preciso: ele considera que afirmar que o pensamento é computação é afirmar que os mecanismos de implementação da psicologia (mecanismos de implementação que devem existir para qualquer ciência não básica) são computacionais e não por exemplo imediatamente biológicos. As leis intencionais da psicologia são implementadas por processos computacionais. Processos computacionais são definidos sobre objectos sintacticamente estruturados, as Representações Mentais. É o carácter sintáctico das

[373] Cf. FODOR 1983 e FODOR 1990, Parte II.
[374] Cf. FODOR 1990: 3.

Representações Mentais que caracteriza a Hipótese da Linguagem do Pensamento. Relativamente à semântica das Representações Mentais, Fodor tem, como Dretske e a partir de Dretske, trabalhado a ideia segundo a qual o significado é informação e nada mais do que informação. Assim, as propriedades semânticas são externalistas e latas. Isto significa que para Fodor o que dá às Representações Mentais o conteúdo que elas possuem são relações causais, nomológicas, com aquilo que cai sob elas e que está "fora da cabeça".

O problema é evidentemente conjugar todas estes teorias com a TRM que Fodor há muito vem a desenvolver. Pelo menos aparentemente a ideia de conteúdo intencional *lato* é difícil de conciliar com a existência de *leis psicológicas intencionais, cuja implementação é computacional.* Ora, se a aproximação entre conteúdo e informação é relativamente pouco polémica na discussão, usualmente junta-se-lhe a ideia de que o conteúdo é determinado por um ingrediente inferencial a mais, que lhe dá o carácter holista. No entanto é precisamente a esta relação entre intencionalidade e holismo, à primeira vista tão natural, que o atomismo da teoria informacional-causal se opõe. Para Fodor é óbvio que a teoria da explicação psicológica não é compatível com uma metafísica holista do significado. De facto, a ideia central da semântica do papel inferencial é errada para Fodor, em grande parte devido à circularidade admitida por quem ao mesmo tempo aceita que o pensamento é computação, i.e. relação causal entre símbolos que respeita o conteúdo, e que o conteúdo está ligado ao papel inferencial, i.e. a computações. Uma semântica do papel inferencial conjuntamente com uma teoria computacional do pensamento só pode produzir circularidade. Fodor pretende evitar a circularidade com a ideia de que o conteúdo é informação e nada mais.

Evidentemente o grande problema para uma teoria nómico-informacional advém do facto de esta confiar ao próprio mundo a tarefa de determinar o conteúdo semântico das Representações Mentais. O primeiro obstáculo é o facto de o funcionamento sintáctico das representações mentais e a informação que constitui o seu conteúdo, em princípio harmonizados, poderem deixar de estar harmonizados de várias maneiras. É necessário explicar como é isso possível. Eles podem deixar de estar harmonizados essencialmente devido a uma *situação de Putnam* (por um lado, numa Terra Gémea, onde as mesmas Representações Mentais correspondem a conteúdos diferentes, por outro no uso de um "mesmo conceito" entre um perito e um leigo, por exemplo "ALCOÓL" ou "AÇÚ-

CAR", situação na qual existe uma diferença de conteúdo) ou devido a uma *situação de Frege* (em que representações diferentes como "ESTRELA DA MANHÃ" e "ESTRELA DA TARDE" correspondem a um mesmo referente no mundo). É esse grau de liberdade ou desarmonia das Representações Mentais relativamente ao conteúdo, considerado como um "apontar para o exterior", que Fodor pretende adaptar à TRM. Consegue-o através da ideia de que Modos de Apresentação asseguram mecanismos que garantem que as situações de Putnam e as situações de Frege não ocorrem (ou pelo menos não ocorrem muito frequentemente).

As situações de Putnam são as mais manejáveis: «Do ponto de vista da semântica informacional [*aquilo que se passa*] é perfeitamente normal: o facto de os meus conceitos de olmo e de ácido terem o conteúdo que têm depende da existência de mecanismos que os correlacionam fiavelmente com instanciações de *olmidade* e *acidez*, respectivamente»[375]. As situações de Frege são mais complicadas e é para as tratar que Fodor afina a noção de Modo de Apresentação (MAP), a noção mais directamente concorrente com o mundo nocional da TSI. Um MAP é uma Representação Mental, um símbolo em Linguagem do Pensamento, individuável pela forma. Com a mais recente reformulação da TRM em termos de MAP Fodor vem a declarar que a ideia de conteúdo estrito é afinal supérflua.

A intenção de Fodor em *Concepts* é fundir a ideia de Turing segundo a qual pensamento é computação, a ideia de Dretske segundo a qual o significado é informação, com as ideias de Frege acerca do sentido, considerado como modo de apresentação. Segundo Fodor, «a ideia básica da TRM é que a história de Turing acerca da natureza dos processos mentais providencia os candidatos a modos de apresentação que a história de Frege acerca da individuação de estados mentais independentemente requer»[376]. No medida em que para Fodor uma teoria dos conceitos é o núcleo da teoria do conteúdo[377] exemplificar-se-á aqui por meio da teoria dos conceitos os princípios da teoria fodoriana do conteúdo. A tese central de *Concepts* é que os conceitos são particulares mentais. Note-se que para Fodor uma teoria do conteúdo nunca poderia ser uma teoria do uso da linguagem (que por sua vez se reduziria a uma teoria das capacidades linguísticas que por sua vez seriam constrangidas por aquilo que é manifesto no comportamento do falante) sob pena de não ser minimamente *explicativa*. Para Fodor, os conceitos que um sistema tem são independentes das capacidades epistémicas deste[378] (considerar que os conceitos não são independentes das capacidades epistémicas seria um movimento em direcção à semântica do papel inferencial que Fodor, como se afirmou, rejeita). O problema, como já foi dito, é que se o significado consiste não no papel inferencial mas em informação – i.e. se o que dá conteúdo a representações mentais é alguma relação causal-nomológica com as coisas que caem sob elas, que as causam – representações co-referenciais, deveriam ser "sinónimas" e nem sempre o são. São estes os casos de Frege. É com este resíduo que Fodor procura lidar, sem poder considerar que o ingrediente a mais tem alguma coisa a ver com inferências que o sistema está preparado para fazer. Fodor propõe a noção de MAP para distinguir conceitos diferentes mas co-referenciais (como ÁGUA e H_2O). Um MAP de Fodor é algo que está na cabeça. Os MAP de Frege, que eram sentidos (*senses, Sinne*), serviam para explicar como é que era possível ter um mas não outro de dois conceitos co-referenciais. Estes MAP eram no entanto considerados por Frege como objectos abstractos, não mentais. Fodor pretende manter a ideia segundo a qual são os MAP que explicam como é possível ter um mas não outro de dois conceitos co-referenciais. No entanto pensa que para manter coerentemente essa ideia é necessário deixar de considerar os MAP como sentidos[379] e como objectos abstractos não mentais[380]. Os

[375] FODOR 1994: 35.

[376] FODOR 1998: 22.

[377] Aliás, Fodor pensa que o que decidirá a questão entre teorias informacionais do conteúdo e as que chama teorias do papel inferencial é o facto de apenas as primeiras possibilitarem uma teoria atomista dos conceitos, e Fodor pensa que há razões independentes para acreditar que só uma teoria atomista será uma boa teoria dos conceitos.

[378] Nos termos de Fodor, as duas questões (a questão acerca dos conceitos que se tem e a questão acerca das capacidade epistémicas que se tem) são conceptualmente e metafisicamente independentes.

[379] Segundo Fodor, a única coisa que sabemos sobre sentidos é que sinónimos os partilham. Se modos de apresentação (MAP) são sentidos e a única coisa que distingue conceitos co-referenciais, conceitos sinónimos devem ser idênticos, sendo impossível pensar um sem pensar o outro (este é o teste de substituição para modos de apresentação). Mas pode-se duvidar que João compreenda que As são Bs mesmo se não se duvida que João compreenda que As são Cs, sendo Bs e Cs sinónimos (cf. FODOR 1998: 16). Basicamente Fodor defende que se conceitos com o mesmo sentido podem ser diferentes MAP, MAP não podem ser sentidos. Individuar MAP deve ser como individuar formas de palavras mais do que como individuar significados.

[380] Em FODOR 1998 Fodor apresenta um argumento complicado no sentido de provar que MAP não podem ser objectos abstractos. Retém-se aqui apenas que Fodor pensa que se MAP podem individuar conceitos (a função para que foram criados....) é precisamente por serem objectos mentais, disponíveis para serem causas próximas de processos mentais, ao contrário dos referentes. De acordo com Fodor, Frege pensa erradamente que MAP têm que estar fora da mente (ser objectos abstractos) para os conceitos serem públicos. Uma teoria informacional deixa a "publicidade" a cargo do mundo.

MAP de Fodor são veículos do pensamento, veículos com que se pensa sem que, evidentemente, se pense acerca deles. São estes veículos que permitem conceber como se pode ter dois conceitos coreferenciais sem se saber. Uma vez afastadas as considerações de Frege sobre objectos abstractos e a sua captura de modo a admitir que os MAP devem ser algo que faz diferença em organismos, Fodor defenderá que os MAP são Representações Mentais sentenciais.

A teoria dos conceitos é em parte uma teoria dos MAP, em parte uma justificação do anti-descritivismo. O atomismo conceptual evita o inatismo – ao qual Fodor é aparentemente conduzido por não admitir uma aquisição indutivo-inferencial de conceitos – ao apoiar-se numa tese anti-descritivista (a que Fodor chama não-cognitivista) da posse dos conceitos[381]. Para Fodor, como para Millikan, possuir um conceito não é conhecer alguma coisa, mas sim ser capaz de a detectar e de a identificar. De resto, a explicação da aquisição "indutiva" de conceitos por teste de hipóteses só é plausível se se assumir uma história "cognitivista ou descritivista" da posse de conceitos, segundo a qual possuir um conceito é conhecer alguma coisa. Para Fodor, uma vez que adquirir um conceito não é aprender indutivamente coisa alguma mas sim ficar nomologicamente ligado (*locked*, termo que Fodor vai buscar à etologia) a algo, a posse de um conceito é independente da posse de outro conceito. Para passar a ter um conceito o que é preciso é ter o tipo certo de experiências de modo a ficar ligado a alguma coisa no mundo. Essa experiência não consiste, no entanto, no teste de hipóteses.

Em *Concepts*, Fodor exemplifica estas teses servindo-se do exemplo de DOORKNOB, um primitivo desinteressante que é o exemplo por excelência da teoria dos conceitos apresentada. Mantém-se o inglês do exemplo porque o termo tem que ser primitivo, o que não aconteceria com PUXADOR DE PORTA. "Doorknob" não é uma coisa natural no mundo e o termo que lhe corresponde não é sequer aparentemente um termo primitivo. DOORKNOB é precisamente indefinível de uma maneira não interessante (i.e. sem que isso envolva qualquer coisa de profundo, ou de metafisicamente primitivo). O explicação da aquisição do conceito DOORKNOB por algum sistema cognitivo passa por alguma coisa que se passa "na cabeça" do sistema em função da interacção do sistema com

[381] Os conceitos são RM, as RM não têm estrutura: são átomos (embora sejam constituintes de representações mentais estruturadas sobre as quais se cumprem as computações).

doorknobs. Se ser um doorknob não é ser uma coisa natural, a propriedade "*ser um doorknob*" é acerca de nós, humanos, é qualquer coisa acerca das nossas mentes. Isso não significa que não exista a propriedade *ser um doorknob* a que um humano pode ficar ligado, adquirindo o conceito DOORKNOB. Precisamente, os humanos adquirem o conceito DOORKNOB ficando ligados a essa propriedade. Existem aliás muitíssimos conceitos de aparência semelhantes. Por definição, para Fodor, ter um conceito é, como se disse, estar ligado à propriedade correspondente no mundo e se esse conceito é um conceito de aparência, a propriedade é constituída pela maneira como as coisas que a têm "embatem em" sistemas cognitivos, por exemplo humanos.

Mesmo os empiristas sempre admitiram o inatismo do *sensorium* e o que Fodor faz para explicar a aquisição de conceitos é apenas generalizar a condição do *sensorium*. O inatismo do *sensorium* não é inatismo de alguma coisa com conteúdo intencional. Tudo o que tem que ser inato para que o conceito VERMELHO seja adquirido são mecanismos (fisiologia) que determinam que as coisas vermelhas nos apareçam assim. Do mesmo modo, ficar ligado a uma propriedade só requer que as coisas que a possuem nos apareçam assim. O inatismo associado ao atomismo informacional não é um inatismo de conceitos (de algo "intencional") mas de mecanismos (o que parece natural dada a inspiração etológica do processo de "ficar ligado"). Ao contrário da tese que se tornou um motivo de chacota relativamente a *The Language of Thought*, nem mesmo os conceitos primitivos têm que ser inatos. Em geral a moral de *Concepts* é que talvez não existam sequer Representações Mentais inatas, embora tenha que haver muito de inato para que sistemas adquiram conceitos. O que é importante numa teoria não cognitivista dos conceitos é explicar a possibilidade de conceitos *primitivos* e o que Fodor propõe é que eles são átomos, Representações Mentais sem estrutura. Em *Concepts* Fodor faz notar que os conceitos de géneros naturais, objecto usual da discussão acerca do conteúdo de conceitos, não servem senão para obscurecer esta: conceitos de géneros naturais são muito tardios no desenvolvimento conceptual humano, têm que emergir do funcionamento descrito relativo a conceitos que ligam os sistemas a propriedades que são dependentes da mente e importantes para o sistema.

O atomismo informacional e o anti-descritivismo que este envolve são núcleos essenciais da teoria do conteúdo de Fodor. É o

atomismo informacional que afasta a semântica do papel inferencial e permite o enquadramento da questão do conteúdo na TRM. Isto parece um avanço em relação à teoria do conteúdo ligada à TSI, que não inclui uma teoria da individuação fina do conteúdo, exactamente devido ao facto de o interpretativismo ser holista.

2.2.3.4 *Sentencialismo, eliminativismo ou interpretação.*

A individuação de veículos internos e o apoio da explicação psicológica que ela permite é o maior ponto de discórdia entre Dennett e os três autores aqui referidos, uma vez que há inegáveis pontos de consenso entre as teorias. Antes de mais, e essencialmente, elas confluem no *externalismo*. Este redunda pelo menos nas seguintes teses: (1) significar não é um funcionamento epistémicamente controlado nem necessariamente relacionado com a consciência ou com a consciência de si, (2) o mundo (exterior) é o maior responsável pela fixação do conteúdo de representações mentais, de uma forma que passa ao lado do sistema no qual ocorrem os veículos do significado. Para além deste consenso mínimo, a discussão foca o estatuto dos veículos do significado e é aí que as posições divergem. Nas condições referidas pela TSI quanto ao mundo nocional, o significado não parece poder realizar trabalho e é nesse sentido que se pode falar da impotência do significado ou epifenomenalismo que é a última palavra da TSI. É certo que nenhum dos restantes autores propõe que os significados causem directamente o que quer que seja. No entanto, tem-se Ícones Intencionais Internos, veículos internos que transportam informação, a que alguns, nomeadamente Fodor, consideram legítimo chamar representações e que são para-o-sistema, para guiar o comportamento, independentes de qualquer interpretação e como tal evocáveis em algum tipo de *explicação*. A diferença entre Dennett por um lado e Fodor, Millikan e Dretske por outro, reside no facto de Dennett pensar que a terminologia das representações e do conteúdo, aplicada a ocorrências no interior de sistemas cognitivos, é sempre já interpretativa. Repetindo o princípio atrás referido, não é possível ter "texto" (neste caso, natureza de representação e conteúdo de representação) sem interpretação e é isso que as noções de Fodor, Dretske e Millikan presumem emboras eles não o admitam. Por outro lado, apenas a ideia de de representações reais no sistema,

independentes de qualquer interpretação ou atribuição, possibilita a intencionalidade intrínseca que a TSI exclui e os restantes autores admitem.

Recapitulando, o posicionamento da TSI quanto à intencionalidade envolve uma posição quanto à natureza de representação e uma posição quanto ao conteúdo de representações, as quais fogem de várias maneiras ao realismo. A posição quanto à natureza de representação é definida contra a TRM de Fodor, que é considerada como preconceituosamente realista e sentencialista. Neste momento e ainda tendo Fodor em mente, já se viu como é que uma posição realista e sentencialista quanto à natureza de representações se prolonga numa teoria do conteúdo. De acordo com Fodor, os modos de apresentação são Representações Mentais sintaticamente estruturadas que estão na cabeça e cuja instanciação tem o conteúdo que tem em virtude de relações nómico-informacionais com o mundo. Do ponto de vista assim definido, os problemas da posição da TSI quanto à intencionalidade são claros: padrões ditos "reais" mas que afinal não passam de estados intencionais atribuídos e mundos nocionais interpretativos atribuídos de forma holista não podem ter eficácia causal local, ao contrário do que acontece com modos de apresentação sentenciais (ou, de resto, também com os veículos de Dretske ou com os Ícones Intencionais de Millikan), o que os torna inúteis e redundantes do ponto de vista da explicação psicológica. Ora, se as descrições intencionais não são explicativas porque não desaparecem elas simplesmente da teoria da cognição? Não seria mais coerente que o interpretativista fosse um eliminativista acerca da intencionalidade? É por essa razão que Fodor põe em causa o entusiasmo relativamente às descrições intencionais de alguém como Dennett, que não pretende falar de causas, nega a possibilidade de explicação do comportamento e nega a possibilidade metafísica de leis intencionais. Aos olhos de Fodor, seria preferível que Dennett, em vez de defender a utilização de descrições e previsões intencionais não nómicas se rendesse definitivamente à "Califórnia do Sul", i.e. ao eliminativismo, à substituição das explicações intencionais pela neurociência como propõem os Churchland, ou pela teoria física de outro substracto de mentalidade. Ao utilizar todo o aparato do realismo intencional (as crenças, a racionalidade) sem conceder realidade a nada, Dennett seria a demonstração viva da impossibilidade de compromisso quanto ao problema da explicação psicológica: apenas o realismo inten-

cional e o eliminativismo permitem posições totalmente desprovidas de ambiguidade perante o problema da explicação psicológica.

Mas precisamente, a TSI situa-se num meio termo: não aceita *explicações* intencionais mas também não aceita eliminar a panóplia das noções intencionais. No entanto e essa foi a razão pela qual se descreveu aqui outras posições teleofuncionalistas além da de Dennett, o teleofuncionalismo permite uma posição acerca da intencionalidade que não é sentencialista, não fecha a porta à explicação (mesmo com as salvaguardas de Millikan acerca das ciências biológicas), é naturalista e não evoca a consciência para oferecer uma teoria do significado. Simplesmente para isso é necessário um realismo acerca de veículos e funções que é problemático para a TSI.

O que parece estranho em primeiro lugar quando se compara o teleofuncionalismo que a TSI alberga com outras teorias teleofuncionalistas é o facto de Dennett continuar a evocar a interpretação para a compreensão do conteúdo, nomeadamente através da ideia de mundo nocional. Dennett não é o único filósofo da mente que propõe que a teoria da mente deve ter a forma de uma teoria da interpretação: essa é uma posição quineana muito difundida. No entanto os autores que também apresentam a teoria da mente sob a forma de uma teoria da interpretação (por exemplo Davidson) não pretendem, ao contrário de Millikan, Dretske, Fodor e Dennett, elaborar uma teoria naturalista da psicologia. Uma teoria *naturalista* da interpretação parece, pelo menos à primeira vista, uma abordagem híbrida. O problema do conteúdo de estados mentais parece mais tratável do ponto de vista realista – ao contrário do que se passa com a TSI, não existe nas posições realistas uma cláusula que especifica a existência de pelo menos um intérprete, pelo menos um ponto de vista, que permite a atribuição de intencionalidade, e que fica aí para ser explicada, não se vê bem como. Como nota Fodor, «a maioria dos filósofos que abordam a intencionalidade a partir da evolução são correspondentemente realistas(...) acerca do conteúdo (é o caso de Millikan e de Dretske)»[382]. Curiosamente, não é esse o caso de Dennett, embora ele frequentemente remeta o desenvolvimento das suas próprias posições para o trabalho de Millikan, que o teria despertado do sono semi-dogmático do Racionalismo do Significado. Para os autores cujas posições se pretendeu aqui contrastar com a de Dennett, a intencionalidade é

[382] FODOR & LEPORE 1992: 145.

real, e sendo real, como é previsível que ela não faça parte da "lista do físico" (a intencionalidade não aparecerá ao lado de *spin* e *charm* – como diz Fodor «*it doesn't go that deep*[383]») ela terá que ser fundamentalmente outra coisa (informação, Linguagem de Pensamento, Ícones Intencionais) mas alguma coisa no interior dos sistemas cognitivos eles próprios e nunca apenas interpretação. Evidentemente, do ponto de vista de Dennett, são os seus adversários que não reconhecem o grau de interpretação inerente a estas noções. E de facto, o grande problema da TSI não é o interpretativismo, que permite uma descrição adequada do estatuto do mundo-para-o-sistema, cuja análise se continuará no próximo capítulo. Mas há algo que falta à TSI e que os restantes autores, não interpretativistas, reconhecem. Provisoriamente nomear-se-á isso que falta como *Do-It-Yourself-Understanding*, a expressão com a qual Dennett ironicamente caracteriza «o processo alquímico pelo qual mera informação se transforma no entendimento genuíno característico de apenas alguns sistemas cognitivos»[384]. Este entendimento genuíno acerca do qual Dennett ironiza é o entendimento do sistema para o próprio sistema e ele não pode ser ignorado.

Duas coisas têm que ser corrigidas ou aperfeiçoadas na TSI tal como ela neste momento se configura: a cegueira perante o entendimento genuíno (essa correcção ficará para o Capítulo 6) e uma ambiguidade quanto ao *design* que se procura esclarecer em seguida mas acerca da qual a última palavra se encontra também no Capítulo 6. De acordo com a TSI a atribuição de intencionalidade é justificada pelo *design*. No entanto, também de acordo com a TSI o reconhecimento de *design* pressupõe a intencionalidade. A raíz do problema da posição de Dennett quando comparada com as posições de outros teleofuncionalistas é, aliás, a sua ambiguidade quanto ao estatuto do *design*.

2.3 *Dennett e o design: o foco das tensões internas da TSI.*

2.3.1 *Oscilação entre design real e design como interpretação.*

Na medida em que a teoria dennettiana do conteúdo semântico remete a questão do conteúdo para o funcionamento de dispositi-

[383] FODOR 1987: 97.
[384] DENNETT 1998b: 60.

vos com funções, ela é uma teoria teleofuncionalista entre outras. No entanto, as teorias teleofuncionalistas são em geral realistas quanto a representações internas e o interpretativismo da TSI impede este último passo. De facto, as oscilações de Dennett entre realismo e interpretativismo quanto às representações internas devem-se a conflitos inerentes à sua teoria das funções ou do *design*. Dennett tanto considera o *design* "relativo" a uma interpretação (quando afirma por exemplo que a explicitação da função específica de um determinado dispositivo supõe a EI e que o adaptacionismo é isso mesmo) como considera que o *design* é real, i.e. uma característica de dispositivos físicos resultantes de selecção natural ou desígnio humano, como organismos ou artefactos.

No que diz respeito à relatividade do *design*, a aproximação entre a EI e a ED é a seguinte: o «adaptacionismo e o mentalismo (a TSI) não são teorias no sentido tradicional (...) são pontos de vista ou estratégias que servem para organizar dados, explicar correlações e gerar questões para colocar à natureza (...) se fossem teorias de molde clássico, a objecção segundo a qual elas cometem petições de princípio (...) seria irrefutável»[385]. Quer a EI quer a ED são, então, apenas estratégias interpretativas e não teorias acerca da "verdadeira natureza" de alguma coisa. Ora, alguma coisa que não existe, que não é verdadeiramente real, algo cuja natureza é interpretativa, não pode ser *explicativa*.

Em *Darwin's Dangerous Idea* (DDI) o objecto propriamente filosófico de Dennett é precisamente a questão do *design* nos seus vários aspectos (a fonte do *design*, o espaço do *design*, o estatuto do *design*). Dennett pretende antes de mais defender o adaptacionismo das críticas que lhe são dirigidas, mostrando que é realmente possível falar de funções biológicas. O problema é que a interpretação que Dennett faz do darwinismo o compromete com uma resposta dupla, ou ambígua, à questão "será o *design* relativo ou real?" e essa ambiguidade repercute-se depois na teoria do conteúdo. Considerando os produtos da selecção natural, Dennett afirma por vezes que o seu *design* depende de uma interpretação, outras vezes que o *design* é real, produzido por processos sem intenção, e obviamente sem intérprete. Da mesma forma, é notório que ao longo dos seus escritos acerca de intencionalidade Dennett afirma alternadamente que (i) a racionalidade é uma interpretação idealizante

[385] DENNETT 1987: 265.

e que (ii) a racionalidade é explicada pelo *design*. Aparentemente, a própria definição de racionalidade como *design* óptimo ameaça fazer colapsar a postura "intencional" (quineana, interpretativa) numa postura realista: negar que um dispositivo desenhado para a racionalidade seja realmente racional seria como negar que um olho desenhado para ver realmente veja. No entanto e por outro lado, Dennett afirma que a especificação das funções dos dispositivos físicos adaptados "está nos olhos do intérprete". É a partir dessa perspectiva que considera que existe uma indeterminação da função biológica, do mesmo modo que existe uma indeterminação do conteúdo mental. Em suma, é difícil saber qual é exactamente o alcance da afirmação que Dennett faz frequentemente segundo a qual o *interpretativismo está para a psicologia como o adaptacionismo está para a biologia*: ela tanto pode significar que as funções biológicas dependem de uma interpretação como que os dispositivos psicológicos são adaptações reais.

As tensões internas à TSI são para muitos críticos sintomas de incoerência. A instabilidade da concepção do *design* mostraria a impossibilidade de conjugar uma posição teleofuncionalista com uma postura quineana que considera a linguagem mentalista meramente interpretativa. Poder-se-ia perguntar a Dennett o seguinte: se em última análise a possibilidade de adopção da EI é remetida para o *design*-para-a-racionalidade de determinados sistemas – o que supõe que o *design* é pelo menos por vezes real – então porque não se há-se considerar a intencionalidade também ela real? Esta é evidentemente a posição de alguém como Fodor. Fodor comenta da seguinte forma o modo como a TSI considera que o estatuto hermenêutico das descrições intencionais deriva do estatuto hermenêutico das funções biológicas: «Não queremos insistir que esta história é circular, mas também não queremos insistir que não é»[386]. Por outro lado, se a referência ao *design* biológico não envolve um realismo acerca de funções, uma vez que falar de funções implica já uma hermenêutica dos artefactos ou dispositivos naturais, se quer a EI quer a ED dependem de um intérprete, porque não hão-de ser todas as estratégias, incluindo a EF, entendidas como dependendo de um ponto de vista interpretativo? Esta seria a sugestão de alguém como R. Rorty, que considera que a solução está implícita na própria ter-

[386] FODOR & LEPORE 1992: 148. Ao que se acrescenta: «No doubt interpretation can do a lot – hermeneutics is everywhere these days. Maybe interpretation can somehow determine teleology or selectional history (though with such friends, Darwin hardly needs enemies)».

minologia de Dennett, na alusão a "estrátegias" (EI, ED, EF) que é parte integrante da TSI. Aparentemente, se o estatuto do *design* fosse desambiguado, a TSI não subsistiria, sendo obrigada a inclinar-se para uma destas posições (ou para o realismo intencional ou para uma teoria da interpretação generalizada). Para que a sua posição não fosse instável, Dennett necessitaria de distinguir claramente o interpretativismo quanto à intencionalidade do realismo quanto ao *design*, o realismo quanto ao *design* do realismo envolvido na EF, o relativismo intencional do relativismo quanto ao *design*, o relativismo quanto ao *design* do relativismo quanto à própria EF. Ora, o que se verifica é que a TSI fornece argumentos para fazer e para desfazer estas distinções.

Dennett queixa-se frequentemente de que a sua posição acerca de intencionalidade é mal compreendida por não ser enquadrável em etiquetas. Propôs mesmo a expressão "realismo moderado" para afastar a etiqueta de "instrumentalismo" que tinha sido colada à TSI. No entanto o problema pode não estar nas etiquetas e sim na posição de Dennett. Em alternativa, é possível que o verdadeiro problema não seja a posição de Dennett quanto ao estatuto do *design* mas o próprio estatuto do *design*: a ambiguidade na TSI reside no facto de procurar conceber que a atribuição de *design* aos objectos com *design*, quando comparada com a atribuição de intencionalidade é realista, mas quando comparada com a EF é de certa forma "interpretativa". Talvez essa ambiguidade corresponda à própria realidade, sendo uma dificuldade inerente ao pensamento das funções, que em virtude da posição teleofuncionalista se alarga até ao pensamento do conteúdo.

Em suma, o apelo à evolução inclina a TSI para o realismo e empurra a EI em direcção à ED, o que é a posição normal dos teleofuncionalistas. No entanto, como L. Rudder Baker nota, não se pode consistentemente supor que a racionalidade é alcançada por selecção natural sendo um realista acerca dos produtos da selecção natural e um instrumentalista acerca da racionalidade, nem se pode deixar de verificar, por outro lado, que «o preço de tomar as características do *design* como dependentes de uma estratégia (é tornar-se) instrumentalista acerca das teorias da selecção natural»[387]. O que parece certo é que o apelo teleofuncionalista à evolução para dar conta das categorias psicológicas coloca a intencionalidade, a racionalidade e o *design* no mesmo barco, inclinado para um lado ou para o outro.

[387] RUDDER BAKER 1994: 339.

2.3.2 Darwin's Dangerous Idea: *a ideia de Darwin e o evolucionismo generalizado. Realidade e relatividade do design, de novo.*

A fundamentação da teoria do conteúdo é remetida pela TSI para a biologia (ou para um evolucionismo generalizado, uma vez que os princípios de uma teoria geral do *design* valem igualmente para sistemas físicos não biológicos). Todo o desenvolvimento de *design*, natural ou artificial, é regido por princípios darwinistas. A capacidade humana de utilizar significados, i.e. a mente, resulta dos mesmos processos darwinistas de criação de *design* que regem a existência de todas as entidades com *design*. Em DDI, Dennett explicita os princípios de uma teoria geral do *design* que subjazem à sua teoria da mente. A ideia de Darwin a que o título de DDI alude é a *ideia de evolução por selecção natural como processo algorítmico*[388]. Ela oferece uma explicação da origem de todo o *design*, unificando o domínio das ciências da vida e da cognição (i.e. o domínio teleológico) com o domínio das ciências físicas, com consequências que, de acordo com Dennett, por vezes passam despercebidas ou são voluntariamente evitadas. Considerar a evolução por selecção natural como um algoritmo, é considerá-la como um processo formal de cuja implementação se pode esperar resultados, um processo mecânico, neutro quanto ao substracto, sem qualquer inteligência subjacente, e que de forma fiável produz resultados. É a partir deste processo que todo o *design* surge por acumulação.

Em DDI Dennett pretende que todo o *design*, todas as formas funcionais e as suas operações num ambiente, inclusive o *design* para o mental, explora um mesmo espaço das possibilidades. Dennett chama "espaço do *design*" ao conjunto formado por todas as formas funcionais actuais já existentes e que existirão. Uma das implicações do darwinismo é a unicidade do espaço do *design*, o que significa que tudo aquilo que é desenhado (*designed*) e actual está nesse espaço unido a tudo o resto. Corpos, mentes, produtos das mentes (desde pensamentos a artefactos) estão unidos no espaço do *design* sendo a criação do *design* um resultado de movimentos de exploração desse espaço. Um aspectos decisivo da ideia de Darwin é, repi-

[388] DENNETT 1995: «Here, then, is Darwin's dangerous idea: the algorithmic is the level that best accounts for the speed of the antelope, the wing of the eagle, the shape of the orchid and all the other occasions for wonder in the world of nature. (...) No matter how impressive the products of an algorithm, the underlying process always consists of nothing but a set of individually mindless steps succeeding each other without the help of any intelligent supervision; they are automatic by definition: the workings of an automaton».

ta-se, permitir pensar que também o *design*-para-o-mental pode vir à existência em função de um processo mecânico e sem propósito. Deixa assim de ser inimaginável a existência de "teleologia actual" sem pensamento anterior, deixando de exercer-se a atracção irresistível do argumento segundo o qual se existe teleologia tem que ter existido/existir pensamento (a ideia de *"mind-first"*),[389] que ainda enredava por exemplo D. Hume nos *Diálogos sobre a Religião Natural*[390]. Aliás, o perigo, a corrosão associada ao ácido universal de Darwin é precisamente a ausência de mentalidade (*mindlessness)* do processo algorítmico que produz *design*, e nomeadamente *design* para a mente. Darwin descobriu o carácter algorítmico da produção do *design*, um processo que decorre na Terra há milhões de anos e que envolve todos os seres vivos, os artefactos, as linguagens, os sistemas inteligentes artificiais, etc. A ideia de Darwin não se circunscreve à biologia: a biologia é apenas o domínio do *design* natural (uma das teses fundamentais de DDI é aliás que a IA, cujos produtos se situam no mesmo espaço de *design* que os produtos da inteligência natural, é um descendente directo da ideia de Darwin).

As ideias de Darwin sobre o *design* explicam como pode este ser criado sem que no início tenha havido inteligência ou intenção. Elas supõem não apenas que o espaço do *design* é único, como que o *design* pode ser acumulado, perdido, melhorado, que podem existir movimentos forçados no espaço do *design*, que a inovação de *design* é mais dispendiosa do que o reaproveitamento, etc. São estas ideias que sustentam aquilo a que Dennett chama *o assalto de Darwin à pirâmide cósmica*, a ideia segundo a qual haveria uma orientação do base para o topo da árvore da vida[391], sendo superior aquilo que estivesse no topo. O darwinismo não nega evidentemente a possibilidade de sofisticação do *design*: existe uma dimensão vertical da pirâmide que se traduz por exemplo no facto de uma mente humana estar, nela, acima da mente de um pássaro. O que Darwin afirma é que tal sofisticação não é função de *skyhooks* e sim de "guindastes".

[389] A ideia de *mind-first* faz supor que tudo aquilo que tem função (i.e. *design*) revela uma intenção e portanto deve ter tido origem em algo de intencional, i.e. numa mente.

[390] Cf. HUME [1779].

[391] A árvore da vida é um diagrama das trajectórias no tempo das relações de descendência entre todos os seres vivos alguma vez existentes (DENNETT 1995: 85). À existência de cada organismo corresponde uma linha de tempo, da qual partem ou não outras linhas. Não é objecto de controvérsia o facto de toda a vida alguma vez existente no planeta se situar neste gráfico. O que é controverso são os princípios que permitem explicar os padrões do gráfico, por exemplo os princípios relativos à especiação. Segundo Dennett tais padrões só se tornam identificáveis com apoio em idealizações.

Skyhooks e guindastes são alternativas quanto aos instrumentos para o movimento no espaço do *design*. *Skyhooks*[392] ou ganchos do céu seriam dispositivos imaginários capazes de elevar a partir de cima. Guindastes são dispositivos intermediários – como a simbiose, o sexo e a aprendizagem – que auxiliam e aceleram o desenvolvimento do *design* através da evolução por selecção natural. Afirmar que a selecção natural é um algoritmo, é afirmar que não são necessários *skyhooks* para explicar o *design*. É no entanto uma nostalgia de *skyhooks* que Dennett encontra nos autores mais insuspeitos (desde S.J.Gould a N. Chomsky e J. Fodor) e um dos propósitos de DDI é esclarecer a diferença entre o bom e o mau reducionismo na consideração do *design*. O mau reducionismo oblitera a importância dos guindastes na produção de *design* e a dimensão "vertical" da pirâmide, enquanto o bom reducionismo (aquele que o próprio Dennett defende) apenas sublinha a inexistência de *skyhooks* na evolução por selecção natural.

Se é verdade que todo o *design* explora um mesmo espaço de possibilidades, no entanto não é verdade que apenas o *design* actual seja possível. Para explicar os padrões no espaço de *design* actual é preciso considerar tipos de possibilidades "encaixados" uns nos outros: a possibilidade lógica, a possibilidade física e a possibilidade biológica e histórica. Noutras palavras, para conceber o estatuto das coisas desenhadas actuais é preciso concebê-las sobre um fundo no qual se encontram as maneiras como as coisas poderiam ter sido, não poderiam ter sido e têm necessariamente que ser. A possibilidade lógica é definível como inexistência de contradição. O âmbito da possibilidade física é definido pelas leis naturais. Existem, é certo, problemas acerca da relação e da distinção entre as duas. No entanto, a ideia de Darwin é uma ideia acerca da possibilidade biológico-histórica, ou seja acerca da possibilidade do *design*. A possibilidade biológica é pensada por Darwin como *acessibilidade num espaço de busca na "biblioteca de Mendel"*. Esta é definida em DDI como o espaço lógico de todos os genomas. A necessidade biológica é correlativamente identificável com movimentos forçados no espaço do *design*. Considerando a necessidade de alguma coisa, o facto de haver uma só maneira como ela pode ser, ou poderia ter sido, pode ter razões profundas (se tais razões forem físicas) ou mais superficiais, se tais razões forem históricas, nomeadamente biológi-

[392] Por excelência um processo *mind-first* (DENNETT 1995: 76).

cas. Há pelo menos duas maneiras de alguma coisa ser biologicamente impossível: a violação de leis biológicas, se estas existirem, e o mero afastamento histórico de uma possibilidade. A possibilidade histórica, nomeadamente a possibilidade biológica, diz assim respeito a *oportunidades*, ao facto de a maneira como as coisas historicamente passaram a (ter que) ser não ser a única maneira como as coisas poderiam ter sido. É essa contingência que está na origem do *design*.

Estes são alguns dos constrangimentos fundamentais dos princípios darwinistas acerca de origem do *design*. Em DDI Dennett alarga-se na exploração filosófica do darwinismo, mas o que interessa para o problema da fundação do significado no *design* é discernir em que medida, de acordo com os princípios darwinistas assim explicitados, o *design* é relativo ou real. Ora o que acontece é que a realidade do *design*, a sua independência de qualquer interpretação, diz respeito à *produção algorítmica*, enquanto que a relatividade do *design* diz respeito à *identificação de padrões* no espaço do *design*. Dennett considera que, exactamente como os padrões reais da intencionalidade os padrões no espaço do *design* (por exemplo as especiações) apenas são visíveis subindo de nível (em relação à EF) e utilizando idealizações[393]. É esta posição, que isola quer na EI quer na ED um elemento de interpretação que conduz Dennett aos problemas de circularidade na concepção das relações entre EI e ED. Ela não nega a realidade do desenho material produzido por um processo "algorítmico", simplesmente liga o reconhecimento dos padrões desenhados à utilização de idealizações. O desenvolvimento mais aprofundado desta ideia, que se considera ser a última palavra da TSI acerca de cognição, será retomado no Capítulo 6.

[393] Recorde-se o exemplo das duas caixas negras evocado na primeira parte deste capítulo. Ele refere-se tanto ao estatuto dos padrões da intencionalidade como ao estatuto dos padrões do *design*.

Capítulo 3

Pensamentos conhecendo outros pensamentos. Dennett e os debates da teoria da consciência (dos anos 70 aos anos 90): de Brainstorms *a* Brainchildren *passando por* Consciousness Explained.

> «Podemos com efeito pensar, sentir, querer, recordarmo-nos, agir em todas as acepções do termo, sem ter consciência de tudo isso (...) Então para que serve a consciência, se ela é supérflua para o essencial da existência?»
> Nietzsche, *Gaia Ciência*, Do génio da espécie, 354

> «How on earth could my thoughts and feelings fit in the same world with the nerve cells and molecules that made up my brain?»
> Dennett, *Consciousness Explained*[394]

3.1 *A consciência como problema dos anos 90. A posição heterodoxa de Dennett: os* qualia *aniquilados, o Teatro Cartesiano desmontado, os* zombies *declarados inconcebíveis. Teorias empírico-especulativas da consciência. Um outro prisma: o problema metafísico da consciência fenomenal e da sua irredutibilidade.*

No capítulo anterior assumiu-se a independência do problema do conteúdo relativamente ao problema da consciência. Essa independência não é de modo algum pacífica. Mais especificamente, não é pacífico que a *intencionalidade genuína* seja independente da consciência. No entanto, Dennett rejeita a própria noção de intencionalidade genuína, o que lhe permite defender a referida independência e mesmo propor uma ordem entre a teoria do conteúdo,

[394] DENNETT 1991: xi.

que deve ser tratada em primeiro lugar na teoria da mente, e a teoria da consciência, que virá depois. Neste capítulo analisar-se-á a teoria dennettiana da consciência. A sua forma mais elaborada é o Modelo dos Esboços Múltiplos (MEM). Ao longo do presente capítulo procurar-se-á expor a teoria dennettiana da consciência, justificá-la e também verificar a sua coerência relativamente à outra metade da teoria da mente, a teoria da intencionalidade. Como se verá, ao contrário do que acontece com a TSI, o MEM faz apelo a características dos substractos físicos que implementam processos mentais, concretamente a características dos processos neuronais. Assim, para além dos problemas à frente apontados (como a possibilidade de os modelos apresentados não constituirem uma *teoria* da consciência e a possibilidade de os modelos não capturarem a consciência *fenomenal*) não é sequer claro que as duas metades da teoria da mente (a TSI e o MEM) se articulem sem problemas. É mesmo possível defender que elas se contradizem e apresentam visões diferentes e incompatíveis quanto àquilo que é fundamental na realidade[395].

A posição de Dennett quanto à consciência só se torna compreensível num contexto que se procurará em seguida esclarecer. Na introdução da antologia *The Nature of Consciousness,* organizada por N. Block, O. Flanagan e G. Güzeldere[396], chama-se a atenção para a estranheza da história da investigação científica e filosófica da consciência nos últimos cem anos: «Talvez não exista outro fenómeno em relação ao qual o pêndulo da credibilidade intelectual tenha oscilado a tais extremos (...) no espaço de aproximadamente um século, fazendo-o passar por tempos de exaltação e por tempos em que foi encarado como um tabu»[397]. Nem o afastamento do behaviorismo na filosofia e na psicologia nos anos 60, que produziu o cognitivismo psicológico e a sua justificação filosófica, o funcionalismo, reintroduziu de facto a consciência na teoria da mente. Apenas nos anos 90 se dá o definitivo retorno à respeitabilidade do

[395] RUDDER BAKER 1994. Dennett encara essa possibilidade e justifica-se da seguinte maneira: «The intentional stance presupposes (or fosters) the rationality and hence the unity of the agent – the intentional system – while the Multiple Drafts Model opposes this central unity all the way. Which (...) is the right way to conceive of a mind? It all depends on how far you are. The closer you get, the more the disunity, multiplicity and competitiveness stand out as important.» (DENNETT 1991: 458).

[396] BLOCK, FLANAGAN & GÜZELDERE 1997.

[397] GÜZELDERE 1997: xi.

problema da consciência. A prova desta nova respeitabilidade é a atenção que lhe é dedicada nas neurociências, na neuropsicologia, na psicologia cognitiva e na filosofia da mente. Nesta última, a consciência parece mesmo ter-se tornado o tópico da década[398]. Ora, entre os filósofos, a tentativa de abordagem da consciência a partir de um ponto de vista naturalista conduz frequentemente à proclamação de uma última fronteira: a consciência seria o problema limite para uma metafísica materialista devido ao carácter especial dos *qualia* e ao facto de – pelo menos para algumas entidades – ser como alguma coisa ser. Dennett aparece neste contexto como uma excepção: ele nega a existência de um problema filosófico especial da consciência, insurge-se contra as ideias de abismo explicativo *(explanatory gap*[399]*)* e de problema difícil (*hard problem*[400]*),* e descontrói aplicadamente a noção que usualmente sustenta o abismo entre a consciência e o resto, bem como a dificuldade do problema, precisamente a noção de *qualia*[401].

[398] Cf GÜZELDERE 1997: xi-xii. Alguns exemplos de obras marcantes, sobre a consciência, escritas por filósofos nos anos 90 são: DENNETT 1991, *Consciousness Explained,* McGINN 1991, *The Problem of Consciousness,* SEAGER 1991, *The Metaphysics of Consciousness,* FLANAGAN 1992, *Consciousness Reconsidered,* SEARLE 1992, *The Rediscovery of Mind,* DRETSKE 1993, *Conscious Experience,* CHURCHLAND 1995, *The Engine of Reason,* DRETSKE 1995, *Naturalizing the Mind,* CHALMERS 1996, *The Conscious Mind* e SEARLE 1997, *The Mistery of Consciousness.* A própria antologia *The Nature of Consciousness* (BLOCK, FLANAGAN & GÜZELDERE 1997), na qual aparecem vários textos provenientes ou derivados das obras referidas, é uma referência marcante, bem como a obra *Explaining Consciousness – The Hard Problem* (SHEAR 1997), um conjunto de textos escritos em torno da obra de David Chalmers e provocados pela ideia defendida por este segundo a qual a consciência seria o "problema difícil" da teoria da mente (a noção de problema difícil é adiante explicada). Uma terceira obra conjunta de referência é o já referido livro *The Mistery of Consciousness,* no qual se encontra reunido um conjunto de recensões que J. Searle fez de obras acerca da consciência (obras de F.Crick, G. Edelman, R. Penrose, D. Dennett, D. Chalmers e I. Rosenfield) acompanhadas de trocas de argumentos entre Searle e Chalmers, entre Searle e Dennett e de uma conclusão do próprio Searle.

[399] A expressão é de Joseph Levine (cf. LEVINE 1983, *Materialism and Qualia: The explanatory gap.*) Após a nomeação do abismo resta saber se a existência deste representa um problema epistemológico ou um problema ontológico. O próprio J. Levine pensa que a consciência não coloca um problema ontológico ao materialismo (cf. LEVINE 1997).

[400] A expressão é de David Chalmers (cf. CHALMERS 1996) e pretende contrastar o problema da consciência (ou experiência fenomenal, ou subjectividade sentida) com os restantes problemas da cognição, relativos a habilidades e funções cognitivas, tais como a discriminação e a categorização, o processamento e integração de informação, o controlo do comportamento, a capacidade de reportar verbalmente estados internos, a distinção entre a vigília e o sono, etc. De acordo com Chalmers, «a distinção entre os problemas fáceis e o problema difícil é (uma distinção) conceptual, e não uma pretensão de que os dois não têm nada em comum» (CHALMERS 1997: 411).

[401] O texto em que Dennett encara directamente a noção é DENNETT 1988e, *Quining Qualia.* Escolheu-se a análise do conceito de *qualia* neste artigo como ponto de partida para a exposição da teoria da consciência. A escolha justifica-se pelo facto de Dennett chegar a considerar a sua filosofia

Antes de se analisar a posição de Dennett quanto aos *qualia*, que será tomada ao longo de todo este capítulo como ponto de referência da teoria da consciência e emblema do verificacionismo desta, far-se-á um excurso descritivo por alguns marcos importantes da investigação empírico-especulativa recente, procurando colocar em paralelo conceitos de consciência aí sugeridos. Essa investigação por um lado e as abordagens filosóficas que consideram a consciência como o problema limite para o materialismo por outro constituem o contexto no qual importa considerar a forma mais desenvolvida da teoria dennettiana da consciência, i.e. o Modelo dos Esboços Múltiplos.

3.1.1 *Alguns marcos da investigação empírico-especulativa.*

O artigo *Towards a Neurobiological Theory of Consciousness* de F. Crick e C. Koch[402] é um marco da investigação neurobiológica da consciência sobretudo devido ao facto de o problema da consciência ser aí directamente abordado como um problema científico legítimo. No artigo é defendida uma hipótese acerca do correlato neuronal da consciência, que consistiria em oscilações de 40Hz com um papel importante na ligação (*binding*) num todo dos vários tipos de informação processados no cérebro. F. Crick[403] explica da seguinte maneira (a partir do estudo da visão) o alcance da hipótese defendida. Antes de mais, o correlato neuronal da consciência aparece como solução para um determinado problema, o *binding problem*, que consiste no seguinte: nós temos experiência visual unificada de objectos, no entanto, os traços dos objectos vistos –como cor, forma e movimento – são tratados por regiões diferentes e separadas entre si do córtex visual. O problema que se coloca consiste em saber como é possível que do tratamento desunificado e distribuído da informação se passe a uma experiência consciente unificada de objectos. A estratégia de Crick consiste em perguntar como é que os neurónios envolvidos numa particular percepção são temporariamente activados como unidade. A hipótese dos 40

como um "absolutismo da terceira pessoa" (cf. CHALMERS 1997: 384, para uma expressão de concordância com tal veredicto), constituindo portanto aparentemente o problema da subjectividade sentida o obstáculo maior.

[402] CRICK & KOCH 1990.

[403] CRICK 1994.

Hz sugere como solução o disparo sincronizado de neurónios numa determinada frequência. Este disparo sincronizado seria assim o procurado correlato neuronal da consciência. Como o tálamo[404] parece ter um papel central na consciência – esta parece estar dependente de circuitos ligando o tálamo ao córtex – a definição de consciência sugerida é a seguinte: o correlato neuronal da consciência é o *disparo sincronizado, numa determinada frequência, de neurónios dos circuitos que ligam o tálamo ao córtex*[405].

O trabalho do neuropsicólogo Larry Weiskranz sobre visão cega é outra referência da investigação científica da consciência. A visão cega, uma patologia neuropsicológica na qual os sujeitos não têm (afirmam não ter) experiência consciente de estímulos visuais determinados comportando-se no entanto, nomeadamente em testes de discriminação, como se tivessem essa experiência, parece ser um caso de dissociação entre cognição e consciência, permitindo isolar a consciência[406]. Seja o que for a consciência, ela é aquilo que falta na visão cega. Um sujeito com visão cega tem, aparentemente, percepção visual sem ter consciência de que a tem, sendo por isso um *zombie* parcial. A visão cega parece ser um contra-exemplo para uma teoria funcionalista da consciência, pois mostra que a função pode estar presente na ausência de algo específico, de algo a mais, que é a consciência, que seria então *um sentir ou um saber dos produtos da função cognitiva por um sujeito.*

A obra *A Cognitive Theory of Consciousness*[407] de Bernard Baars é por sua vez um marco do estudo da consciência na psicologia cog-

[404] O tálamo é uma estrutura sub-cortical do diencéfalo cujo papel é essencial na (re) transmissão de informação para os hemisférios cerebrais.

[405] Para um resumo da hipótese, cf. SEARLE 1997, Capítulo 2. Searle acusa F. Crick de cometer vários erros filosóficos na exposição feita em CRICK 1994. Nomeadamente, Searle acusa Crick de não compreender o problema dos *qualia* (que confunde com os problemas da privacidade e da comunicação das experiências) e de confundir dois sentidos de "redução": o sentido de explicação causal (que é o que Crick faz) e o sentido de eliminação. Para além do mais, a expressão "correlato neuronal" utilizada por Crick é perigosamente dualista, embora ser dualista não seja certamente a intenção de Crick.

[406] Cf. WEISKRANZ 1986. A visão cega é devida a lesões no córtex visual. Um paciente com visão cega afirma não ter experiência visual sendo capaz de superar testes que pressupõem a posse da informação visual cujo conhecimento é negado. Não há nada de bizarro na posse dessa informação (i.e. na sua presença no cérebro do paciente) uma vez que a lesão é apenas cortical e a restante maquinaria visual está intacta. Para investigações mais recentes, cf. também WEISKRANZ 1996. Os pacientes com visão cega são sujeitos a testes de discriminação forçada (*forced-choice discrimination*), persuadidos a entrar numa espécie de jogo, em que se lhes pede por exemplo que tentem adivinhar se uma barra é horizontal ou vertical mesmo que (supostamente) não vejam nada (WEISKRANZ 1988: 187). Cf. também HUMPHREY 1995: 95-100, FLANAGAN 1992: Capítulos 6 e 7 e DENNETT 1991: 322-333.

[407] BAARS 1988. Cf. também BAARS 1997.

nitiva. Baars apresenta um modelo funcionalista que lhe permite definir a consciência como o *espaço de trabalho global num sistema de processadores de informação inteligentes e distribuídos*. Quando os processadores acedem ao espaço de trabalho eles difundem ou emitem (*broadcast*) os conteúdos processados para o sistema na sua totalidade. Essa é a característica dos conteúdos conscientes num sistema: aquilo de que o sistema é consciente é aquilo que no sistema é globalmente acessível para o próprio sistema. Os conteúdos da consciência são os *conteúdos do espaço de trabalho global*. O efeito de difusão permite que os conteúdos deixem de estar isolados para se tornarem mutuamente acessíveis sendo assim susceptíveis de uso generalizado e não apenas especializado. Sem a acessibilidade mútua dos conteúdos no espaço de trabalho global estes "não seriam como nada" para o sujeito. O espaço é global não apenas no sentido funcional – i.e. é um "espaço" onde tudo pode entrar em contacto com tudo – como também no sentido anatómico de ser distribuído pelo córtex, envolvendo mesmo outras regiões do cérebro[408].

O biólogo G. Edelman é o autor de um outro livro marcante na teoria da conciência, *The Remembered Present: A Biological Theory of Consciousness*[409]. Nessa obra, G. Edelman propôs-se estender a uma teoria da consciência a sua explicação do desenvolvimento de categorias perceptuais por meio do darwinismo neuronal[410]. Se F.Crick parte da visão para chegar à consciência, Edelman parte da categorização. As ideias fundamentais de Edelman são as ideias de Mapas, Selecção de Grupos Neuronais e Reentrada[411]. Os mapas são camadas de neurónios no cérebro cujos pontos estão sistematicamente relacionados com pontos numa camada de células receptoras, por exemplo da superfície da pele ou da retina. Com a ideia de *Selecção* de Grupos Neuronais Edelman sugere que o desenvolvimento cerebral não deve ser pensado como função do impacto do ambiente no organismo. O cérebro está geneticamente equipado desde o nascimento com grupos neuronais abundantes e desenvolve-se através de um mecanismo darwiniano de selecção natural cujas unidades são grupos neuronais e não células individuais. O cérebro não é, assim, um mecanismo de instrução mas de selecção. A Reentrada é o processo pelo qual sinais paralelos circulam para a frente e para trás entre mapas (o processo não é identificável com o *feedback* na medida em que prevê a existência de muitos canais paralelos a funcionar simultaneamente).

[408] O que significa por exemplo a utilização do mesmo *hardware* utilizado pela memória. Baars sugere o ERTAS como "correlato neuronal" (para usar a expressão de Crick) do seu modelo funcional. ERTAS é uma sigla para *extended reticular-thalamic activating system*. A sugestão de Baars apoia-se no facto de a estimulação numa área do tronco cerebral se estender ao córtex (o animal fica alerta, orienta-se, desperta se estiver a dormir). Essa área do tronco cerebral (a formação reticular, sistema de activação reticular) veio a ser considerada apenas parte de um sistema ascendente mais amplo que integra o tálamo e o córtex. A sigla sublinha que o sistema, embora se estenda, está centrado no tálamo.

[409] EDELMAN 1989.

[410] Cf. EDELMAN 1987, *Neural Darwinism: A Theory of Neuronal Group Selection*.

[411] A exposição que se segue apoia-se em SEARLE 1997, Cap. 3.

Edelman utiliza estas três ideias para abordar o desenvolvimento da consciência a partir daquilo a que chama *consciência primária*. O cérebro tem um problema a resolver: perante um mundo não etiquetado é preciso desenvolver categorias e abstracções. Existem muitos inputs de estímulos para cada "categoria" prospectiva. No entanto, após muitos estímulos, padrões de grupos neuronais serão seleccionados em mapas. A partir daí, sinais similares activarão os grupos seleccionados. Os mapas não funcionam isoladamente mas em grupos, ligados por canais de reentrada. Cada mapa pode utilizar para as suas operações discriminações feitas por outros mapas. Um mapa pode extrair bordos, um outro mapa movimentos, um outro mapa pode, devido ao processo de reentrada, extrair a forma a partir dos bordos e do movimento. O resultado são representações unificadas de objectos, mesmo que se encontrem distribuídas por muitas zonas do cérebro. Quando existem mapas em todo o cérebro sinalizando entre si Edelman fala de *mapeamento global*. É o mapeamento global que permite ao sistema ter categorias perceptuais, generalizar e coordenar percepção e acção.

Nenhum dos processos descritos até ao momento é consciente. Para se chegar à consciência é ainda necessário distinguir entre a *consciência primária* (que Edelman identifica com ter sensações, categorizar perceptualmente) e a *consciência de nível elevado* (que inclui auto-consciência e linguagem). Para Edelman o maior problema é obter uma teoria da consciência primária, já que a consciência de nível elevado será depois "extraída" a partir desta. Edelman propõe então que a existência da consciência primária requer a existência de:

(1) Memória (entendida como processo activo de categorização e não apenas armazenamento de informação).

(2) Um sistema para a aprendizagem (i.e. um modo de valorizar certos estímulos em detrimento de outros).

(3) A habilidade de discriminar o *self* em relação ao *não-self*, que não é ainda a auto-consciência mas o facto de o sistema nervoso ser capaz de discriminar o organismo do qual é parte em relação ao resto do mundo. Esta distinção "existe" aliás na anatomia cerebral, na diferença entre áreas de registo de estados internos e áreas dedicadas ao tratamento de sinais externos.

(4) Um sistema para categorizar eventos sucessivos no tempo e para formar conceitos (Edelman sugere um substracto neurobiológico comum a estas duas capacidades).

(5) Um tipo especial de memória, envolvendo interacções entre o sistema 4 e os sistemas 1, 2 e 3 de modo a ser possível a memória de valores associados a categorias passadas.

(6) Um conjunto de conexões de reentrada entre o sistema especial da memória e os sistemas anatómicos dedicados às categorizações perceptivas.

A partir destas ideias, Edelman propõe que se considere a consciência como sendo o *resultado de uma memória recursivamente comparativa na qual categorizações self / não self prévias são continuamente relacionadas com as categorizações perceptuais presentes e a sua sucessão de curto prazo, antes de tais categorizações se tornarem parte da memória.*

Um outro trabalho teórico acerca da natureza da consciência que se impôs na década de noventa foi o trabalho do neurocientista António Damásio[412], em grande parte dedicado a explorar a ligação entre consciência, emoções e corpo próprio, simultaneamente nos níveis neurobiológico, das arquitecturas cognitivas e fenomenológico. O problema de Damásio não é exactamente a consciência fenomenal dos filósofos (o ser como alguma coisa ser) mas a subjectividade (do *self*, do eu), o enraízamento corpóreo dessa subjectividade e o acompanhamento da actividade mental por um sentimento de si corporeamente apoiado. Damásio considera que o problema da consciência tem uma dupla face: ele é por um lado o problema do filme dentro da cabeça (o decorrer temporal da consciência e seus conteúdos) e por outro o problema da pertença desse filme (o sentido do *self*, a reportação da fenomenologia a uma entidade que se sente ser). É este último problema o foco da atenção de Damásio (o primeiro problema poderá, segundo Damásio, ser abordado por meio de modelos como o acima

[412] DAMÁSIO 1994 e DAMÁSIO 1999.

referido modelo do espaço de trabalho global de Baars ou mesmo por meio do MEM de Dennett). Damásio considera que a subjectividade tem raízes corpóreas estratificadas e diversificadas, i.e. que o cérebro tem o corpo em mente de diversas maneiras e a diferentes níveis. Em *O Erro de Descartes*[413] o problema da corporeidade da subjectividade conduz Damásio a uma teoria da emocionalização da mente. Se um dos sentidos do erro cometido por Descartes é o dualismo espírito-corpo, o outro sentido é a intelectualização do mental, a separação entre uma razão que seria pura e as emoções. Pelo contrário, Damásio sublinha a importância das emoções em fenómenos de racionalidade prática tais como decisões e planeamento das acções próprias, bem como as características de patologias neuropsicológicas (como as provocadas por lesões nos lobos frontais[414]) ligadas com a ausência de um funcional normal das emoções. Uma vez que na ausência de um normal funcionamento das emoções o mecanismo da racionalidade prática é perturbado, Damásio defende que a natureza construiu o "aparelho da racionalidade" (que nos permite nomeadamente decidir e planear) não por cima dos mecanismos biológicos de regulação do corpo mas a partir deles e com eles. Este é um dos aspectos da corporeidade da subjectividade[415]. Por corporeidade da subjectividade deve ainda entender-se uma teoria neurobiológica mais ou menos localizacionista do *self*[416] e do eu. Damásio distingue as fundações neurobiológicas do *self* das "localizações" do eu autobiográfico da identidade pessoal, que considera ser uma realidade de nível funcional e cortical mais elevado. É sobre o *self* que vem estabelecer-se o eu linguístico.

[413] Cf. DAMÁSIO 1994.

[414] É o caso de Phineas Gage, exemplar em DAMÁSIO 1994 (cf DAMÁSIO 1994: Capítulo 1).

[415] Este tema do "*body-minded brain*", central em DAMÁSIO 1999 e que é já explorado no Capítulo 10 de DAMÁSIO 1994. A ideia central é que o corpo contribui com um conteúdo essencial para o funcionamento da mente normal.

[416] I.e. existem determinadas regiões cerebrais ligadas à constituição do *self* como núcleos do tronco cerebral, o hipótalámo e estruturas basais, o córtex da ínsula, o córtex somatossensorial. Elas são a base neuroanatómica da distinção entre *self* e não-*self* (cf. DAMASIO 1999: 155). Os núcleos do tronco cerebral regulam estados do corpo e mapeiam sinais corporais. O hipotalámo intervem na representação "actual" do corpo mantendo um registo do estado de várias dimensões do meio interno (como a circulação de nutrientes, a concentração iónica, o PH, etc). O córtex da insula e o córtex somatossensorial mantêm representações do corpo. Damásio liga estas regiões ao que chama proto-*self*. Do facto de ser possível apontar estas estruturas não se segue que a distinção entre si e não-si (envolvendo *self* e eu) seja consciente e pessoal em todos os níveis envolvidos. Para uma abordagem filosófica de uma noção idêntica de proto-*self* ou proto-eu, cf. GIL 1998, Parte 1, 3.

Embora o objectivo geral de Damásio em *The Feeling of What Happens* seja analisar o acompanhamento do pensamento de outros conteúdos por um pensamento do *self* e/ou do eu, i.e. por um "sentimento" de si, ele pretende precisamente distinguir as fundações neurobiológicas do *self* das "localizações" do eu autobiográfico da identidade pessoal. O sentimento de si tem em cada um de nós uma versão nuclear e uma versão sofisticada, ligada à consciência alargada (*extended consciousness*) e ao eu. O eu está associado a uma reunião das memórias daquilo que acontece. Só ela permite que um eu se mantenha e se reconheça como o mesmo ao longo do tempo e que exista uma consciência de si propriamente *pessoal*. Uma das funções do cérebro é, assim, manter uma narrativa constante e contínua na qual o eu é a personagem principal. Embora o eu constitua, relativamente ao *self*, uma realidade de nível funcional e cortical mais elevado, o eu e a consciência alargada vêm estabelecer-se sobre o *self*. Parte do interesse de Damásio é precisamente explorar a transição entre representações de si ainda não conscientes e consciência propriamente dita. A transição não é no entanto uma definitiva "passagem" para uma consciência pura, que seria um saber--que-se-sabe-neutro, como se não fosse de subjectividade incorporada nenhuma. Pelo contrário, a consciência como aparição de si a si, que pode ser pessoal, não se destaca nunca de um fundo que acompanha a vida consciente e que é constituído pelas representações de si a vários níveis, mantidas pelo cérebro. Damásio sugere assim que o cérebro é uma audiência cativa de representações de si, representações que o próprio cérebro constantemente mantém. Esse facto é essencial para a aparição da aparição, i.e. para o surgimento da consciência como um *sentido de si próprio no acto de conhecer*. Esta é a resposta de Damásio à segunda face do problema da consciência (o problema da pertença do "filme"). A importância do corpo (mais especificamente da representação do corpo no e pelo cérebro desse corpo) na constituição da consciência conduzem Damásio à sua tese central, segundo a qual a consciência se inicia com um sentimento, um sentimento de si e de saber. Este sentimento é, segundo Damásio, gerado pela maneira como o cérebro sente a carne do corpo próprio depois de o cérebro ter agido sobre esta. É com a perda da inocência acerca da existência que está aqui em causa que se paga a possibilidade de uma existência melhor ou mais rica: «O sentimento daquilo que acontece é a resposta a uma questão que nunca pusemos, e é também a moeda num trato faus-

tiano que nunca poderíamos ter negociado. A Natureza fê-lo por nós»[417]. O aspecto do problema da consciência que Damásio põe em relevo seria esquecido em abordagens como a de Dennett: Damásio chega a considerar que o MEM é o modelo de uma consciência des-apropriada (*selfless consciousness*)[418].

Vindo do exterior de áreas disciplinares biológicas e psicológicas, o físico e matemático R. Penrose[419] defendeu que o nível e o tipo de teoria requeridos para a explicação da consciência são totalmente diferentes dos até agora mencionados. Quanto ao tipo de teoria, a explicação da consciência envolverá de acordo com Penrose a mecânica quântica e a lógica matemática (nomeadamente o teorema de Gödel), e não neuroanatomia, neurofisiologia ou psicologia cognitiva. Quanto ao nível ou escala da teoria, o nível dos neurónios não é o nível apropriado para a explicação da consciência. Segundo Penrose os neurónios apenas amplificam o nível mais baixo no qual reside a explicação física última da mente consciente. A ideia chave de Penrose é a seguinte. Os seres humanos conscientes são capazes de habilidades cognitivas que os computadores não conseguem simular. Segue-se que o entendimento humano não pode ser um actividade algorítmica. O teorema de Gödel mostra que existem verdades em sistemas matemáticos que não podem ser demonstradas nesses sistemas. Uma versão específica do teorema de Gödel (a irresolubilidade do problema da paragem) pode ser utilizada para provar que o comportamento humano consciente nem sequer pode ser simulado em computador. Ora, os neurónios são computáveis (i.e. as suas características podem ser simuladas em computador). Como a consciência tem características não computáveis, os neurónios não podem explicar a consciência. Para explicar a consciência é necessária qualquer coisa não computável, nomeadamente efeitos que se encontram ao nível sub-neuronal, no citoesqueleto das células, nos microtúbulos dos neurónios. Compreender estes efeitos requererá uma revolução na física; Penrose alude a uma teoria da gravidade quântica. A sua ideia acerca da consciência é portanto que esta poderia ser uma *manifestação dos estados quânticos internos (do citoesqueleto) dos neurónios e do envolvimento destes estados na inter-*

[417] DAMÁSIO 1999: 316.
[418] DAMÁSIO 1992.
[419] Cuja incursão pelos problemas da consciência se iniciara em PENROSE 1989.

acção entre os níveis quântico e clássico de actividade. A hipótese de Penrose acerca da consciência convoca assim a lógica matemática e a física quântica[420].

Uma vez que foram aqui enumerados alguns marcos da investigação empírico-especulativa da consciência, convém ainda sublinhar que foi especialmente importante para uma certa acessibilidade empírica das questões da consciência o desenvolvimento recente de técnicas de imagiologia, que permitiram nomeadamente explorar a ligação entre aspectos fenomenológicos e aspectos neuronais da experiência em sujeitos vivos e conscientes, humanos e não humanos, em tempo real [421].

Como é suficientemente claro a partir deste panorama breve, as sugestões especulativas quanto ao próprio teor de uma teoria da consciência apresentadas a partir do trabalho empírico são muito diferentes e mesmo incompatíveis. O próprio nível dos princípios explicativos sugeridos (cognitivo-funcional, neurobiológico, microfísico) varia. Apenas a constatação da enorme difusão recente do interesse pelo problema é segura neste contexto[422]. No entanto, acompanhando essa constatação, filósofos como D. Chalmers ou J. Searle podem sempre afirmar que por mais que o tipo de teorias a que se aludiu encontrem os processos correlacionados com a consciência elas (i) *não explicam a razão por que tais processos físicos devem originar subjectividade e* (ii) *não elucidam o carácter dessa mesma subjectividade.* As teorias referidas apenas apontam – como um facto bruto e utilizando critérios pragmáticos como o controlo do comportamento e a capacidade de reportar verbalmente – a *correlação* entre determinados processos físicos e a consciência. Como se verá neste capítulo, a teoria da consciência apresentada por Dennett, embora se apresente como uma teoria filosófica da consciência, faz algo de muito semelhante.

[420] A relação entre a física quântica e a consciência estabelece-se devido à importância que a consciência pode ter nas próprias fundações da teoria quântica: o colapso da função de onda seria provocado, de acordo com as interpretações mais comuns da teoria quântica, pela medição, que supõe a consciência. A consciência determinaria indeterminações quânticas.

[421] Cf. BUCKNER & PETERSEN 1998.

[422] De um ponto de vista mais sociológico e procurando dar conta da proliferação recente dos debates em torno da teoria da consciência, essa grande difusão manifestou-se por exemplo nas conferências de Tucson, no aparecimento das revistas *Consciousness and Cognition* e *Journal of Consciousness Studies* e da Association for the Scientific Study of Consciousness (ASSC), etc. Para dois panoramas do estudo recente do problema da consciência elaborados por filósofos, cf. FLANAGAN 1998 e BLOCK 1994.

3.1.2 *Quining Qualia.*

Como se afirmou, são muitos os filósofos que, no contexto do interesse recente pelo problema da consciência, o consideram o problema limite para o materialismo. Dennett opõe-se a esse diagnóstico e ao uso que nele é feito da noção de *qualia.* Uma parte essencial (a que se chamará desconstrutiva) da teoria dennettiana da consciência é precisamente a eliminação dos *qualia.* Os *qualia* que Dennett procura eliminar são, à partida, definidos de forma muito pouco consensual, como propriedades intrínsecas[423], inefáveis[424], privadas[425], imediatamente (e incorrigivelmente) conhecidas[426], da maneira como as coisas nos (a)parecem[427]. A aniquilação destes *qualia*[428] é uma peça básica da teoria da consciência[429]. Os críticos são unânimes na classificação da posição como eliminativista. O "materialismo eliminativo" de Dennett acerca dos *qualia* opõe-se, segundo o próprio, a um materialismo reducionista, que procuraria identificar o item problemático nos termos de uma teoria materialista fundacional. A diferença entre materialismo eliminativo e materialismo reducionista é considerada por Dennett co-

[423] Entende-se por *carácter intrínseco* o facto de alguma coisa (aqui, os *qualia*) existir independentemente de qualquer interpretação que dela seja feita, sendo portanto não-relacional, não disposicional. Quando Dennett nega o carácter intrínseco dos *qualia* ele está a afirmar que nada existe na experiência (com essas características) que seja independente da forma como o sujeito está subsequentemente disposto a comportar-se ou a acreditar.

[424] Entende-se por *inefável* aquilo que é inexprimível, incomunicável, que apenas é possuído por quem experimenta. Se Dennett não defendesse, como defende, uma teoria tão intelectualista dos *qualia* poder-se-ia falar aqui de conteúdo não conceptual, de apercebimento fenomenal, não linguístico, e por isso mesmo (por definição e não problematicamente) inexprimível e incomunicável.

[425] Entende-se por *privado* aquilo que é insusceptível de ser acedido e/ou confirmado por outrem, insusceptível de testes exteriores, que são assim considerados de importância insignificante para o caso. Na teoria da consciência de Dennett os testes exteriores são pelo contrário extremamente importantes.

[426] Entende-se por *imediatidade* a não inferencialidade, a intimidade, o contacto (*acquaintance*), a característica daquilo ao qual o experienciador tem acesso directo. Por incorrigibilidade entende-se a infalibilidade de um saber. Dennett liga quase sempre a questão da incorrigibilidade à memória da experiência e portanto a um acesso presente a *qualia* de experiências passadas, evidentemente pressupondo que não existe um acesso à experiência que não seja um acesso ao conjunto formado pela experiência-mais-memória. Os opositores de Dennett tenderão a acentuar o presente da experiência e a incorrigibilidade do acesso presente e não a "incorrigibilidade" das memórias de experiência.

[427] Para um comentário crítico a cada ponto da definição dennettiana de *qualia* resultando numa defesa da sobrevivência do Teatro Cartesiano e dos *qualia*, cf. LORMAND 1994.

[428] Em geral traduzir-se-á assim a expressão "*quining qualia*", já que a tradução preserva quer o sentido da expressão quer a homofonia. Por vezes falar-se-á também de "desqualificação dos *qualia*", como o próprio Dennett faz em DENNETT 1991: 369.

[429] Ou melhor, do aspecto desconstrutivo, prévio, da teoria da consciência. O aspecto construtivo da teoria está ligado a conceitos com origem no pensamento acerca de máquinas computacionais.

mo apenas táctica, mas a questão não é tão indiferente assim. Na alternativa joga-se a própria definição do problema da consciência e a relação deste com a metafísica materialista e o problema da redução teórica. Na definição de Dennett, os *qualia* são propriedades de segunda ordem, propriedades do conhecimento que os sujeitos têm dos conteúdos da sua própria experiência consciente. A experiência de um sujeito tem o *quale* F se e só se o sujeito julga que a sua experiência tem o *quale* F. O problema dos *qualia* é assim convertido num problema epistemológico acerca de autoridade introspectiva e infalibilidade, e portanto em geral num problema de (auto) acesso. Assim sendo, o eliminativismo acerca de *qualia*, a surgir, surge por razões *epistemológicas*. No entanto, para a maioria dos autores, os *qualia* são estados fenomenais, ligados ao sentimento de si, pelos quais é como alguma coisa para uma entidade ser e é relativamente à natureza desses estados que se coloca a questão *ontológica* da redução. No primeiro caso trata-se do estatuto daquilo que é reportado mediante o acesso a si de um sistema a que se chama vulgarmente introspecção, no segundo caso o que está em causa é o facto de «o introspectado ser sentido»[430]. Repetindo, no primeiro caso o problema é epistemológico (um problema acerca do que fazemos quando enunciamos conhecimento), no segundo caso o problema é ontológico ou metafísico (um problema acerca da locação no mundo tal como ele é de alguma coisa de real).

Esclareça-se antes de mais que os dois sentidos de consciência em causa –auto-acesso e consciência fenomenal[431] – não são mutuamente exclusivos. Um defensor da consciência fenomenal pode perfeitamente admitir que esta está usualmente ligada ao auto-acesso, sem ter necessariamente que considerar que o auto-acesso oferece conteúdos incorrigíveis. De facto, a ideia de Dennett segundo a qual os *qualia* estariam ligados a, ou seriam o nome para o facto de, o auto-acesso oferecer conteúdos incorrigíveis representa uma intelectualização da noção de *qualia* que não é de todo forçosa. Mais: assim como nada obriga o qualófilo[432] a ligar *qualia* e incorrigibilidade (ao contrário do

[430] Cf. PINTO 1999: 118-119. Aí se afirma que, na perspectiva dos filósofos adversários de Dennett, «este fenómeno não envolve qualquer saber de origem introspectiva, sendo antes um facto especialmente bruto (...) pelo qual pode aparecer uma instabilização da correspondência entre o publicamente observado e o acontecimento privado».

[431] Cf. a distinção de Ned Block entre consciência de acesso (*access-consciousness)* e consciência fenomenal (*phenomenal consciousness),* por exemplo em BLOCK 1994a.

[432] Como é frequente nestas discussões (cf. DENNETT 1991) falar-se-á de agora em diante em qualófilos e qualófobos, respectivamente aqueles que consideram a noção de *qualia* imprescindível numa teoria da experiência e aqueles que a rejeitam, como é o caso de Dennett.

que Dennett dá como garantido), nada o obriga a negar que os *qualia* tenham outras propriedades além do facto de serem sentidos em primeira pessoa (i.e. nada obriga a negar que os *qualia* tenham propriedades não acessíveis em primeira pessoa). Evidentemente, como nota N. Block, se se permite que seja um adversário dos *qualia* (como Dennett, por exemplo) a defini-los, os *qualia* possuirão apenas a propriedade de serem sentidos[433]. Se nada obriga o qualófilo ao "isolamento" dos *qualia*, ou melhor, a considerar que existe uma diferença relevante e clara entre *qualia* e outros estados mentais e que os estados fenomenais são estados exclusivamente sentidos (a-conceptuais, dir-se-ia), nada o impede também à partida de considerar que os estados representacionais têm propriedades fenomenais acerca das quais podem existir crenças e memórias enganosas sem que isso torne a noção ilegítima. Parece, em suma, que qualquer qualófilo encontra facilmente um caminho de fuga perante a estratégia intelectualista e eliminativista de Dennett. Procurar-se-á em seguida, analisar cuidadosamente uma particular formulação dessa estratégia.

No texto que dedica directamente à questão dos *qualia*, o artigo *Quining Qualia*, Dennett apresenta quinze casos ou *intuition pumps* destinados a demonstrar que não existem *qualia.* Destes casos, o mais frequentemente citado e aquele que porventura melhor evidencia o que está (para Dennett) em causa na discussão acerca de *qualia*, é o caso de Chase e Sanborn, os provadores de café[434]. Embora se vá aqui tomar este caso como exemplar, convém sublinhar que o caso clássico do espectro invertido (intersubjectivo)[435], bem como a sua variação

[433] Cf. BLOCK 1994.

[434] DENNETT 1997: 625.

[435] O *espectro invertido* é a situação na qual dois sujeitos experienciam cores complementares numa mesma situação, partilhando no entanto o vocabulário para cores e as capacidades de discriminação de cores. Alguém que admita a possibilidade do espectro invertido perde a possibilidade de evocar dados que permitam decidir se as cores são vistas da mesma maneira de um sujeito para o outro. O proponente inicial da situação foi J. Locke, que no *Essay Concerning Human Understanding* considerou a possibilidade de um mesmo objecto produzir na mente de homens diferentes diferentes ideias de cor ao mesmo tempo (cf. LOCKE 1690, II.xxxii.15). Para uma compilação das variações possíveis da situação, cf. o Cap. 7 de CHALMERS 1996, *Absent Qualia, Fading Qualia, Dancing Qualia.* Para uma experiência de pensamento quase sempre mencionada a este propósito (a Terrra Invertida de Block), cf. BLOCK 1997a, *Inverted Earth.* Na Terra Invertida de Block todas as coisas têm a cor complementar à sua cor terrestre e o vocabulário também está invertido. As duas diferenças cancelam-se mutuamente e o discurso sobre cores na Terra Invertida seria idêntico ao discurso sobre cores na Terra. O problema que interessa Block começa com a transposição de alguém da Terra para a Terra Invertida e consiste em saber o que aconteceria ao conteúdo intencional dos pensamentos do transportado uma vez que o conteúdo qualitativo, mediante rearranjos feitos por cientistas previamente à viagem, permanece o mesmo.

contemporânea, o espectro invertido intra-subjectivo por intervenção neurológica manipulatória, fazem parte dos referidos quinze casos. O propósito das análises de Dennett é sempre o mesmo: mostrar que a diferença entre inversão dos *qualia* e inversão das disposições reactivas dos sujeitos não poderia ser subjectivamente apercebida. Ora, se não existe qualquer diferença apercebida entre *qualia* e disposições reactivas, os supostos *qualia* não são senão disposições reactivas.

Chase e Sanborn são dois provadores de café cujo trabalho consiste em assegurar que o sabor do café que provam se mantém o mesmo. Ora, Chase e Sanborn afirmam ambos terem perdido o prazer que anteriormente experimentavam com o sabor do café que provavam e que preferiam a todos os outros cafés. Chase pensa que deixou de apreciar o café porque se tornou um apreciador de café mais sofisticado, embora o café que prova mantenha o sabor inalterado. Sanborn considera que deixou de apreciar o café não porque o saber deste se tenha alterado mas porque os seus sensores gustativos (*tasters*) se alteraram. Em suma, as experiências de ambos mudaram do gosto para o "desgosto", mas a mudança é interpretada diferentemente. Segundo Chase foram os "conceitos" que mudaram, segundo Sanborn foram as "intuições". O ponto de Dennett é que nem Chase nem Sanborn podem, ao contrário do que pensam, estabelecer intrasubjectivamente qual dos casos é o seu caso (o caso em que os "conceitos" mudaram ou o caso em que as "intuições" mudaram) nem exprimir a diferença entre os dois casos. As restantes experiências de *Quining Qualia* apresentam uma estrutura semelhante. Coloca-se o dilema: "O que é que se deve afirmar, A ou B?" para em seguida se afirmar que a pergunta é sem sentido. Aquilo que se passa no caso de Chase e Sanborn, passar-se-ia também, por exemplo, no caso de uma inversão intrapessoal do espectro. O sujeito no qual e para o qual a mudança ocorresse não poderia discernir a diferença, não poderia saber se estaria a experienciar uma mudança nos *qualia* ou nas disposições reactivas. Dennett conclui da análise dos casos de *Quining Qualia* que quando o sujeito é incapaz de aperceber uma diferença, nenhuma diferença existe. A palavra de ordem da doutrina dos *qualia* é que "Diferenças que não fazem diferença não são diferenças", i.e. simplesmente não existem. Todos os argumentos de Dennett acerca de *qualia* recaem sobre o princípio segundo o qual "Quando se trata de fenomenologia, o que não é apercebido não existe, já que existir fenomenologicamente é ser apercebido". O princípio funda o operacionalismo em primeira pessoa metodologicamente essencial a uma teoria da consciência[436] e é reportado a considerações sobre a incorrigibilidade[437].

É muito contestável que a fenomenologia possa ser caracterizada como incorrigibilidade[438]. É perfeitamente possível que Dennett esteja certo quanto à necessidade de eliminar ou aniquilar os *qualia* tal como os define sem que a mesma conclusão se aplique à fenomenologia. O problema de Dennett relativamente à fenomenologia é, como se verá, considerar que não existem objectos fenomenológicos reais mas apenas objectos *meramente* intencionais. Esse problema aparece aliás na sequência das ambiguidades relativas ao mundo nocional referidas no Capítulo 2. É pelo facto de a teoria dos *qualia* intencionalizar e intelectualizar a consciência que Dennett se vê obrigado a encarar o problema da consciência exclusivamente como um problema epistemológico. Para Dennett, só é possível fazer sentido da fenomenologia como um espaço-lógico-cujas-propriedades-são-constituídas-pelas-crenças-do-sujeito. No entanto, entre aquilo que se argumenta quanto aos *qualia* e aquilo que

[436] O operacionalismo em filosofia da ciência é, como se sabe, a posição empirista de acordo com a qual os conceitos teóricos se definem através de operações empíricas, sendo portanto as proposições acerca de entidades teóricas de facto proposições acerca da experiência.

[437] Um ligeiro desvio: de acordo com o *Philosophical Lexicon* compilado por Dennett e disponível on-line em www.blackwell/publishers.co.uk/lexicon/ o termo *"rorts"* aplica-se a *"incorrigible reports"*, e dele proviria *"rorty"*, significando *"incorrigible"*... Em DENNETT 2000, Dennett refere-se a R. Rorty (cf. RORTY 1970, RORTY 1972a e RORTY 1972b), i.e. às incursões de Rorty pela filosofia da mente nos anos 70, como sendo o antecessor não suficientemente reconhecido em exposições anteriores da sua própria posição quanto à incorrigibilidade. Rorty seria o responsável pela exploração da ideia segundo a qual a incorrigibilidade é a marca do mental. Como dizia Rorty nos anos 70 (RORTY 1970: 414) «What makes an entity mental is not whether or not it is something that explains behavior, and what makes a property mental is not whether or not it is a property of a physical entity. The only thing that can make either an entity or a property mental is that certain reports of its existence have the special status that is accorded to e.g. reports of thoughts and sensations – the status of incorrigibility». Como Rorty nota em seguida, essa incorrigibilidade não é infalibilidade: não é que os relatos não possam estar errados, o que se passa é que «certain knowledge claims about them cannot be overriden» (RORTY 1970: 413). Rorty passa da caracterização da incorrigibilidade para um tipo de materialismo eliminativo (aquele que afirma que nesse caso o conhecimento em terceira pessoa pode sobrepôr-se ou ser melhor do que o conhecimento em primeira pessoa dos estados mentais próprios) que Dennett não aceita, e a que chama (cf. DENNETT 2000) *"the Churchlandish alternative"*. De acordo com Rorty, «if it came to pass that people could explain behavior at least as well by reference to brain states as by reference to beliefs, desires, thought and sensations, then references to the latter might simply disappear from the language» (RORTY 1970: 421). Rorty não deixaria assim nenhum ponto de ancoragem para o mental, o que é indefensável para Dennett.

[438] Cf. FOX 1994 para uma exploração exaustiva desta crítica à noção dennettiana de *qualia*. Fox defende que o mundo fenomenal é o "fim da linha" e não requer observador ou conhecimento alguns, e que aquilo que Dennett está a atacar quando ataca os *qualia* é algo de criado pelo próprio Dennett.

se argumenta quanto à fenomenologia vai uma larga distância. Transpô-la implica atribuir ao argumento de Dennett quanto aos *qualia* um alcance que na opinião da maioria dos críticos ele não possui. Se os argumentos de Dennett acerca de *qualia* provam alguma coisa eles provam que o auto-acesso não oferece conteúdos incorrigíveis. Ora, isso não é sequer polémico. Nenhum defensor da consciência fenomenal fica automaticamente comprometido com a defesa da incorrigibilidade do conhecimento dos conteúdos da experiência nem aliás, ao contrário do que Dennett pressupõe, com uma distinção absoluta entre consciência e não-consciência ou com a ligação da consciência a um lugar único ou um centro no cérebro[439].

Como se verá neste capítulo, é da indiferenciação estabelecida por princípio entre fenomenologia e *qualia* que nascem as várias ambiguidades da posição de Dennett quanto à consciência. É essa indiferenciação que faz com que o destino dos *qualia* e o destino do Teatro Cartesiano estejam ligados. De facto, a primeira parte ou parte desconstruccionista da teoria da consciência de Dennett é um caso contra os *qualia* e contra o Teatro Cartesiano. O Teatro Cartesiano é a imagem metafórica do mental como um lugar onde tudo se põe conjuntamente para um sujeito. De acordo com Dennett, ou os *qualia* e o Teatro Cartesiano se sustentam em conjunto ou caem em conjunto. Se os *qualia* e o Teatro Cartesiano se sustentam, as três ideias a seguir listadas são, de acordo com Dennett, legítimas:

(1) Existe uma apresentação ou aparição pura, a-conceptual, que estabelece uma diferença de natureza entre (a) parecer e julgar.

(2) Existe um momento absoluto e um lugar único da consciência (ou nos termos de Dennett, existe uma "linha da meta", sendo que a ordem de chegada a essa linha marca a entrada na consciência de um determinado conteúdo, i.e. a sua apresentação no palco do Teatro Cartesiano). A unicidade e singularidade do fluxo de apresentação ou aparição ficam assim estabelecidas, bem como uma diferença natural entre eventos pré-experienciais e pós-experienciais.

(3) Existe uma distinção nítida entre consciência e não-consciência, entre estar ou não estar no palco do Teatro Cartesiano.

Se as três ideia não se sustentam, não há lugar para *qualia* ou Teatro Cartesiano na teoria da consciência. Resta saber se da oposição a estas três ideias se segue realmente a eliminação da experiência fenomenal, o apagamento da consciência como sentir de si

[439] Cf. BLOCK 1997c para uma crítica – apropriadamente intitulada *Begging the Question Against Phenomenal Consciousness* – ao modelo de consciência defendido por Dennett.

e do pensamento, ou mesmo alguma prova da ilegitimidade da ideia de "teatro" como representação dirigida para dentro, para-si, num dado sistema cognitivo. Note-se que a existência de um sentido interno produzindo tal representação não está necessariamente ligada a um lugar único no sistema cognitivo. Defender-se-á – a partir da "metade construtiva" da teoria da consciência do próprio Dennett – que a resposta a todas essas questões é negativa e, especificamente, que não há nada a objectar a um "Teatro Cartesiano impreciso"[440]. De qualquer modo, e para além dos problemas epistemológicos focalizados por Dennett na noção de *qualia*, para os (muitos) filósofos que consideram a teoria da consciência o núcleo da teoria da mente e a experiência fenomenal o problema limite, o problema mais importante e apaixonante que se coloca a uma concepção materialista do mental e por isso mesmo um problema muito especial, a abordagem que Dennett leva a cabo do problema da consciência é fundamentalmente mal dirigida, não apenas por razões metodológicas mas por razões profundas, relacionadas com o que se entende por "metafísica" e por "teoria".

3.1.3 *Um outro prisma: Metafísica e fundamentalidade da consciência.*

A abordagem dennettiana da consciência é muito pouco ortodoxa no contexto da filosofia da mente. É importante compreender porquê, comparando-a com outras abordagens filosóficas, igualmente materialistas, do problema da consciência. Para muitos críticos de Dennett, por exemplo F. Jackson[441], a abordagem que Dennett faz do problema da consciência é extraviada antes de mais porque é uma forma de evitar dar uma resposta à questão metafísica obrigatória para o materialista, aquela a que Jackson chama a *truth-maker question*. Esta questão é para F. Jackson a seguinte: "o que é que faz com que seja verdadeiro que, num mundo tal como este mundo é, alguma entidade seja consciente?" (ou "o que é que na natureza material de uma entidade faz com que seja necessário que esta seja consciente?"). Uma tal questão deve, segundo Jackson, ser respondida com o lugar – no quadro traçado em termos de ingredientes fundamentais que uma metafísica propõe acerca daqui-

[440] LORMAND 1994: 148.

[441] JACKSON 1993.

lo que existe – de alguma específica característica[442]. É sobretudo muito importante, segundo Jackson, não confundir a *truth-maker question* com o problema acerca do que estamos a fazer quando falamos de estados e eventos conscientes e lhes atribuímos propriedades. Esta última questão é apenas uma questão linguística, metodológica, epistemológica, não é uma questão metafísica acerca da natureza do mundo. Ora Jackson nota acertadamente que é o problema acerca do que estamos a fazer quando falamos de estados e eventos conscientes e lhes atribuímos propriedades, que ocupa Dennett em grande parte da sua teoria da consciência, pelo menos certamente na doutrina acerca de *qualia*. Para Jackson isto significa que Dennett rebate uma questão metafísica sobre questões epistemológicas e linguísticas bem menos fundamentais conseguindo assim passar ao lado do verdadeiro problema da consciência. É aliás essa a razão que conduz muitos dos críticos de Dennett a considerar que embora a teoria apresentada faça luz sobre um grande conjunto de fenómenos cognitivos, ela não é propriamente uma *teoria* da consciência no sentido de teoria que os filósofos tipicamente têm em mente. É exactamente essa a opinião de F. Jackson. É certo que Dennett, por seu lado, pensa que oferecer uma teoria no sentido que os filósofos tipicamente têm em mente, i.e. «preencher a fórmula (x) (x tem experiência consciente se e só se ...) e defendê-la dos contra-exemplos propostos não é um bom método para desenvolver uma teoria da consciência»[443]. Pelo contrário, não se terá uma teoria da consciência enquanto não se tiver um *modelo* que

[442] Acerca do que Jackson entende por metafísica, cf. JACKSON 1997: 483-484. De acordo com Jackson, «Metaphysics is about what there is and what it is like» (JACKSON 1997: 483). Isto significa que a metafísica procura completude na compreensão da maneira total como as coisas são e, mais especificamente, procura formular uma teoria englobante de alguma coisa (por exemplo a mente) em termos de «a limited number of more or less fundamental notions» (JACKSON 1997: 483). A metafísica dita séria, que Jackson opõe à elaboração de longas listas, «is simultaneously discriminatory and putatively complete, and the combination of these two factors means that there is bound to be a whole range of putative features of our world up for either elimination or location» (JACKSON 1997: 484). O materialismo ou fisicalismo é uma tese metafísica, aliás defendida e explorada por Jackson, que se caracteriza por ser «the very opposite of a big list metaphysics» (JACKSON 1997: 484). A questão metafísica colocada pelo materialista por exemplo em relação à consciência será então «What then is for some putative feature to have a place in the story some metaphysics tells in its favored terms? One answer is for the feature to be *entailed* by the story told in the favored terms» (JACKSON 1997: 484. O itálico é de Jackson, e esta noção de *entailment* traduz um "requerimento de necessidade" para a colocação de uma determinada característica. Este é um requerimento lógico e constitui o quadro para o tratamento da questão metafísica nos termos (de alguma variedade) da superveniência (cf. PINTO 1999: 34-64).

[443] DENNETT 1991: 459.

explicite as funções envolvidas. Esse modelo é um modelo mecânico, pertinente do ponto de vista científico-natural e não resulta de um trabalho filosófico de análise conceptual[444].

O (pelo menos suposto) extravio levaria, assim, Dennett a ignorar a possibilidade de a teoria da consciência requerer que alguma coisa fundamental fosse acrescentada à ontologia. Esse "algo de fundamental" estaria ligado ao facto de ser como alguma coisa ser uma entidade consciente, i.e. à característica que T. Nagel capturou com o seu célebre *What is it like to be ...?* [445]. A questão nageliana poderia ser respondida por certas entidades e não por outras, sendo o facto sintomático de uma diferença ontológica importante. "Ser como alguma coisa ser" (uma determinada entidade) deve ser contrastado com o facto de o processamento de informação, o controlo do comportamento, o cálculo, a inferência, ou a execução de inúmeras outras habilidades cognitivas não ser necessariamente sentido ou experienciado como alguma coisa. A característica de "ser como alguma coisa ser" não é identificável com os *qualia* tais como Dennett os define. Para filósofos como T. Nagel, J. Searle ou D. Chalmers a acima mencionada diferença entre ser como alguma coisa, subjectivamente, ser e a cognição enquanto processamento de informação que pode ocorrer "no escuro", i.e. sem qualquer sentimento de

[444] Ao contrário do que pensa F. Jackson, que defende com toda a clareza o papel da análise conceptual na defesa e elaboração do materialismo, inclusivamente no que diz respeito à consciência (cf. JACKSON 1997 e JACKSON 1998). Assim, «We should not accept the necessary a posteriori identifications by metaphysical fiat, but rather as flowing out of a piece of conceptual analysis combined with the relevant empirical information. We thus agree with those physicalists who see it as a prime task to provide a suitable analysis of consciousness as an essential preliminary task to making the physicalist identification they seek. Part of what makes the problem of consciousness so hard is that there is yet no analysis that comes close to holding the field» (BRADDON-MITCHELL & JACKSON 1996: 142). Evidentemente, por mais que Dennett negue a necessidade de análise conceptual, a sua teoria da consciência pressupõe um trabalho prévio de análise do que está a ser explicado (ou então nem haveria disparidade por exemplo entre Jackson e Dennett quanto ao que deve ser explicado quando se trata da consciência).

[445] O *what is it like to be...?* estabelece, para Nagel, uma irredutibilidade do ponto de vista de primeira pessoa ao ponto de vista de terceira pessoa. O ponto de partida é o seguinte: «Fundamentally an organism has conscious mental states if there is something that it is like to be that organism – something it is like for the organism (...) «I want to know what it is like for a bat to be a bat» (NAGEL 1974: 166-69). O propósito central de *What is it like to be a bat?* (NAGEL 1974) é estabelecer que por mais que saibamos sobre a neurofisiologia do morcego nunca poderemos saber, a partir desse conhecimento, como é ser um morcego (ou qualquer outra entidade consciente). Nagel pretende assim estabelecer a irredutibilidade da fenomenologia à neurofisiologia ou a qualquer outra teoria física e portanto a (pelo menos por enquanto) incompreensibilidade do fisicalismo. Este último aspecto significa que, na imagem de Nagel (NAGEL 1974), a afirmação "A mente é o cérebro" é tão incompreensível para nós como o teria sido para um pré-socrático (ou para um contemporâneo que não saiba física) a afirmação "A matéria é energia".

ser ou percepção dos processos cognitivos próprios, é o primeiro passo para o reconhecimento de algo que Dennett simplesmente ignora ou nega e que deve ser considerado o primeiro dado da teoria da consciência: a *subjectividade*. Esta poderia ser um traço irredutível da realidade, apenas apercebível em primeira pessoa, não identificável com "propriedades de segunda ordem" do apercebimento. Essa subjectividade de estados e processos mentais conscientes, uma característica que nenhum outro fenómeno natural possui, deve, de acordo com T. Nagel, ocupar «um lugar tão fundamental em qualquer visão do mundo credível como a matéria, a energia, o espaço, o tempo e os números»[446]. Nas palavras de J. Searle, devido à consciência, devido a ser como alguma coisa ser, sabemos mais acerca de como o mundo é. Nomeadamente, sabemos que o mundo é tal que contem subjectividade como um «elemento fundamental (*rock-bottom*)»[447]. É ilustrativo comparar estas declarações de princípio com a taxativa posição de Fodor, abordada no capítulo anterior, segundo a qual se a intencionalidade é real, então deve ser outra coisa qualquer, o que abre a possibilidade de resolver a questão da intencionalidade de forma reducionista e eliminativa. O ponto importante sublinhado pelos autores referidos é que o mesmo não se pode afirmar da consciência: afirmar que a consciência deve ser considerada fundamental é admitir que ela não poderá ser explicada nos termos de alguma coisa mais simples.

As posições de filósofos como Jackson, Chalmers, Nagel ou Searle, que procuram enfrentar o problema metafísico da consciência, não são simples formas de um novo misteriosismo (*new misterianism*)[448]. Este está, pelo menos numa primeira abordagem, mais associado ao piedoso pessimismo epistemológico de um Colin McGinn, para quem uma explicação adequada da consciência estará para sempre fora do alcance do entendimento humano, devido ao fechamento epistemológico do problema da consciência para os humanos. McGinn compara o fechamento epistemológico do problema da consciência para os humanos, a inacessibilidade da consciência fenomenal à explicação, com o fechamento epistemológico por princípio da mecânica quântica para um macaco[449], insistindo na relação entre a irresolubilidade do problema da consciência devido a uma (não particularmente misteriosa, como é claro pela comparação acima) limitação epistémica (*epistemic boundedness*). As posições de Jackson, Chalmers, Searle e Nagel acerca da irredutibilidade da consciência não dizem respeito a propriedades de segunda ordem do apercebimento nem a limitação cognitiva mas à forma como o mundo é, àquilo que fundamentalmente existe e à *truth-maker question,* na linguagem de Jackson. Ora, Dennett não encara explicitamente tais questões nas abordagens que faz do problema da consciência.

Mas a verdadeira pedra de toque da teoria proposta por Dennett, a marca máxima da heterodoxia, é a sua vulnerabilidade (ou indiferença) perante a ameaça do *zombie*. Perante modelos de consciência como os de Dennett, que serão analisados ao longo deste capítulo, e especialmente perante o MEM, poder-se-á sempre perguntar, como pergunta Chalmers, «*Why doesn't all this information-processing go on in the dark, free of any inner feel?*»[450]. A mais repetida acusação à teoria da consciência de Dennett – de acordo com a qual ela pura e simplesmente passa ao lado do problema da consciência na medida em que ignora a consciência fenomenal e a experiência sentida[451] – justifica a insensibilidade de Dennett aos *zombies*. Mas, precisamente, Dennett, ao contrário da maioria dos filósofos interessados no problema metafísico da consciência, não leva a sério a ameaça do *zombie*. A sua defesa consiste em declarar que a ideia de *zombie* é inconcebível, que o *zombie* é um tópico aberrante, discutido pelos filósofos com uma seriedade que envergonha a disciplina e que resume só por si uma má abordagem do problema da consciência. No artigo *The Unimagined Preposterousness of Zombies*[452] a

[446] NAGEL 1986:7-8.

[447] SEARLE 1992: 93-95.

[448] Evidentemente o próprio Dennett pensaria que sim. A expressão "*new misterianism*" é de Owen Flanagan, que a utiliza por exemplo em FLANAGAN 1992: 8-11 em relação a Colin McGinn (Nagel é considerado agnóstico). O *new misterianism* opõe-se ao *old mysterianism* que é o dualismo. Na opinião de Flanagan, «New misterianism can be shown to be at least premature, and perhaps some agnostics can be converted» (FLANAGAN 1992: 11).

[449] Cf. McGINN [1989] e McGINN 1990. A ideia de *cognitive closure* proposta por McGinn é inspirada por Nagel e Fodor. Fodor fala de *epistemic boundedness*, para fazer uma distinção entre tipos de problemas semelhante à conhecida distinção chomskyana entre mistérios e problemas (ao contrário do que acontece com os *mistérios*, os *problemas* podem ser *resolvidos*, vale a pena trabalhar sobre eles, não estamos epistemicamente limitados para os encarar).

[450] CHALMERS 1997:13.

[451] Cf. por exemplo BLOCK 1994b. Block é aliás um crítico de Dennett que raia a falta de caridade interpretativa. Block sublinha constantemente que a teoria de Dennett pode ser uma teoria do acesso, uma teoria do *software*, uma teoria da monitorização, uma teoria do eu, mas não é uma teoria da consciência. É claro que para Dennett uma teoria da consciência deve ser uma teoria do acesso, do *software*, da monitorização e do eu ou *não será uma teoria da consciência.*

[452] DENNETT 1998i.

questão é muito simplesmente afastada de maneira típica: se o *zombie* filosófico é de facto comportamentalmente indistinguível de um ser consciente, então ele é mesmo indistinguível e o defensor dos *qualia* tem que admitir que um *zombie* falará sobre os seus *qualia* e sobre como é sentir-se ser, terá crenças sobre *qualia*, sobre si próprio e sobre a sua vida interior, etc. Ora, esse *zombie* somos nós, e portanto (uma vez que é impossível estabelecê-la) a distinção entre *zombies* e seres conscientes não é uma verdadeira distinção. Os *zombies* filosóficos são simplesmente inimagináveis enquanto distintos de seres conscientes. Ou os *zombies* são inimagináveis, ou somos todos *zombies*. De qualquer modo, o conceito é inútil, pois não existem *zombies* lado a lado com outra coisa (os seres conscientes) de uma forma que seja pensável.

Poder-se-ia caracterizar as relações difíceis de Dennett com a consciência usualmente chamada fenomenal ainda de um outro modo. Mais do que ignorar a questão da consciência fenomenal, Dennett rebate-a pura e simplesmente sobre outras questões, nomeadamente sobre a questão do acesso a si de um sistema, que determina a disponibilidade de conteúdos para o controlo racional da acção e da fala, sobre a questão do eu, ligada ao uso de uma representação de si num sistema, e sobre a questão dos estados mentais de segunda ordem[453], estados mentais acerca de outros estados mentais, cuja possibilidade se liga à monitorização interna de um sistema. Aliás e como se verá, o(s) acesso(s), o eu e os estados mentais de segunda ordem são assuntos tratados de forma esclarecedora na parte construtiva da teoria da consciência, i.e. a partir de um modelo funcionalista.

Dennett é o caso típico do filósofo que N. Block tem em mente quando se refere a filósofos que deslizam de forma tácita da noção de consciência fenomenal «para um ou outro destes conceitos cognitivos ou intencionais ou funcionais (de consciência)»[454]. Na verdade, enquanto teórico da consciência, Dennett assume-se igualmente como o behaviorista- caracterizado-por-Jackson[455], i.e. alguém que responde à *truth-maker question* através da *superveniência psi-*

[453] A idcia de *higher-order thought* é explorada por David Rosenthal (cf. por exemplo ROSENTHAL 1997) e será de facto importante para Dennett em CE, como se verá mais à frente. Toda uma parte (Parte X) da antologia BLOCK, FLANAGAN & GÜZELDERE 1997 é dedicada às *Higher-Order Monitoring Conceptions of Consciousness.*

[454] BLOCK 1994a: 213.

[455] JACKSON 1993.

cofísica[456]. Para um tal behaviorista, se dois organismos são comportamentalmente exactamente idênticos eles serão, necessariamente, psicologicamente idênticos. O facto de aceitar a caracterização de Jackson significa que Dennett se define como materialista aceitando o mínimo do materialismo como tese metafísica: uma ideia de *superveniência.* Ora, é precisamente no quadro da superveniência que decorre uma grande parte das discussões filosóficas actuais sobre consciência. A formulação mais simples da superveniência enquanto relação entre características ou propriedades básicas da realidade (por exemplo físicas) e outras características que sobrevêm a estas faz apelo à *indiscernibilidade* (se duas coisas são indiscerníveis quanto às suas propriedades físicas (propriedades-F), não haverá entre elas diferenças quanto às propriedades supervenientes (propriedades-S); as coisas são S em virtude de serem F e quaisquer coisas que sejam idênticas em termos de propriedades-F serão idênticas em termos de propriedades-S) e à *dependência assimétrica* (são as propriedades-S que dependem das propriedades-F e não o contrário), e não envolve considerações sobre causalidade[457]. Ao contrário de um grande número de filósofos que se dedicam ao problema da consciência, Dennett não vê qualquer interesse na exploração teórica da ideia de superveniência, por exemplo na exploração da diferença entre a superveniência lógica e a superveniência natural, que precisamente cria "espaço" para a caricaturada noção de *zombie*[458].

Como virá a tornar-se claro, é a amálgama voluntária (ou o esclarecimento, conforme a perspectiva) da consciência com conceitos cognitivos como o acesso, o eu ou os pensamentos sobre pensamentos que justifica a insensibilidade de Dennettt perante os *zombies,* a consciência fenomenal e a discussão da superveniência. Nos termos de D. Chalmers, para um funcionalista como Dennett, que

[456] Melhor dizendo, Dennett não faz nenhuma questão de distinguir o behaviorismo do funcionalismo ou do eliminativismo, considerando que a sua teoria da consciência é susceptível de todas essas classificações. Ora, este é precisamente o ponto que perturba Jackson, que pensa que Dennett não é de todo explícito quanto ao tipo de complexidade (comportamental, funcional ou outra) que implica ou acarreta necessariamente a consciência.

[457] Cf. PINTO 1999: 5-15.

[458] Por exemplo de acordo com Chalmers (CHALMERS 1996), os *zombies* são logicamente possíveis mas não naturalmente possíveis, e essa condição é relevante no que diz respeito ao estatuto da consciência fenomenal no nosso mundo. Dennett, no entanto, simplesmente aglutina a possibilidade física ou natural com a possibilidade lógica nas suas discussões (cf. CHALMERS 1996: 370).

dá como garantida a superveniência *lógica*[459] da consciência às propriedades funcionais de um sistema cognitivo, o problema da consciência não chega a ser de facto um problema. Dennett tem ainda uma outra defesa perante as acusações de insensibilidade relativamente à experiência fenomenal e à possibilidade de *zombies:* esta consiste em declarar que não existem questões de facto relativas à consciência fenomenal que escapem aos modelos funcionalistas que apresenta. A profissão de fé de verificacionismo relativamente à consciência que subjaz a esta afirmação baseia-se no facto de, no que respeita à consciência, o *an-sich* simplesmente ser o *für-mich.* Se existir fenomenologicamente é aparecer, nada pode haver de consciente e oculto, e Dennett considera que os seus modelos dão conta do estatuto de tudo aquilo que aparece. É por isso mesmo, devido a tal pretensão de exaustividade, que é essencial ter como referência a posição de filósofos que consideram o problema da experiência fenomenal o problema limite para o materialismo e que levam a sério a ameaça do *zombie*[460] para caracterizar a teoria da consciência defendida por Dennett. A especificidade da posição de Dennett reside precisamente no facto de ela impedir qualquer suposição de um carácter especial do problema da consciência.

Concluir-se-á a presente tentativa de delinear o outro prisma de abordagem filosófica da consciência com o resumo, feito por D.

[459] I.e. alguém para quem a relação de "necessitação" entre as propriedades básicas e as propriedades supervenientes (no caso a consciência) é explicitável unicamente em termos lógicos: seria *contraditório* que entidades descritas como tendo determinadas propriedades (os funcionamentos que o MEM caracteriza) não fossem conscientes.

[460] Cf. PINTO 1999 para uma discussão abrangente do problema da experiência fenomenal na filosofia da mente, estruturada em torno da utilização da noção de superveniência para caracterizar o materialismo e para o tratamento teórico da experiência fenomenal. Duas referências cruciais são a obra de J. Kim sobre a superveniência (cf. nomeadamente KIM 1993) e o tratamento da fenomenalidade da experiência por D. Chalmers (CHALMERS 1999) a partir dela. De acordo com o sumário do trabalho referido (PINTO 1999: 3-4) procede-se a: (1) Apresentação do conceito de superveniência e das suas variedades, (2) Apresentação do materialismo ou fisicalismo como posição metafísica, (3) Exposição dos dilemas do funcionalismo na filosofia da mente, dilemas estes que não conduzem ao abandono do funcionalismo mas sim à explicitação do seu compromisso com a variedade local de superveniência e à necessidade de integração do funcionalismo numa "teoria plenamente fisicalista" dos estados mentais, (4) Reunião da variedade global da superveniência com o teleofuncionalismo, (5) Caracterização do materialismo como tese de superveniência global do género lógico, (6) Problema criado ao materialismo pela ocorrência de experiências fenomenais (no âmbito de cuja análise se procede ao afastamento de dissoluções do problema do género da dennettiana, de que se tratará daqui em diante, cf. PINTO 1999: 122-123) e finalmente, (7) Razões para preferir uma explicação fisicalista da experiência, exame da noção de superveniência como "saída" do materialismo. Uma investigação deste género é o exemplo do que está em causa em abordagem filosóficas da consciência "alternativas" à abordagem dennettiana.

Chalmers[461], dos tipos de argumentos que em conjunto sustentam o caso oposto ao de Dennett. Este fica, a partir de agora caracterizado como o tipo de abordagem da consciência que *requer a apresentação de um modelo funcionalista, pressupõe a superveniência lógica e o verificacionismo, e considera que isso é tudo o que é necessário para explicar a consciência.* Pelo contrário, os argumentos em seguida listados sustentam a pretensão de um "abismo" (*gap*) explicativo e ontológico entre a consciência e o resto e a existência um problema difícil da consciência. Esses argumentos, que voltaremos a encontrar em várias versões ao longo do capítulo, dizem respeito a: (1) possibilidade lógica de *zombies*[462], (2) possibilidade lógica de espectros invertidos[463], (3) assimetria epistémica entre o conhecimento do facto de se ser consciente e o conhecimento de todos os outros factos[464], (4) o caso de Mary, a neurocientista da cor (o Argumento do Conhecimento de F. Jackson)[465] e (5) ausência de uma análise não circular daquilo que é "experiência"[466]. Dennett precisa de desenvolver uma posição perante todos estes argumentos de modo a fundamentar o seu caso.

[461] CHALMERS 1996: 94.

[462] Que Dennett nega.

[463] Que Dennett nega.

[464] Como Chalmers nota (CHALMERS 1996: 102), «Eliminativism about conscious experience is an unreasonable position only because of our own acquaintance with it. If it were not for this direct knowled, consciousness could go the way of the vital spirit».

[465] Cf. JACKSON [1986], *What Mary Didn't Know*, onde é apresentado o chamado Argumento do Conhecimento (*knowledge argument*). O propósito do Argumento do Conhecimento é mostrar que o fisicalismo (definido como a doutrina segundo a qual conhecer todos os factos físicos é conhecer tudo o que há para conhecer) é falso. Especificamente, o caso de Jackson pretende mostrar que conhecer todos os factos físicos acerca de visão de cor não é suficiente para saber como é ver cores. A ideia é que Mary, uma neurocientista da cor, que vivera isolada da experiência de cor, fechada num quarto preto e branco, aprendera tudo o que há para saber acerca da natureza física do mundo e no entanto aprende alguma coisa de novo quando sai cá para fora e pela primeira vez vê cores. O argumento segue a linha de NAGEL 1974, que também argumenta que mesmo conhecendo os factos físicos não saberemos ainda como é a experiência subjectiva (não saberemos por exemplo como é para um morcego ser um morcego). Com o Argumento do Conhecimento Jackson pretende estabelecer a falha da superveniência lógica da consciência fenomenal sobre os factos físicos. Jackson acentua (JACKSON 1986) que o caso de Mary não pretende sublinhar nem uma incapacidade de imaginação (antes, ela pura e simplesmente não poderia saber) nem uma incompletude, por falta de desenvolvimento lógico das consequências, do conhecimento de Mary. Mary descobre que, apesar do seu conhecimento de todos os factos físicos, ela tivera uma concepção empobrecida da vida mental, e nomeadamente da vida mental das outras pessoas. Nem todos os factos são portanto factos físicos: existem outros factos acerca da vida mental das pessoas que não são factos físicos.

[466] As duas partes não se compreendem, falam de coisas diferentes. Cf. CHALMERS 1997: 385 «Dennett challenges me to provide "independent" evidence (presumably behavioral or functional evidence) for the postulation of experience. But this is to miss the point: conscious experience is not postulated to explain other phenomena in turn; rather it is a phenomenon to be explained in its own right.».

Chamar-se-á em geral à teoria da consciência apresentada por Dennett uma *teoria deflacionária*, i.e. uma teoria que desvaloriza ou diminui aquilo que se pensava existir. O aspecto deflacionário relaciona-se não apenas com a desconstrução da ideia de fenomenologia e com a eliminação dos *qualia* mas também com a pretensão segundo a qual a consciência não envolve uma "descontinuidade" ou abismo relativamente ao resto da realidade física. Em suma, e de novo nos termos de D. Chalmers, a estratégia de Dennett é tentar justificar que o *hard problem* da consciência fenomenal não é senão mais um "problema simples" como o são as questões respeitantes a processamento de informação, categorização, controlo de acção, acesso interno e produção de relatos verbais de estados internos. Não há razão para proclamar mistérios (acerca da consciência fenomenal), há sim necessidade de elaborar modelos. Ora, quando se trata de elaborar modelos, aquilo que a teoria dennettiana da consciência oferece é, nas palavras de J. Searle, um exemplo de IA Forte. Na descrição que J. Searle faz do MEM[467], este corresponde à pretensão segundo a qual ser consciente é implementar um programa que instala virtualmente serialidade numa máquina paralela que evoluiu por selecção natural, o cérebro. E de facto o modelo de consciência defendido em CE propõe que a consciência de cada indíviduo consciente consiste não numa "aparição fenomenal" mas na operação de uma Máquina Virtual[468] instalada no cérebro (este é de resto mais um aspecto do "deflacionismo" dennetiano: a realidade da consciência é de acordo com o MEM de alguma forma uma realidade *virtual*).

Ao longo de todo o presente capítulo procurar-se-á expor as proposta de Dennett quanto à consciência e avaliá-las, concedendo à partida que existem bons motivos para defender a separação do tratamento teórico da intencionalidade e da racionalidade relativamente ao tratamento da consciência. Dennett não é o único a defender esta separação: é bastante plausível pensar que a inteligência e a intencionalidade estão muito mais espalhadas do que a consciência. Uma situação como a vitória de Deep Blue sobre Kasparov num jogo de xadrez (que aconteceu em 1997, mas já era esperada há muito mais tempo) torna a existência de entidades muito inteligentes, com "vidas internas" ou "psicologias" complexas, e que são no entanto totalmente inconscientes, incontornável para o teórico da cognição. De facto vivemos rodeados por elas. Definitivamente não é como alguma coisa ser Deep Blue, e no entanto há toda uma teoria cognitiva a ser feita acerca das suas capacidades. É talvez por isso que, como nota frequentemente J. Fodor, dividir e conquistar, concentrarmo-nos na intencionalidade e ignorar a consciência, tem mostrado ser uma estratégia de investigação notavelmente bem sucedida[469]. Fodor "divide e conquista" deixando a consciência de lado. Dennett divide também mas pretende levar a conquista mais longe. Pretende subsumir por princípio a teoria da consciência numa teoria da intencionalidade, arriscando-se a fazer desaparecer a aparição da consciência. É essa extensão máxima da posição de Dennett que estará em causa.

Se o problema da consciência se tornou *o problema* para a filosofia da mente nos anos 90, o facto é que Dennett tem vindo a apresentar modelos de consciência desde os anos 60. Passar-se-á em seguida à análise do desenvolvimento desses modelos, retomando o incipiente modelo apresentado no Capítulo 1. Verificar-se-á que embora os rebatimentos do problema da consciência fenomenal sobre os problema relativos ao acesso, à monitorização, ao controlo e ao eu, bem como o rebatimento do problema metafísico da consciência sobre o problema epistemológico e linguístico dos relatos de estados mentais próprios, coloquem Dennett em polémica aberta com muitos filósofos, desse rebatimento resultam várias contrapartidas benéficas. Nomeadamente, os modelos de consciência propostos fornecem instrumentos para a análise de um grande número de problemas empíricos da ciência cognitiva, desde a produção de linguagem, às imagens mentais, ao sonho, à visão da cor e às patologias neuropsicológicas, como a visão cega (*blindsight)* ou a desordem de personalidade múltipla. Para analisar a evolução da teoria da consciência de Dennett ao longo de mais de trinta anos serão utilizados os seguintes fios condutores: (1) os modelos funcionalistas de consciência apresentados em BS e CE, (2) o posicio-

[467] SEARLE 1997: 97-115. A opinião de Searle sobre a teoria da consciência proposta em DENNETT 1991 é assim resumida: «The net effect is a performance of Hamlet without the Prince of Denmark» (SEARLE: 100). Searle afirma ainda que Dennett defende que todos somos *zombies*.

[468] "Máquina Virtual" é um termo utilizado na ciência de computadores para referir estruturas "temporárias" feitas de regras e não directamente de materiais, nomeadamente programas.

[469] Cf. também CHALMERS 1996:83: «There is no argument analogous to the argument against the supervenience of consciousness showing that intentionality cannot supervene logically on physical and phenomenal properties. Indeed, conceivability arguments indicate that intentional properties must be logically supervenient on these if such properties are instantiated at all, and epistemological arguments lead us to a similar conclusion. So there is no separate ontological problem of intentionality».

namento metodológico face às questão empíricas da consciência que estes modelos justificam e (3) as implicações do modelo mais elaborado, o modelo de CE, relativamente aos problemas filósóficos standard da consciência (formulados em termos de *qualia e zombies*, *explanatory gap*, *hard problem versus easy problems*, *what-it's-like*, Mary-a-neurocientista-da-cor, Quarto Chinês, etc)[470].

A estratégia adoptada por Dennett para enfrentar o problema da consciência está patente na própria estrutura de CE e consiste em deixar para o fim o momento em que se encara a formulação e a discussão filosóficas ortodoxas do problema da consciência. Dennett começa sempre pela teoria empírica, pelo ponto de vista de terceira pessoa, partindo não de análises conceptuais mas de teorias da evolução biológica da consciência e de modelos da evolução no cérebro. É certo que a teoria envolve também, ou pressupõe, um trabalho metodológico, e não já propriamente empírico, de análise das relações entre os relatos verbais que os sujeitos fazem dos seus estados mentais, e os "responsáveis causais" por esses estados mentais. Na forma madura de CE, esta é a função do método heterofenomenológico, um método descritivo supostamente neutro. De forma reveladora, o mundo heterofenomenológico identifica-se com o mundo nocional do capítulo anterior. Além disso, de acordo com as ideias propostas, a consciência está longe de ser um traço único, especial e básico do mundo, como o é para filósofos como Nagel ou Chalmers e em geral para qualquer pessoa para quem a consciência não seja o resultado contingente da instalação de linguagens naturais em *gadgets* biológicos de natureza local como os cérebros humanos.

3.2 *Modelos de consciência e natureza das experiências:* Brainstorms *(1978).*

«Having an inner life is (...) a matter of having a certain sort of functional organization» (...) «I propose to construct a full-fledged "I" out of sub-personal parts by exploiting (...) notions of access», Daniel Dennett[471]

«I suggest that a theory of consciousness should take experience as fundamental», David Chalmers[472]

[470] Terminologia de, respectivamente, J. Levine, D. Chalmers, T. Nagel, F. Jackson e J. Searle.
[471] DENNETT 1978: 71, seguido de DENNETT 1978: 154.
[472] CHALMERS 1997: 19.

O psicólogo Philip Johnson-Laird afirmou uma vez que qualquer teoria da mente que se pretenda científica deve tratar a mente como um autómato[473]. No que se segue procurar-se-á expor aquilo que Dennett faz nesse sentido com os seus modelos de consciência. Em BS[474] Dennett procura aperfeiçoar a noção de consciência como auto-acesso linguisticamente apercebido e incorrigível apresentada em C&C. O aperfeiçoamento passa pela proposta de um modelo funcionalista, explicitado em *Toward a Cognitive Theory of Consciousness.* Dennett importa da psicologia cognitiva para a filosofia da mente o uso de diagramas, notório em psicólogos cognitivos como Philip Johnson-Laird ou Bernard Baars[475], uso esse ele próprio inspirado nos fluxogramas da ciência de computadores. O modelo apresentado em *Toward a Cognitive Theory of Consciousness*, com os seus sub-sistemas funcionalmente individuados (ou módulos) e o fluxo de informação entre eles, é disso exemplo. No modelo a consciência não aparece como mais um módulo mas antes como difusão ou "publicação" dos eventos em larga medida inconscientes que guiam o comportamento do sistema. Através do modelo apresentado, Dennett propõe-se construir o eu consciente (ou acesso pessoal) a partir de partes sub-pessoais. A intenção que preside ao modelo é, pelo menos em parte, criticar a importância exagerada concedida à mente consciente e à sua eficácia causal na geração da acção. Apesar disso, o "componente-controlo" tem ainda um lugar de relevo no modelo de BS, o que não acontecerá no modelo de produção por pandemónio[476] de CE. Homúnculos e acessos (pessoal e sub-pessoais) são fundamentais no modelo. A consciência é identificada com o acesso pessoal. Este distingue-se de dois outros tipos de acesso sub-pessoal, o acesso computacional e o acesso público, que retomam a distinção estabelecida em C&C entre apercebimento-2 e apercebimento-1 e dizem respeito a relações entre partes de sistemas (por oposição a relações entre o sistema global e estados de coisas no mundo).

Por definição, «Aquilo de que eu estou consciente é aquilo a que eu tenho acesso, ou (para colocar o ênfase no sítio devido) aquilo a que *Eu* tenho acesso. Vamos chamar a este tipo de acesso o acesso da consciência pessoal para acentuar que o sujeito desse acesso (que

[473] JOHNSON-LAIRD 1983: 477.
[474] Sobretudo na Parte III (*Objects of Consciousness and the Nature of Experience*)..
[475] Cf. por exemplo BAARS 1988 e JOHNSON-LAIRD 1983.
[476] O termo é de Oliver Selfridge e o seu uso será generalizado em DENNETT 1991, nomeando antes de mais um processo *acentrado*.

exaure a consciência) é a pessoa e não alguma das partes da pessoa»[477]. Quanto à noção de acesso computacional, a própria expressão remete para aquilo que a inspira: a programação estruturada em sub-rotinas. O acesso computacional diz respeito ao fluxo de informação entre diferentes rotinas. O acesso computacional não se relaciona directamente com o acesso pessoal (por exemplo, uma pessoa não tem acesso pessoal à informação tratada no seu sistema visual, embora o acesso computacional entre partes sub-pessoais do sistema visual deva existir para que a pessoa possa estar a ver). É sobre a noção de acesso público (que é também um acesso sub-pessoal, entre partes do sistema) que recai agora o privilégio da linguagem na teoria da consciência. Para explicar a noção de acesso público é de novo evocada a ideia de rastros de programas: «Frequentemente é útil a um programador ter acesso ao que o computador está a fazer, de modo a que o progresso do computador com o programa possa ser monitorizado: para isto providencia-se um rastro (*trace*) do programa, de modo a que o computador possa imprimir, passar para fora (*print out*), informação acerca dos passos intermédios das suas próprias operações»[478]. Acesso público deste tipo envolve uma sub-rotina com acesso computacional a determinados resultados do processamento interno e um rastro exterior desses resultados, no caso humano por exemplo a fala. Embora o acesso computacional e o acesso público sejam ambos sub-pessoais, eles têm domínios diferentes. O primeiro diz respeito àquilo que está a ser utilizado para controlar o comportamento do sistema, o segundo àquilo que está disponível para ser relatado. Que se trata de coisas diferentes era já, recorde-se, o ponto da distinção entre apercebimento-2 e apercebimento-1. Pelo facto de sermos criaturas linguísticas, o acesso público parece-nos próximo do acesso pessoal ou mesmo idêntico a ele. Dennett defende no entanto que o "sujeito" do acesso público não é o eu (ou melhor, defende que o acesso público não tem "sujeito", i.e. que o querer-dizer não é agenciado por um eu). Com estas três noções de acesso, parte-se para a modelação da consciência num sistema físico organizado. Supostamente, se uma entidade física organizada em departamentos intercomunicantes estivesse "ligada" tal como o fluxograma que apresenta o modelo exemplifica[479] seria como alguma coisa ser essa entidade.

[477] DENNETT 1978g: 150.

[478] DENNETT 1978g: 151.

[479] Cf. DENNETT 1978g: 155 para o diagrama original. O diagrama que se apresenta foi simplificado, ignorando nomeadamente a expressão gráfica dos diferentes níveis nos quais a relação entre componentes acontece.

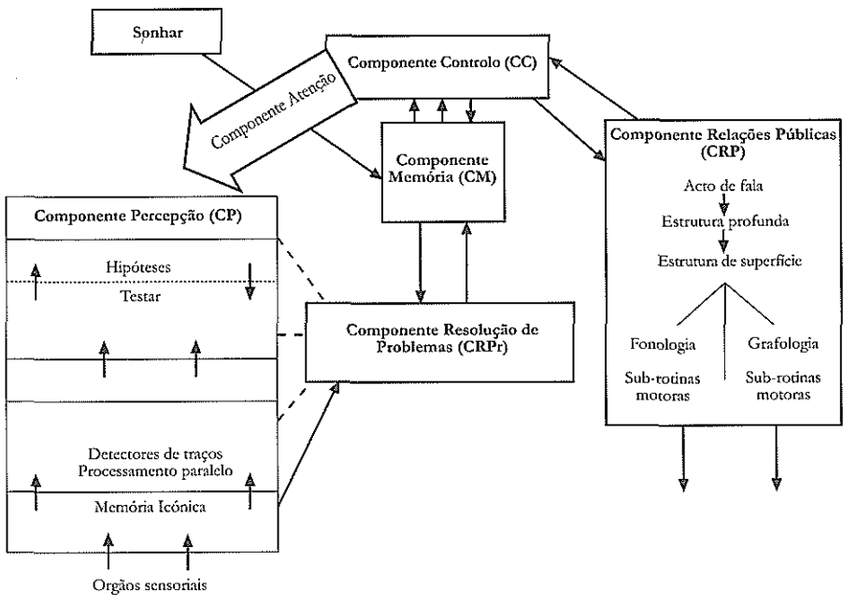


FIG 2. O modelo funcionalista de consciência de *Brainstorms*

Em concreto, no modelo apresentado, existem seis caixas que representam seis sub-componentes funcionais[480]. No extremo do output, existe um *Componente Relações Públicas* (CRP). O input deste componente são intenções semânticas, i.e. o componente recebe "ordens" para cumprir actos de fala e executa essas ordens. Este componente (CRP) recebe as ordens de um componente executivo (o *Componente Controlo*, CC). O CRP só tem acesso à informação através de um particular armazenamento em memória, a que se chama CM, o *Componente Memória*. Dados estes componentes, por exemplo uma rotina de introspecção passar-se-ia da seguinte maneira: o CC "decide" entrar na sub-rotina de introspecção, e "dirige uma questão" ao CM. Quando a resposta chega pode "avaliá-la" (censurá-la, interpretá-la, inferir a partir dela) ou passá-la directamente para o CRP. O resultado destes passos é uma ordem

[480] Excluindo a caixa "sonhar", que se reproduz devido à análise dos sonhos que se fará mais à frente.

de fala dada ao CRP, que executa a ordem e produz um acto de fala[481].

É claro que um item deve estar no CM para poder ser acedido pelo CRP como conteúdo de uma intenção semântica (embora estar no CM não seja condição suficiente para tal). Logo, é preciso saber o que vem a estar no CM. Ora, grande parte do que está no CM chega da Percepção. O modelo necessita assim de um quarto componente, um *Componente Percepção*, CP. Nem todo o tratamento computacional, incluindo a "geração e teste de hipóteses" que a percepção envolve e que o modelo considera, chega ao CM. A análise perceptual é aliás governada por instruções do CC, que tem que gerir os recursos cognitivos e dirigi-los para aquilo que é mais importante. Esta adscrição de recursos cognitivos constitui a Atenção. O *Componente Atenção* (CA), o quinto componente do modelo, está apenas indirectamente ligado com a consciência. Aquilo que chega ao CM do CP provem de muitos níveis da análise perceptual. Muito desse material não será sequer "publicitado" (pelo CRP). Dennett não identifica o que é apercebido com o que pode ser "publicitado", em contrapartida estabelece que se alguma coisa pode ser "publicitada" ela foi apercebida, i.e. registada no CM. Além do que chega de CP, chegam ao CM resultados de um *Componente de Resolução de Problemas* (CRPr, o sexto componente). O próprio CC também envia informações (planos, intenções, etc) para o CM.

As condições da consciência como acesso pessoal são então as seguintes de acordo com este modelo:

(1) Como já se defendera em C&C, *o acesso computacional para controlo do comportamento é independente do acesso computacional para publicitação.* Tal facto justifica a não-fiabilidade da evidência introspectiva (=acesso público) quanto ao que se passa no sistema e está envolvido no controlo do comportamento (=acesso computacional).

(2) O CRP pode receber "ordens para actos de fala" e não conseguir executá-las por não ter como, por não ter por exemplo palavras. O CC pode lidar com a situação dirigindo recursos para o CP, e produzindo uma nova percepção. Um processo assim conduz ao enriquecimento perceptivo e explica a sofisticação da capacidade de discriminação (por exemplo de um provador de vinhos ou de um pintor) por revisão perceptual.

[481] Flanagan (FLANAGAN 1992) comenta que os relatos de experiência consciente são como *press releases* de um porta voz governamental que está fora do círculo de decisão e que é um dos últimos a saber.

(3) *A questão do formato da informação* (nomeadamente da sua codificação proposicional ou imagética) é adiada a este nível. O modelo de BS é feito "do topo para a base" (*top-down*), como é usual na psicologia cognitiva, e considera apenas o *conteúdo* de estados e eventos conscientes, deixando de lado a estrutura dos veículos. Esta é uma questão de engenharia, acerca do tipo de sistema físico que cumpriria as funções descritas, embora bastante mais abstracta do que por exemplo questões acerca da neurofisiologia do sistema. Em BS a questão do formato das representações será retomada no artigo *Two Approaches to Mental Images* e a decisão de adiamento supõe que uma questão empírica de "engenharia psicológica" não pode por princípio ser respondida através do acesso pessoal. Como a questão também não é também directamente neurofisiológica, é impossível abordá-la exclusivamente da base para o topo (*bottom-up*). Dennett pensa que questões semelhantes requerem abordagens aprioristas (como o são por exemplo as abordagens da IA), embora não abordagens aprioristas em primeira pessoa, e chama a atenção para a continuidade entre a questão epistemológica geral (a questão kantiana de duas faces "Como é possível o conhecimento?" e "Como é que qualquer entidade pode experienciar o que quer que seja?") e as disciplinas empiricamente constrangidas, mas também aprioristas, da cognição, muito especialmente a IA. Aliás, a referida continuidade é analisada em BS, no artigo *IA as Philosophy and as Psychology*[482], e considerada como uma continuidade entre os problemas epistemológicos, problemas psicológicos da informação e problemas da engenharia dos veículos do conteúdo.

A IA é assim, como a filosofia, uma investigação abstracta dos princípios da psicologia. Para caracterizar o que se passa num sistema cognitivo, como inquérito apriorista feito do topo para a base, toma sistemas cognitivos globais (por exemplo pessoas) e divide-os

[482] DENNETT 1978h. Dennett nota aí que se a IA e a filosofia parecem aos psicólogos ser domínios aparentados é porque em ambas as disciplinas se investiga a priori o que é possível em termos de cognição com pouca atenção aos resultados experimentais (entendidos por exemplo como tratamento numérico de resultados de performances ou modelos neurofisiológicos). A IA é para Dennett a evidência da vantagem de uma estratégia *top-down* no estudo da cognição. Os seus programas são experiências de pensamento. O perigo correlativo é a suposição de sub-sistemas com capacidades miraculosas, o tomar de necessidades condicionais por constrangimentos gerais, a restrição artificial da análise a sub-sistemas sem conexão com outros, a restrição de performances de sistemas a fragmentos rduzidos de um domínio natural de cognição: «What is needed is a technical language with which to disciple one's imagination» (Z. Pylyshyn, citado em DENNETT 1978h: 118). Dennett afirma por sua vez que «The discipline provided by computers is undeniable (...) the use of the technical language Pylyshyn speaks of can cripple an imagination in the process of disciplining it».

em sub-sistemas ou agentes ou homúnculos com funções especializadas (como a decomposição que o modelo de *Toward a Cognitive Theory of Consciousness* apresenta, embora os homúnculos deste sejam ainda demasiado sofisticados). Os agentes são por sua vez constituídos por outros agentes mais "estúpidos", e assim sucessivamente, até se chegar a componentes mecânicos que permitirão pura e simplesmente dispensar (*dismiss*) os homúnculos ou agentes. Só uma abordagem assim pode mostrar a solução para o problema de Hume.

Quanto à questão decisiva (*Seria como alguma coisa ser esta organização?*) Dennett defende que a resposta é afirmativa. Para alguém que lidasse com a entidade a partir de fora, haveria certamente a percepção de que para ela é como alguma coisa ser. Ela diria como é ser e descreveria o que se passa consigo. À objecção segundo a qual a ilusão aí envolvida se prova pelo facto de que o interior do sistema é "escuro" e nada nesse interior se sente ser, Dennett responde que exactamente o mesmo acontece com o interior do nosso cérebro. Nós somos o primeiro "modelizado" por este modelo, uma organização que tem acesso consciente aos resultados do processamento cognitivo mas não aos próprios processos, e que dá por si a "querer dizer", a gerar episódios proposicionais, sem qualquer ideia acerca do que causa esses estados.

Para Dennett – mais uma prova da persistência da inspiração ryleana no seu pensamento[483] – quando se procura comparar este modelo funcionalista com aquilo que nós somos, constatamos que o tipo de evidência de que dispomos para pensarmos naquilo que somos consiste exactamente nos nossos próprios pronunciamentos não estudados. Em geral, episódios proposicionais, pensamentos de que p, são o caminho do auto-conhecimento e esgotam o apercebimento imediato que temos de nós mesmos. Objectar-se-ia talvez que o nosso auto-apercebimento imediato envolve também entidades fenomenais, como por exemplo imagens mentais. Dennett afirma no entanto que as imagens mentais atestadas em experiências psicológicas como as de R. Shepard[484] são "supostos" (*posits*) análogos aos que Hume[485] considerava explicarem a continuidade e a persistência independente do pensamento (não apercebidas) do mundo real exterior. Lista-se em seguida os princípios gerais que subjazem ao modelo funcionalista de BS:

[483] RYLE 1949: 194.

[484] Dennett refere-se aqui a SHEPARD & METZLER 1971. Os estudos de imagens mentais serão analisados no ponto 3.2.2.

[485] HUME [1739].

(1) *O apercebimento é mais alargado do que a "experiência consciente"e é simplesmente identificável com disponibilidade na memória.*

(2) O conteúdo do apercebimento é tudo aquilo que entra no CM. *No entanto, esse conteúdo pode degradar-se ou sofrer interferências antes de ser recuperado para acesso público*, já que a expressão ou "publicitação" de qualquer conteúdo envolve um acesso sub-pessoal entre o CM e o CRP. Logo, não é possível afirmar sem margem de dúvida que aquilo que alguém dá por si querendo dizer mostra infalivelmente o que foi / está a ser experienciado.

(3) *O acesso pessoal (ou consciência) acontece através do acesso entre o CM e o CRP sem observação interna ou introspecção.* O sentimento de autoridade introspectiva advém do facto de serem as intenções semânticas (concebíveis como tendo por base a relação funcional: CC→CRP) o standard pelo qual são "medidas" as produções verbais de um sujeito. Se o sujeito diz o que quer dizer, não haverá "injustiça" no acesso à vida interior. Contra o que afirmara em C&C acerca de incorrigibilidade, Dennett considera agora estar a explicar o facto de as pessoas pensarem que são incorrigíveis, o seu sentimento de autoridade subjectiva, e não uma real incorrigibilidade (i.e. uma infalibilidade).

Em geral, o modelo ilustra a tese segundo a qual ter uma vida interior é uma questão de organização funcional e não de uma especial "luz acesa no interior" (esta é evidentemente uma alusão à consciência fenomenal referida no início deste capítulo). Nas palavras do autor «O que me convence de que uma teoria cognitivista poderia capturar todas as características que eu descubro na minha vida interior e que me são caras, não é um argumento e nem é apenas o apelo programático a preservar assim qualquer coisa como a unidade da ciência, mas antes uma tentativa detalhada de descrever para mim próprio exactamente essas características da minha vida e a natureza do meu contacto (*aquaintance*) com elas que eu citaria como fundamentos para afirmar que eu sou – e não meramente pareço ser consciente»[486].

Em concreto, o modelo de BS conduz a um posicionamento metodológico determinado perante o estatuto de fenomenologias empiricamente estudadas. Em BS Dennett analisa os casos dos sonhos (no artigo *Are Dreams Experiences?*[487]), das imagens mentais

[486] DENNETT 1978g: 173.

[487] DENNETT 1978j.

(no artigo *Two Approaches to Mental Images*[488]) e da dor (no artigo *Why You Can't Make a Computer that Feels Pain*[489]).

3.2.1 *Experiência, memória, apresentação, expressão: os sonhos.*

A análise do estatuto de experiência dos sonhos feita em *Are Dreams Experiences?* revela claramente o princípio verificacionista que rege a teoria da consciência e que continuará a ser explorado no modelo de CE e especialmente no artigo *Time and the Observer*[490]. Os problemas abordados por meio da questão dos sonhos são por um lado a relação entre experiência, apresentação, memória e expressão e por outro o problema de saber a que evidência se pode justificadamente apelar para determinar se uma determinada experiência existiu ou não[491]. Estes são problemas gerais para uma teoria da experiência, não problemas específicos dos sonhos, e o seu núcleo é a questão do *peso epistémico do sentimento de autoridade subjectiva*. Numa primeira caracterização o sentimento de autoridade subjectiva marcaria a infalibilidade das crenças acerca de estados de consciência próprios. Num âmbito mais estrito, trata-se de decidir que sentido poderia ter (se é que poderia ter sentido) uma "infirmação" do conteúdo de tais crenças por via da investigação empírica.

É impossível definir os sonhos como experiências que ocorrem durante o sono, como se faz vulgarmente, sem se ser obrigado a constatar que se pressupõe assim a possibilidade de *experiência* numa entidade de alguma forma *inconsciente*. Uma análise do alcance da definição torna-se imperativa. Ela revelará dois pontos desejados por Dennett: (i) os sonhos permitem uma certa substituição da apresentação por memória, "substituição" esta que Dennett pretende alargar à teoria da experiência em geral; (ii) os sonhos permitem clarificar a distinção existente entre a consciência e outras formas de apercebimento com as quais ela é frequentemente confundida.

O pretexto da análise levada a cabo em *Are Dreams Experiences?* são as teses defendidas pelo filósofo N. Malcolm em *Dreaming*

[488] DENNETT 1978i.

[489] DENNETT 1978l.

[490] DENNETT&KINSBOURNE 1992, sobre a relação entre o tempo e a consciência.

[491] O problema da existência de experiência de x também se coloca relativamente à situação de vigília. No caso da vigília a questão da determinação da existência de experiência coloca-se por exemplo em relação a alguma coisa vista durante um muito curto intervalo de tempo (por exemplo, 40 milésimos de segundo).

(1959)[492]. Na opinião de Dennett, na obra referida, Malcolm apenas consegue salvar o sentimento de autoridade subjectiva ligado às asserções de experiência interior cedendo a um conservadorismo epistemológico inaceitável, de acordo com o qual qualquer investigação empírica sobre a subjectividade é "impertinente". Em *Dreaming* Malcolm defende, por razões aprioristas ligadas à sua filosofia da linguagem, que os sonhos não são experiências e que as investigações empíricas do sono REM (um estádio do sono assinalado por movimentos oculares rápidos[493]) não dizem respeito aos sonhos. A intenção de Dennett é mostrar o absurdo destas teses ao mesmo tempo preservando aquilo que considera acertado no verificacionismo de Malcolm.

Dennett crê que mesmo que o em-si da experiência seja por natureza um para-mim, isso não nos obriga a afirmar que o conhecimento dos fenómenos mentais só pode ser apriorista. Mas se considerações objectivas não são pertinentes para decidir acerca da presença de experiência e se ao mesmo tempo muitas vezes as convicções do sujeito não são suficientes para determinar a existência ou natureza da experiência (o que é o caso do sonho, se não se estiver disposto a decretar, por critérios linguísticos aprioristas, que os sonhos decididamente não são experiências), o perigo que espreita é evidentemente o de uma situação em que o teórico da consciência se vê obrigado a admitir fenómenos de subjectividade sem sujeito, *qualia* não apercebidos. Para Dennett, a única alternativa pertinente é o verificacionismo em primeira pessoa.

O *verificacionismo em primeira pessoa* é a ideia segundo a qual (i) é impossível por princípio a consciência de um estímulo na ausência de crença nessa consciência e (ii) é impossível por princípio ao sujeito decidir, relativamente à sua consciência de x, entre aparência e realidade. A pertinência do verificacionismo em primeira pessoa quando se trata de experiência interior deve-se a duas razões: (i) o em-si da experiência, a existência de experiência, só pode ser acedido como um para-mim; (ii) só a própria entidade que se sente ser pode saber como é ser, e pelo menos no caso humano, exprimir como é para si ser, daí a sua autoridade.

[492] Cf. também MALCOLM 1957.

[493] O sono REM é um estádio do sono em que existe activação cerebral das áreas responsáveis pelo processamento visual e dos sistemas motores responsáveis por movimentos oculares rápidos. Supostamente, durante o sono REM existem sensações e percepções vívidas, internamente geradas, pensamentos ilógicos e bizarros e movimentos comandados, embora inibidos. O sono REM alterna com a vigília e o sono não-REM de acordo com uma sequência recorrente. Cf. HOBSON 1999a e HOBSON 1999b.

Há aqui no entanto um problema grave: para que alguém expresse a sua experiência interior deve haver inscrição e recuperação de informação, i.e. memória. Ora, seja o que fôr recordar, não é verdadeiro por definição que o recordado tem que ter acontecido tal como é recordado, i.e. a natureza da memória não é tal que as memórias devam ser verídicas. O problema é que o sujeito ele próprio não tem qualquer possibilidade de distinguir, de entre as suas memórias, as memórias verídicas e as não verídicas, entre o que lhe parece ter sido e o que foi de facto[494]. A vulnerabilidade dos sonhos a este problema reside no facto de eles apenas serem acedidos enquanto memórias de experiências. Como diria Malcolm, o uso da frase «Estou profundamente adormecido e a sonhar» como asserção provocaria um sentimento de estranheza e seria de algum modo auto-contraditório[495]. Se os sonhos apenas são acedidos por recordação subsequente impõe-se saber se o facto de alguém recordar que sonhou basta ou não basta para concluir que sonhou (que experienciou, portanto). Malcolm defende que não, devido ao seu verificacionismo, e é levado a concluir em *Dreaming* (cujo objectivo era a análise conceptual wittgensteiniana do conceito "sonhar") que os sonhos não são experiências. Concluiu ainda que as investigações científicas dos sonhos – que visam estabelecer a relação do sono REM com o sonho e confirmar assim que os sonhos são "experiências que ocorrem durante o sono" – incorrem em violações e pecados conceptuais e em última análise não dizem sequer respeito aos sonhos

As investigações sobre o sonho e o sonho cuja clareza conceptual Malcolm atacou continuaram e revelam que todas as pessoas sonham, todas as noites[496]. Apesar disso, algumas pessoas afirmam convictamente não sonhar, outras sabem que sonharam mas não se

[494] Perante este ponto fixo da argumentação de Dennett, que aparece já em BS – a impossibilidade de distinguir a partir de dentro entre o que parece ser (*seem*) e o que é – D. Chalmers (CHALMERS 1996: 190) afirma que Dennett confunde dois sentidos de "*seem*": o sentido psicológico, cognitivo de *seem* (parecer-me, estar disposto a julgar de uma determinada maneira) e o sentido fenomenal de *seem* (aparecer). Note-se a propósito da possibilidade ou impossibilidade de distinção "a partir de dentro" entre memórias verídicas e não verídicas que a questão dá lugar a interessantes prolongamentos no contexto do problema da Identidade Pessoal, com a noção de quase-memórias. As quase-memórias poderiam ser distinguidas das memórias pelo facto de o sujeito ter o sentimento (ou o saber?) de que elas não lhes pertencem, embora estejam lá como suas (cf. por exemplo PARFIT 1984).

[495] Cf. MALCOLM 1956:18: «Suppose that I am in bed and you come and shake me and ask "Are you asleep?" and that I reply "I am sound asleep" It would be amusing if you took me as claiming that I am sound asleep and then concluded from this that I am sound asleep».

[496] HOBSON 1999a e HOBSON 1999b.

recordam dos detalhes, e todas elas, mais ou menos rapidamente, perdem as recordações dos detalhes. O relato parece mesmo interferir com a recordação e acelerar a perda. A situação de perda é acompanhada frequentemente de uma impressão de que as memórias, fossem quais fossem, estavam "lá", mas deixaram de estar (a situação não é até aqui muito diferente das memórias de experiência em geral). Como é fácil imaginar, o que é problemático é a possibilidade de alguém, a partir de critérios externos como electroencefalogramas e observação de movimentos oculares, poder garantir a outra pessoa que ela sonhou (experienciou, portanto), quando esta, uma vez interrogada, garante que tal não aconteceu (ou numa hipótese futurista, a possibilidade de garantir a alguém que acorda e relata um sonho A que esse sonho não foi sonhado mas inserido e que o sonho verdadeiramente sonhado e experienciado, um sonho B, não está acessível[497]).

Impedir a contestação da autoridade subjectiva, mostrar que a ideia de um conhecimento em terceira pessoa, que entra pelo sujeito dentro para negar as suas convicções, é totalmente impertinente, é o que está no horizonte de Malcolm quando declara que as investigações do sono REM não dizem respeito aos sonhos. Malcolm não afirma, note-se, que as pessoas não sonham (o "critério" dos sonhos é a existência de relatos de sonhos) mas que os sonhos não são experiências. A origem do veredicto de Malcolm é uma afirmação de Wittgenstein nas *Investigações*[498]: «A questão de saber se a memória do sonhador o engana quando ele relata um sonho depois de acordar não pode surgir a não ser que introduzamos um critério completamente novo para o relato concordar com o sonho, um critério que nos dá um conceito de verdade por oposição a sinceridade (*truthfulness*)»[499]. Com esta afirmação Wittgenstein sublinha que não dispomos de um espaço interior de avaliação que nos dê a capacidade de distinguir entre ter realmente sonhado e recordar (enganosamente) que sonhámos. A partir daqui Malcolm defende que não faz sentido falar de uma verdade acerca da experiência do sonhador durante o sonho, obtida por verificações exteriores, que

[497] O filósofo inglês Derek Parfit explora noutro contexto (o da Identidade Pessoal) esta possibilidade das memórias inseridas (Cf. PARFIT 1984, Part III). Poder-se-ia ainda evocar casos de ficção, como os casos de "dissociação ou deslocamento de experiência própria para um outro mundo virtual que não aquele onde o corpo próprio vive", como os explorados em filmes como *Matrix* (Irmãos Wachowsky) e *Existenz* (D. Cronenberg).

[498] WITTGENSTEIN 1953 (Parte II, 11, 213).

[499] DENNETT 1978j: 142.

contradiga o testemunho do sujeito: apenas as expressões que as pessoas produzem dos seus estados privados são critério da existência e natureza destes estados.

Para salvar a autoridade introspectiva Malcolm viu-se no entanto obrigado a negar que os sonhos fossem experiências. O argumento é o seguinte: "experiências" são coisas tais que pode parecer-nos enganosamente que tivemos uma experiência sem a termos tido, uma vez que faz sentido pensar em confirmar ou infirmar as asserções de experiência, independentemente da inclinação subjectiva à afirmação. Pelo contrário, o único critério de uso do conceito "sonhar" é o testemunho de uma pessoa acordada, cujas asserções de experiência interior não podem, posteriormente e a partir de fora, ser confirmadas ou infirmadas. Logo, os sonhos não são "experiências". O argumento de Malcolm depende da conjugação do verificacionismo com uma teoria da significação que tem como referência a linguagem comum e de acordo com a qual (i) existem critérios para conceitos (ii) os conceitos devem ter métodos de verificação ou não terão sentido (iii) existem casos paradigmáticos que cumprem esses critérios e nos quais se pode claramente decidir quanto à aplicação do conceito. É daqui que decorrem as conclusões quanto aos conceitos "sonhar" e "experiência".

No entanto tais conclusões apenas se justificam se: (i) a análise apriorista fôr um método defensável para a abordagem do significado (de termos mentais e da linguagem em geral); (ii) a competência semântica das pessoas (i.e. a forma como elas usam os conceitos) pressupuser de facto aquilo a que Putnam, analisando as teses de Malcolm chama uma "gramática de profundidade"[500]. Essa gramática de profundidade, não apercebida pelo falante comum, revelaria a "dependência lógica" de conceitos em relação a critérios (por exemplo a dependência lógica do conceito "sonhar" em relação ao testemunho de uma pessoa acordada) tendo como referência as fronteiras do sentido estabelecidas pelo uso comum. É devido a esta epistemologia que Malcolm defende que é uma extensão sem sentido do conceito "sonhar" aquela que repousa sobre um critério fisiológico (por exemplo os movimento oculares rápidos). O critério fisiológico envolve não apenas uma mudança conceptual radical como também o uso de um conceito sem métodos de verificação.

Malcolm salva com a sua conclusão o sentimento de autoridade introspectiva. A salvação dá-se no entanto, como já foi dito, *à custa de um conservadorismo epistemológico inaceitável para Dennett*, que considera absurdo defender, como faz Malcolm, que as investigações do sono REM nem sequer dizem respeito aos sonhos. O facto de o em-si da experiência ser um para-mim não obriga a defender que o conhecimento dos fenómenos mentais só pode ser apriorista. Se o nosso autoacesso não for um acesso epistémico a tipos naturais, parece desejável que os conceitos mentais tenham caracterizações técnicas, resultantes de investigação empírica. É claro que o problema deste movimento é o facto de ele colocar fora da pessoa a autoridade acerca da natureza da sua própria experiência.

Como é óbvio a grande diferença de pressupostos entre filósofos da linguagem como Wittgenstein e Malcolm e o próprio Dennett é o facto de este considerar a análise conceptual um mau método para a teoria da mente, que se traduz, no caso em causa, num impedimento arbitrário da consideração de dados empíricos. H. Putnam afirmara já que o principal erro da análise de Malcolm era não prever que os "critérios da significação", inclusive da significação de termos mentais, se modificam ao longo do tempo, através da investigação empírica e dos novos conhecimentos. É claro que se se permitir a entrada de dados empíricos no sentido dos termos mentais, o sentimento de autoridade do sujeito não poderá ser explicado como incorrigilidade quanto a conteúdos epistémicos. A incorrigibilidade terá que ser explicada de outra maneira, nomeadamente através da relação entre memória, apresentação e expressão e é precisamente esse o propósito da análise dos sonhos em *Are Dreams Experiences?*.

Dennett considera nos sonhos processos de apresentação, memória e composição e o ponto que pretende explorar é o seguinte: mesmo que se admita que acontece uma apresentação (entenda-se: consciente) das imagens e narrativas do sonho, correspondente aos eventos neuronais durante o sono REM, é a memória, no sentido de inscrição, armazenamento e recuperação de informação, que possibilita o relato do sonho. Ora a inscrição e o armazenamento podem por princípio ser inconscientes e são certamente independentes da (hipotética) apresentação. Assim, parece possível interferir na recordação sem se ter interferido na "apresentação". Interferindo no armazenamento podem por exemplo ser introduzidas memórias de sonhos não sonhados, que serão recordadas como sendo de sonhos sonhados. A recuperação para relato dá acesso a conteúdos mas não a garantia de que uma apresentação aconteceu

[500] PUTNAM [1962], *Dreaming and the Depth Grammar.*

e nem isso pode ser avaliado pelo sujeito (foi o que Malcolm percebeu quando defendeu que de "Parece-me que foi como alguma coisa sonhar" não é permitido concluir que "Foi como alguma coisa sonhar").

Esta condição (segundo a qual a recuperação para relato dá acesso a conteúdos mas não a garantia de que uma apresentação aconteceu) aplica-se à experiência em geral. Analisemos por exemplo uma recordação, reportada por uma pessoa A, das sua experiências relativas à entrada numa sala cheia de gente. Pergunta-se: Quantas pessoas estavam na sala? A pessoa B estava presente? No relato da pessoa A, como em qualquer relato de experiência, estarão presentes conclusões tiradas com base em informação armazenada. Mas o que é expresso nas conclusões não tem que ter sido apresentado, apenas tem que ter sido inscrita e estar acessível a informação da qual decorrem essas conclusões. Se por exemplo a pessoa A afirma que a pessoa B não estava na sala cheia de gente onde entrou, não se pode daí concluir que ela apercebeu realmente uma ausência, a ausência de B. Como diria Wittgenstein, a ausência de percepção não é percepção de uma ausência. A ausência de B não foi experienciada mas apenas "concluída" a partir da informação disponível.

Quanto ao terceiro processo envolvido nos sonhos, a composição das frequentemente elaboradas narrativas e imagens do sonho, embora se considere usualmente que ela acontece durante o sono, nos períodos do sono REM, poder-se-ia descobrir que não é assim. Pense-se nos casos em que alguém sonha algo que encaixa perfeitamente no que está a acontecer quando se acorda (dá-se um grande estrondo exterior, por exemplo, e o sonhador acorda de um sonho em que sonha que morreu numa explosão). Parece haver uma preparação narrativa do encaixe com a realidade exterior. Excluindo a existência de pré-cognição, esse encaixe poderia explicar-se, sugere Dennett, por hipóteses semelhantes a estas: talvez os sonhos sejam compostos e apresentados muito rapidamente no intervalo entre acordar e estar plenamente desperto, sendo o efeito conseguido com um sistema de atraso da percepção exterior. Ou talvez, nesse intervalo, os sonhos sejam compostos e armazenados em ordem invertida, sendo depois recordados na sequência própria. Ou talvez exista uma biblioteca de sonhos não sonhados no cérebro, com vários fins indexados, prontos a serem inscritos na memória de forma adequada à percepção exterior, para virem a ser relatados. O que importa é que para o sonhador, uma vez acordado e relatando o que lhe parece ser, será impossível por princípio perceber qualquer diferença entre memórias inseridas de sonhos e sonhos experienciados.

Dennett não pretende infirmar com um exemplo avulso e com hipóteses *ad hoc* a ideia segundo a qual os sonhos são experiências, apenas analisar as implicações que a confirmação de uma hipótese empírica semelhante a estas teria. Tal confirmação atestaria que o sentimento de autoridade subjectiva do sonhador não lhe dá qualquer autoridade para decidir entre hipóteses relevantes. Pelo contrário, outro tipo de dados (nomeadamente neurofisiológicos) seriam mais relevantes para a decisão. E note-se que pelo menos de acordo com a última hipótese sonhar não teria sido como nada, embora viesse a parecer ter sido como alguma coisa.

O objectivo do artigo *Are Dreams Experiences?* é defender que a questão "serão os sonhos experiências?" é uma questão teórica em aberto e não uma questão a dirigir ao sonhador. Não é possível decidir a resposta através do apelo à autoridade subjectiva. O caso dos sonhos mostra que, apesar do sentimento de autoridade subjectiva que acompanha as asserções de experiência interior, o conteúdo destas é vulnerável à infirmação empírica. Dennett estende a conclusão a toda a fenomenologia (por exemplo, como se verá à frente, às descrições de imagens mentais e às decisões voluntárias). O sentimento de autoridade subjectiva não assinala qualquer *conhecimento* privilegiado da *natureza* da experiência mental, não sendo sequer uma marca confiável do apercebimento num sujeito: o conteúdo do apercebimento é tudo aquilo que entra na memória e os indícios deste apercebimento não estão necessariamente ou exclusivamente ligados à convicção subjectiva. A haver no modelo de consciência de BS uma "relação conceptual" importante como queria Malcolm, ela é uma relação entre experiência e memória e não entre experiência e convicção subjectiva. O caso dos sonhos mostra que pode existir memória de algo apercebido sem sequer ter havido apresentação: basta ter havido inscrição na memória de um conteúdo que (i) *guie o comportamento actual*. Este é um primeiro indício do apercebimento de x (um conteúdo qualquer) por um sistema. Por outro lado, a informação que guia o comportamento actual poderá ser objecto de acesso público, i.e. estar disponível para ser expressa. Um segundo indício do apercebimento será então (ii) *poder ser expresso*. Mas quando alguém dá por si querendo emitir uma asserção de experiência interior, isso acontece por meio da memó-

ria (CM) e do "querer-dizer" (CC → CRP), sem qualquer espaço de observação interno. Retomando as conclusões de *Toward a Cognitive Theory of Consciousness* Dennett defende que o sentimento de incorrigibilidade que acompanha estas asserções é função dos acessos sub-pessoais inconscientes entre a memória e o sistemas de produção da fala e que os relatos de experiência interior são assim os próprios dados e não relatos de dados. Precisamente os sonhos levam Dennett a alargar este segundo indício de apercebimento a um terceiro: (iii) *a capacidade retrospectiva de relatar*. O quarto indício serão as (iv) *saliências funcionais que a investigação empírica descobrir* (no nosso exemplo a caracterização do sono REM). Estes indícios podem inclusivamente chocar entre si.

Assim, de acordo com a teoria verificacionista da consciência, a consciência não esgota o apercebimento. No entanto, os juízos esgotam a consciência imediata a que outros chamariam "consciência fenomenal". Não pode haver consciência de x sem crença do sujeito na consciência de x, e o sujeito será incorrigível nessa crença. Mas ele é incorrigível porque não tem "espaço de manobra" e não devido a qualquer autoridade epistémica. *Ser incorrigível não é ser infalível.* Convem sublinhar, recordando o início deste capítulo, que Dennett pensa que apenas uma posição verificacionista como esta "impede" *zombies*, espectros invertidos ou epifenomenalismo, ideias que considera absurdas (daí que Dennett afirme que o "superficialismo" verificacionista não é uma posição superficial na teoria da consciência). "Consciência" é apercebimento-de-que-(factos ocorrem) e algo que tem resultados (nomeadamente os conteúdos serem expressos). Poder-se-ia objectar dizendo que o que está a acontecer aqui é que o verificacionismo e a correlativa definição de consciência como um fenómeno de segundo grau estão a "ocupar" o lugar da consciência fenomenal. Dennett defende que um fenómeno como "consciência fenomenal, não epistémica" não existe, mas poder-se-ia contrapor que Dennett apenas pensa que tal fenómeno não existe devido à sua "intelectualização" do apercebimento (traduzida por exemplo numa confusão entre "aperceber maçãs" e "aperceber-se de que existem maçãs"). É por esta razão que por exemplo Dretske classifica Dennett como cognitivista, i.e. alguém para quem a própria percepção é já cognição, e ver ou ouvir ou sentir alguma coisa são espécies de juízos.

Uma particularidade da teoria de Dennett é defender que no limite, em muito do apercebimento que acontece em sistemas co-

mo os humanos, a pura apresentação desaparece e no seu lugar fica a memória, concebida como registo físico. Em virtude desta substituição de apresentação por memória ter-se-á que somos muitos mais "apagados" [501] do que aquilo que nos parece ser o caso (recorde-se comentários sarcásticos de J. Searle e T. Nagel, segundo os quais, respectivamente, se estaria, com a teoria da consciência de Dennett, a assistir a uma representação de Hamlet sem o princípe da Dinamarca, ou perante um livro sobre Picasso que não menciona ou mostra qualquer quadro). Convém no entanto ter consciência da outra leitura dessas críticas: para o próprio Dennett isso não acontece por deficiência da teoria, mas porque é assim que somos. Repita-se que neste quadro o apercebimento tem um âmbito muito maior do que a consciência (considerada como reportabilidade dos conteúdos para um eu), um âmbito a ser determinado pela investigação empírica. A autoridade subjectiva assinala apenas um certo tipo de acesso, não a presença de um género natural que seria a "consciência". Aprendendo com o absurdo da posição de Malcolm, que pretendia decidir a partir do uso linguístico presente se os sonhos são experiências, a conclusão de Dennett é que a noção de critérios de experiência é epistemológicamente ilegítima, derivada de uma insustentável posição apriorista perante aquilo que são questões empíricas abertas.

Juntando as conclusões de *Toward a Cognitive Theory of Consciousness* e de *Are Dreams Experiences?* obtemos o seguinte: um modelo como o proposto em BS conduz-nos a concluir que não dispomos de critérios aprioristas que garantam a existência de experiência num sistema. Assim, não podemos desde já distinguir claramente as coisas para as quais é como alguma coisa ser das outras (aparentemente não podemos fazê-lo nem sequer sendo...). Não podemos fazê-lo porque por um lado é o próprio mundo que é possivelmente vago, sem divisão abrupta entre consciência e apercebimento inconsciente, por outro porque se trata de questão empíricas abertas. As questões que definiriam o reino subjectivo defendido por Nagel são questões «acerca das quais a opinião subjectiva subsequente do sujeito não tem autoridade»[502]. Para

[501] *Cf.* PINTO 1999: 133 *para o uso do termo "apagado". O termo "apagado" traduz o termo zombie* e nomeia uma réplica física de um sistema consciente na qual a experiência fenomenal está ausente. Ser um apagado ou *zombie* neste sentido é evidentemente um caso de tudo ou nada e não de mais ou menos, como se está aqui a supor. A pertinência de qualquer destes termos é a captura da situação em que tudo decorreria no escuro, enquanto Dennett pensa que quando se trata de experiência nunca se está perante situações assim tão claras (ou tão escuras).

[502] DENNETT 1978j: 143.

Dennett esta não é uma objecção e sim uma conclusão, que deve ser estendida da análise dos sonhos à experiência em geral. A diferença entre ser e não ser como alguma coisa sonhar, e experienciar, tornou-se uma questão teórica, científica. Note-se no entanto que se é uma questão teórica em aberto saber se os sonhos caem dentro ou fora da fronteira da experiência e se o propósito de *Are Dreams Experiences?* é alargar certas conclusões acerca dos sonhos à experiência em geral, ficar-se-á a certo ponto perante a afirmação curiosa segundo a qual é uma questão em aberto saber se a experiência *é* experiência...

3.2.2 *Imagens mentais.*

A questão das imagens mentais constitui hoje, como se pode verificar em qualquer manual, uma área clássica da ciência cognitiva, tendo sido objecto de acesa discussão na filosofia e na psicologia. O debate entre dois psicólogos cognitivos, Stephen Kosslyn e Zenon Pylyshyn[503], é exemplar da concretização trazida pela investigação empírica para a questão milenar das relações entre imagens e pensamento. Ele revela no entanto, ao mesmo tempo, o carácter problemático de que a questão continua a revestir-se. Dennett abordara já a questão das imagens mentais em C&C, considerando então que qualquer apelo a imagens sub-pessoais era homuncular e se comprometia com uma testemunha impossível, e que as indeterminações e não-menções nas descrições fenomenológicas de experiências de imagens revelavam o carácter não pictórico mas descritivo das "representações" mentais subjacentes. Em BS Dennett reconhece que as suas afirmações de então, de uma "iconofobia metafísica convicta" tinham sido em parte contraditas pelo trabalho empírico[504]. Crê ainda assim que há uma distinção a fazer entre

[503] Para alguns relatos deste debate e das suas repercussões, cf. GARDNER 1985: 128 e 323-339, FODOR 1975: 187 e BLOCK 1981. Como principais referências das posições dos dois psicólogos tomar-se-á PYLYSHYN 1984, Capítulos 7 e 8, KOSSLYN 1994 (que pretende ser uma "resolução do debate acerca de imagens mentais") e KOSSLYN 1995. Z.Pylyshyn é um psicólogo e cientista de computadores canadiano cujo livro *Computation and Cognition* (PYLYSHYN 1984) constitui um manifesto dos princípios filosóficos do cognitivismo clássico.

[504] Dennett refere-se (cf. DENNETT 1978: 333 n1) a trabalhos como R.SHEPARD & J.METZLER 1971, R. SHEPARD & L.COOPER 1973, A. PAIVIO 1971 e KOSSLYN 1975 sobre imagens mentais. Neste tipo de estudos pede-se aos sujeitos que formem imagens mentais, que respondam a questões sobre elas e que efectuem transformações (como rotações, deslocamentos), etc. Num dos mais citados exemplos –estudado em SHEPARD & METZLER 1971— os sujeitos devem efectuar rotações "imaginadas" a partir de imagens reais de objectos tridimensionais de modo a decidirem

iconofobia e iconofilia *metafísicas e* iconofilia e iconofobia *científicas.* Nomeadamente, apenas a iconofobia científica está aberta a refutação. Dennett admite portanto em BS ter misturado em C&C uma iconofobia metafísica acertada, com uma iconofobia cientifica, que subscreve moderadamente. A análise levada a cabo em *Two Approaches to Mental Images* pretende ser neutra na batalha entre iconófilos e iconófobos "científicos" e esclarecer o âmbito de legitimidade da discussão acerca de imagens mentais.

Os já mencionados psicólogos S. Kosslyn e Z.Pylyshyn são respectivamente exemplos de iconofilia (à qual N. Block chama também pictorialismo[505]) e iconofobia. O seu debate servirá aqui para contextualizar a posição de Dennett[506]. Assumindo que existem representações mentais (suposição já de si bastante polémica, como se viu no capítulo anterior, mas que quer iconófilos quer iconófobos fazem), o debate entre iconófilos e iconófobos gira em torno da seguinte questão: existem ou não diversos *tipos* de representações mentais? Por outras palavras, existirão representações mentais de um tipo específico, as imagens, que supõem um *medium* analógico específico[507]? Ou, pelo contrário, as imagens mentais fenomeno-

quais figuras são representações do mesmo objecto, visto a partir de diferentes pontos de vista. Um dos resultados que mais estimulou os teóricos foi a descoberta de que os tempos de reacção que precedem as decisões acerca da identidade das formas aumentam quando a diferença na orientação das referidas formas aumenta. Os sujeitos demoram mais tempo a decidir no caso em que para responder é necessário efectuar uma rotação de 90° do que no caso em que apenas é necessária uma rotação de 20°, por exemplo (cf. o comentário em KOSSLYN 1995: 276). Os resultados, que se revelaram sistemáticos, foram interpretados como revelando o facto de os sujeitos efectuarem as comparações rodando mentalmente as imagens (a uma "velocidade" constante), permitindo falar do que acontece nesses processos no "espaço mental" de uma forma análoga à usada para falar de processos no espaço físico, por exemplo falando de princípios de conservação, de simetrias e de proximidade entre objectos. Os fenómenos permitiriam formular uma «lei psicofísica simples mas robusta: o tempo necessário para ajuizar da identidade (ou não identidade) é uma função monótona da distância física entre as formas» (GARDNER 1985: 325). Tudo isto é estranho pois as imagens mentais não são objectos físicos reais submetidos às leis físicas mas sim *representações* de imagens. Se as formas mostradas aos sujeitos fossem objectos físicos seriam objectos rígidos, devendo passar por fases intermédias à medida que era feita a sua rotação. Mas à partida nada parece obrigar imagens mentais a comportarem-se de maneira idêntica.

[505] BLOCK 1981.

[506] Vários dos escritos de Kosslyn e Pylyshyn convocados para o resumo do debate e para essa contextualização têm data posterior ao artigo *Two Approaches of Mental Images.* Na medida em que o objectivo de Dennett no artigo é metodológico ou meta-teórico não se considerou o facto especialmente problemático (embora se admita, evocando Pylyshyn, que a decisão é contestável, já que «in cognitive science the gap between metatheory and practice is extremely narrow» (PYLYSHYN 1984: xix)).

[507] A definição de analógico é objecto de discussão e as motivações da definição podem ser, segundo Pylyshyn, dúbias quando se trata de distinguir processos analógicos e digitais em teoria cognitiva. A definição mais comum afirmaria que um processo analógico de representação envolve formas

logicamente descritas pelos sujeitos são apenas "epifenómenos" (um outro termo já de si polémico) de representações de um outro tipo (único)? Iconófobos como Z. Pylyshyn defendem a existência de um tipo único de representações mentais, de "formato proposicional" ou simbólico.

É importante notar que a suposta existência de imagens mentais constitui uma objecção ao "cognitivismo simbólico clássico[508]. Numa caracterização rápida pode-se considerar que o cognitivismo simbólico clássico assenta no postulado de estados mentais simbólicos, com poder causal, o qual é explicado pela existência de relações funcionais / computacionais entre representações mentais e organismo. O nível simbólico dos estados e processos mentais, o nível das relações computacionais, é um nível autónomo em relação ao nível físico que o implementa, e constitui uma arquitectura funcional[509] "cognitivamente impenetrável" (i.e. isolada em relação a "acessos a partir de cima", não acessível ou modificável pelas crenças do sujeito que seriam o nível semântico do sistema). A ideia de níveis separados (físico, simbólico e semântico) é essencial a este paradigma de explicação cognitiva[510]. Ora, o nível simbólico é suposto ser (além de cognitivamente impenenetrável) precisamente *simbólico* (proposicional, digital, computacional) e não imagético.

A discussão acerca de imagens mentais trava-se aparentemente num campo empírico, mas, de facto, um meta-problema espreita, que consiste em decidir se o problema debatido é conceptual ou empírico (note-se que o acima referido "paradigma explicativo" não é propriamente uma descrição neutra mas uma proposta meta-teórica). Dividido o campo entre iconófilos ou pictorialistas e iconófobos ou anti-pictorialistas, comecemos pelo princípio básico do

de representação contínuas e não discretas. Pylyshyn cita por exemplo a definição que Shepard dá no contexto do estudo das imagens mentais de um processo analógico: para Shepard é "analógico" um processo que passa pelos mesmos estádios intermédios por que o processo representado teria passado (PYLYSHYN 1984: 203).

[508] Cf. MACHUCO ROSA, 2002, 1. As ciências cognitivas clássicas, para uma listagem das teses centrais desse cognitivismo clássico, que consiste afinal na contraparte, na ciência cognitiva, do funcionalismo definido filosoficamente por H. Putnam (tal como foi exposto no Capítulo 1) e desenvolvido por Fodor (tal como foi exposto no Capítulo 2), Z. Pylyshyn, D.Marr, etc.

[509] A expressão "arquitectura funcional" é teorizada por Z. Pylyshyn (cf. PYLYSHYN 1984). A questão é importante neste debate.

[510] Por exemplo segundo D. Marr (MARR 1980) seriam distintos um nível 3 (físico ou biológico), um nível 2 (simbólico ou funcional) e um nível 1(semântico). Existe obviamente uma correspondência entre estas ideias centrais no "cognitivismo clássico" e as noções de EF, ED e EI tal como são utilizadas na TSI.

anti-pictorialismo, defendido por Z. Pylyshyn. De acordo com Pylyshyn, existe um único tipo de representações mentais. Isto não quer dizer que as pessoas não têm imagens (fenomenologicamente) mas sim que as "imagens" fenomenológicas são produtos laterais e epifenomenais, sem qualquer função cognitiva, do processamento de representações simbólicas, proposicionais ou sentenciais[511]. Dizer que as imagens são epifenomenais é dizer (na comparação de Kosslyn) que (os traços descritivos das imagens) «não têm nada a ver com a representação utilizada para levar a cabo a tarefa, assim como as luzes acendendo e apagando no exterior de um computador não têm nada a ver com o processamento interno (se forem removidas, este continua idêntico)»[512].

A ideia de Pylyshyn é que processos fenomenológicos imagéticos dependem de conhecimento tácito. Se não existem representações por semelhança mas apenas um *medium* de representações, simbólico e não analógico, as "imagens" serão produto de inferências subdoxásticas baseadas nessas representações proposicionais e não entidades num *medium* analógico, obedecendo a leis "físicas". Em suma, para Pylyshyn, imagens mentais nada têm de específico relativamente a outras representações. O fulcro da tese de "pictorialistas" como S. Kosslyn é, em contrapartida, a existência de representações mentais de um tipo específico, bem como a existência de um espaço para a geração e manipulação das configurações imagéticas. Para um "pictorialista" é um erro acreditar que as representações proposicionais são a língua franca de todos os sistemas cognitivos. A ideia segundo a qual o processamento de informação seria por natureza cego ao conteúdo (*content-blind*) deriva de uma transposição directa para a psicologia cognitiva humana daquilo que se passa nos computadores. No entanto, os computadores são deliberadamente construídos para serem cegos ao conteúdo, não permitindo uma generalização imediata para o caso humano.

[511] Cf. PYLYSHYN 1984: 225, o Capítulo "Mental Imagery and Functional Architecture". Note-se que Pylyshyn (PYLYSHYN 1984: 193) explica o que entende por "natureza sentencial de representações", uma expressão bem menos comprometida do que "natureza proposicional de representações", uma vez que se fala de fenómenos concretos em sistemas físicos e não de entidades abstractas como proposições. O que justifica o termo é a coerência semântica da actividade cognitiva enquanto raciocínio (*reasoning*) que torna pertinente falar de "*truth-bearing expressions*". Mas as representações são concretas, *sentence-analogues*, expressões simbólicas num sistema simbólico interno fisicamente instanciado (a linguagem de pensamento ou mentalês de Fodor, evidentemente). Embora nem tudo seja inferência nos processos que medeiam input e output num sistema cognitivo, este é o caso exemplar, não marcado, diz Pylyshyn.

[512] KOSSLYN 1995: 383.

Nos humanos poderia ter evoluído uma "capacidade imagética" com uma utilidade "ecológica" específica, servindo nomeadamente para lidar com as transformações no ambiente, a qual estaria agora instalada, permitindo os fenómenos que aparecem nos estudos empíricos acerca de imagens.

S. Kosslyn pretende mostrar com os seus estudos experimentais e a sua elaboração teórica das últimas décadas que existe um *medium* imagético, analógico, específico para o tratamento de informação visual e espacial, que fixa as propriedades das representações imagéticas. Pela sua própria natureza, representações pictóricas "incorporam" espaço e assim o modelo de Kosslyn prevê um espaço visual para a geração e manipulação de imagens e para a extracção de informação. Trata-se de processos especificamente pictóricos. Seriam precisamente constrangimentos ligados ao *medium* analógico que explicariam a resistência das imagens mentais ao controlo voluntário dos sujeitos e o facto de elas aparentemente obedecerem a leis próprias. Por exemplo, segundo Kosslyn «nas representações pictóricas (*depictions*) ao contrário das representações proposicionais, a forma (*shape*) do espaço vazio é representada tão claramente como a forma do espaço preenchido e não existe nenhuma representação explícita das relações (como ATRÁS DE)»[513].

Os estudos experimentais permitiram a formulação de princípios como os seguintes: "as imagens mentais são transformadas passo a passo de modo a passarem por estádios intermédios de transformação", "as imagens preservam as distâncias, o que se repercute nos tempos necessários para operar sobre elas transformações tais como rotações", "são necessários tempos diferentes para "atravessar" e navegar em mapas mentais, logo o tempo é função da distância atravessada", etc. Todos estes princípios se encontrariam ligados às características do *medium* analógico.

As teses pictorialistas são sustentadas empiricamente pelos numerosíssimos estudos acerca de geração e inspecção mental de imagens feitos nas últimas décadas. Numa das experiências de Kosslyn[514] mostra-se aos sujeitos um mapa com sete localizações ficcionais: uma rocha, uma árvore, uma praia, um pedaço de relva, um poço, uma cabana e um lago.

[513] KOSSLYN 1995: 286.

[514] Cf. KOSSLYN 1995: 287, onde se relata as experiências levadas a cabo por S. Kosslyn, T.M. Ball & B. J. Reiser e publicadas num artigo significativamente intitulado *Visual Images preserve metric spatial information: evidence from studies of image scanning*, publicado no *Journal of Experimental Psychology* em 1978.

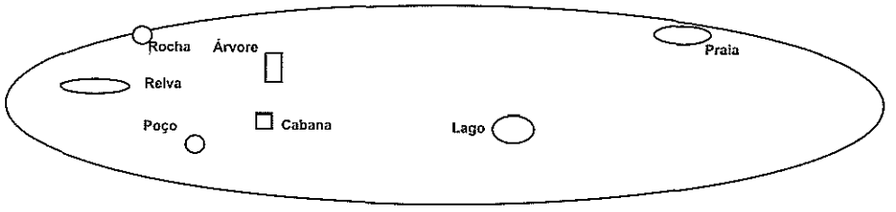


FIG. 2 Um "mapa de Kosslyn"

Pede-se-lhes em seguida que reproduzam mentalmente o mapa e que respondam a certas questões. Nomeadamente, pede-se-lhes que "focalizem" um primeiro lugar e em seguida "procurem" um segundo lugar, imaginando um ponto preto movendo-se de um lugar para outro lugar, e que pressionem um botão quando "chegarem" ao segundo lugar. De novo, como no exemplo atrás referido, o tempo prova ser uma função linear da distância entre os dois lugares. Em contraste, foram feitas questões aos sujeitos acerca da presença no mapa de uma determinada localização. Neste caso nenhum efeito distância/tempo foi encontrado, sendo perfeitamente possível conceber que os sujeitos consultavam representações simbólicas. Noutras experiências verifica-se que se demora mais tempo a inspeccionar um objecto grande do que um objecto pequeno, mais tempo a aperceber os pequenos detalhes do que a aperceber os grandes, que para aperceber pequenos detalhes é aparentemente necessário fazer um *zoom* sobre a imagem, que é mais difícil saber algo acerca de objectos ocultos por outros objectos, etc[515].

Como é óbvio, a dificuldade de interpretação dos resultados deste tipo de experiências reside em distinguir o caso em que os sujeitos, para responderem às questões colocadas, imaginam e consultam imagens do caso em que os sujeitos simplesmente raciocinam. A informação tanto pode estar armazenada sob forma pictórica como sob a forma de outros tipos de estruturas de dados e Kosslyn admite-o. De acordo com a explicação apresentada por Kosslyn, os resultados apresentados nas respostas são possibilita-

[515] Para os relatos e interpretações das experiências conducentes a estas conclusões utilizou-se KOSSLYN 1994 e KOSSLYN 1995.

dos pelas propriedades intrínsecas do *medium* de representação em que os sujeitos imaginam. Recorde-se que a discussão entre iconófilos e iconófobos é acerca do que ocorre ao nível sub-pessoal quando os sujeitos afirmam ter imagens, não acerca do que ocorre a nível pessoal.

Na terminologia de Pylyshyn, a ideia de Kosslyn ao falar de propriedades intrínsecas do *medium* de representação é que a existência de imagens deve ser reportada às propriedades da arquitectura funcional. Mas para Pylyshyn esta tese não passa de um equívoco resultante da tendência para considerar os processos cognitivos em termos dos objectos representados (ao nível semântico) e não em função da estrutura ou sintaxe da representação (a arquitectura funcional). Kosslyn defende, no entanto, firmemente a existência de uma forma de representação quase-pictórica específica, a que chama "*imagery*", e que seria uma característica básica da cognição humana. É preciso sublinhar que Kosslyn não faz afirmações simplistas do género "as pessoas têm imagens na cabeça e os resultados experimentais revelam-no". Afirma antes que as imagens mentais são descrições funcionais[516], caracterizadas por serem essencialmente pictóricas. Kosslyn considera que as "imagens" têm dois componentes e um deles é uma representação de superfície. Esta representação de superfície é uma *"entidade quase-pictórica"* acompanhada pela experiência fenomenológica de ter uma imagem. A entidade quase pictórica corresponde a uma activação temporária na memória. Este componente das imagens é comparado com as mostrações produzidas num écran por um programa de computador operando a partir de dados armazenados na memória: os dados podem ser armazenados na memória num noutro formato mais abstracto e não imagético. Assim, a geração de imagens passa pela interacção entre memória descritiva (de longo prazo) e memória pictórica (de activação transitória). O que interessa Kosslyn é defender que a imagem pictórica existe de facto. Kosslyn e os seus colaboradores desenvolveram modelos computacionais para testar a teoria, que envolvem dois tipos de estruturas de dados. O primeiro tipo de estrutura de dados é uma matriz de superfície (a estrutura de imagens de dados corresponde a uma configuração de

[516] As imagens são representações mentais e «a mental representation is a description at the functional level of analysis of how the brain stores information» (KOSSLYN 1995: 280). A questão é portanto acerca de formatos e da diferença entre "*propositional representations*" com as suas especificidades sintácticas e semânticas e "*depictive representations*" com as suas.

pontos numa matriz). Existe um *medium* espacial para suportar as representações, um *buffer* visual. Uma imagem gerada é um preenchimento de células na matriz. A matriz corresponde à memória visual de curto prazo. O segundo tipo de estrutura de dados são ficheiros de memória de longo prazo, que incluem nomeadamente factos sobre os objectos imaginados, representados em formato proposicional. Estas estruturas de dados são utilizadas em três processos imagéticos: geração de imagens, classificação de imagens e transfomação de imagens. Para Kosslyn, o facto de a simulação funcionar rebate por si só a acusação de homuncularismo.

O trabalho de Kosslyn foi criticado em diversas frentes e acusado de ser pouco preciso, nomeadamente devido ao uso da expressão "quase-pictórico" para falar das imagens, o que é certamente uma boa forma de evitar os problemas inerentes ao termo "pictórico". Mas, basicamente, Kosslyn defende (e com apoio na neurociência cognitiva, que teria, na sua opinião vindo dar um último impulso ao debate acerca de imagens mentais[517]) que existem incontestavelmente apresentações pictóricas de imagens imaginadas, assim como apresentações pictóricas de imagens "reais-actuais", embora o armazenamento da informação correspondente, que permite a geração das imagens, não tenha formato pictórico.

O maior crítico de Kosslyn é precisamente Z. Pylyshyn, que nega absolutamente que as imagens mentais sejam uma forma específica de representação mental que permite aos sujeitos raciocinar e resolver problemas de forma específica. Para Pylyshyn, a cognição é computação e supõe necessariamente um *medium* simbólico. Apenas representações simbólicas garantem produtividade, composicionalidade e racionalidade, e ao contrário do que muitos críticos afirmam, essa conclusão não é devida a uma metáfora tecnológica datada (a metáfora do computador) mas ao entendimento lógico das características da linguagem.

O critério que Pylyshyn propõe para decidir acerca da existência ou não do *medium* analógico de Kosslyn é a impenetrabilidade cognitiva. Uma função é cognitivamente impenetrável se não pode ser alterada via as mudanças induzidas nos fins e crenças de sujeitos. Por definição, de acordo com a ideia cognitivista de níveis de um sistema, a arquitectura funcional e a base física são cognitivamente impe-

[517] Ao pôr em relevo o contraste entre áreas de processamento visual, topograficamente organizadas e também activadas na imaginação, e a ausência de organização topográfica das áreas de armazenamento.

netráveis. A arquitectura funcional é, recorde-se, o mecanismo de processamento de informação no sistema cognitivo, o qual não requer qualquer caracterização "semântica" (se este *medium* é representacional é-o no sentido sintáctico). Aquilo que requer caracterização semântica é pelo contrário cognitivamente penetrável. O argumento básico de Pylyshyn é o seguinte: se o *medium* analógico proposto por Kosslyn fosse parte da arquitectura funcional, os processos imagéticos não deveriam ser cognitivamente penetráveis (assim como os processos cognitivos de computação e inferência não são cognitivamente penetráveis). Trata-se então de interpretar o que é o caso com as imagens mentais e as leis nelas seguidas. Ora, segundo Pylyshyn, aquilo que parece um processo natural que se desenrola de acordo com leis imputáveis ao *medium* (os princípios acima nomeados) revela-se afinal cognitivamente penetrável. Aquilo que acontece com as imagens é manipulável por meio de instrucções verbais. Resta então saber exactamente que tarefa estão os sujeitos a levar a cabo quando imaginam numa das situações experimentais[518]. Para Pylyshyn isto mostra que é conhecimento tácito (e não propriedades de um *medium* analógico) o que permite explicar a situação. Se a formação e a transformação de imagens são processos cognitivamente penetráveis, as características desses processos não são explicáveis fazendo apelo às propriedades do *medium*. Pylyshyn evoca por exemplo experiências que envolvem a imaginação da resolução de problemas físicos comuns (problemas como "de que cor serão as transparências sobrepostas?" ou "O que acontecerá ao nível da água se...?") mostrando que os resultados dependem dos conhecimentos de física dos sujeitos, não de leis do *medium*. Os sujeitos que obtiveram resultados totalmente acertados nas tarefas de imaginação eram sujeitos capazes de declarar explicitamente os princípios teóricos que justificavam esses resultados, como entrevistas posteriores vieram a mostrar. Pylyshyn comenta ainda a situação do mapa de Kosslyn, considerando que também ela explora conhecimento dos sujeitos. Os resultados dependem de etiquetação e interpretação e não de propriedades intrínsecas da mostração de superfície.

A discussão está aberta ainda hoje. Neste momento, e considerando que a questão será retomada mais à frente no contexto do

[518] PYLYSHYN 1984: 233 e 235. Os sujeitos podem estar a (1) resolver um problema usando uma forma de representação ou *medium* prescrito ou a (2) tentar recriar tão precisamente quanto possível as sequência de eventos perceptivos que ocorreriam se o evento estivesse realmente a ser observado. Pylyshyn crê que as instruções conduzem ao caso 2.

modelo de consciência de CE, interessa apenas especificar o que pretende Dennett afirmar acerca de imagens mentais em *Two Approaches to Mental Images*. Repare-se que pretender que existem imagens é uma posição tão comprometida do ponto de vista de Dennett como pretender que existem representações sentenciais reais. Portanto, embora Dennett se incline para o lado anti-pictorialista da discussão (o lado de Pylyshyn) ele nunca poderia assumir um compromisso com representações sentenciais identicamente realista. O propósito de *Two Approaches to Mental Imagery* é, como o propósito de *Are Dreams Experiences*?, metodológico. Pretende-se antes de mais esclarecer a confusão conceptual na discussão acerca de imagens mentais. Quer no caso dos sonhos quer no caso das imagens está-se perante o problema da autoridade dos sujeitos relativamente à sua própria fenomenologia. Embora Dennett nem sequer ponha em causa o facto de a questão das imagens mentais ser empírica, ele defende que a controvérsia não pode ser resolvida por meio de mais dados empíricos: trata-se de um problema conceptual. A questão da existência e natureza das imagens mentais é empírica na medida em que se uma pessoa afirmar que as imagens mentais não são eventos físicos ou estruturas no cérebro e sim items no espaço fenomenal que obedecem a uma lógica diferente da lógica do espaço físico, há razão suficiente para afirmar com segurança que essa pessoa está errada. Não é preciso ciência mas apenas filosofia para prová-lo. O trabalho da filosofia é precisamente afastar doutrinas como essa, deixando o caminho desimpedido para a discussão entre a iconofilia e o iconofobia científicas. O que gera afirmações como a mencionada, e afinal constitui o problema para o filósofo, é a presunção de autoridade das pessoas acerca das suas imagens mentais, a presunção de poderem decidir acerca de si próprias se têm ou não têm imagens mentais. Ora, como se viu através de Z. Pylyshyn, a abordagem científica das imagens mentais está tão preparada «para encontrar imagens mentais não imagéticas como os físicos estavam preparados para descobrir átomos divisíveis»[519]. Dennett propõe-se portanto combater a intuição segundo a qual o que quer que a ciência descubra, o sujeito tem o direito de defender a todo o custo que tem imagens mentais. A sua proposta é um behaviorismo lógico interno, a que dá a seguinte formulação: "dizer estou a ter imagens mentais significa dizer

[519] DENNETT 1978i: 180.

que estou a ter uma certa multiplicidade-β". A multiplicidade-β é o "efeito psicológico de α", a diferença que a existência de α faz na vida cognitiva, seja o que fôr α. A letra α é usada como termo neutro para "imagem mental que o sujeito acredita ter", sem qualquer compromisso acerca da existência de representações mentais imagéticas. É possível assim definir imagens mentais como construtos lógicos a partir da multiplicidade-β, definindo multiplicidade-β de uma forma neutra, como sendo a diferença cognitiva (mas não imagética) que ter uma imagem mental faz no sujeito. O que é importante aqui é que as imagens assim definidas não têm que ter características imagéticas (ser "*image-like*")

É no contexto da discussão acerca de imagens mentais que se inicia em BS a sistematização dos princípios do que virá a ser em CE o método heterofenomenológico. O problema é saber que estatuto conceder àquilo que as pessoas afirmam acerca das suas imagens mentais. As asserções das pessoas são tratadas intencionalmente e não, por exemplo, como sons. Mas a abordagem científica das fenomenologias difere da comunicação normal, pois supõe, ao contrário da comunicação normal, especulação acerca da causa das crenças expressas. A abordagem científica das fenomenologias requer a introdução da noção de crente-como-criador-de-um-mundo. O exemplo proposto é o exemplo de Feenoman, o deus de uma tribo estudada por antropólogos. Para o crente, Feenoman existe. Para o antropólogo, aquilo que existe seguramente é a crença em Feenoman. A descrição das crenças pode conduzir o antropólogo a verdades acerca da vida mental dos crentes. No entanto, para o antropólogo deve manter-se claro que a crença em Feenoman estabelece a realidade do feenomanismo (entendido como crença em Feenoman) e não a realidade de Feenoman. No mundo criado pelo crente «não existem objectos reais mas apenas objectos intencionais»[520]. A abordagem fenomenológica substitui assim as "causas" procuradas na abordagem científica pela descrição dos objectos intencionais.

A abordagem fenomenológica tem como objectivo racionalizar as multiplicidades-β das pessoas descrevendo objectos intencionais. Pode como tal ignorar dados empíricos acerca do suporte físico da cognição, regendo-se apenas pelo "texto", i.e. pelos protocolos dos sujeitos, pelas declarações introspectivas acerca de imagens mentais. A atitude do (hetero) fenomenólogo será por definição diferente da

[520] DENNETT 1978i: 181.

atitude dos sujeitos, para quem aquilo que é acreditado é tido como verdadeiro. Para o fenomenólogo aquilo que é acreditado podem ser "meros objectos intencionais" que o sujeito toma como reais.

Está-se perante as duas abordagens que se procurava: de acordo com a abordagem A, as imagens mentais são constituídas com autoridade pelas multiplicidades-β (obtem-se assim a fenomenologia), de acordo com a abordagem B, as imagens mentais são as causas normais das multiplicidades-β. Além destas duas abordagens *legítimas* da vida mental uma terceira abordagem ilegítima aflora continuamente, a qual reúne a "autoridade" da abordagem A com a procura da "causalidade" da abordagem B, obtendo a ideia de "imagens mentais incorrigivelmente conhecidas e causalmente eficazes, num *medium* que é um espaço fenomenal". Assim se gera a inabalável convicção de que as pessoas teriam autoridade introspectiva acerca do que as suas imagens realmente são. Das três abordagens referidas, *apenas esta última é espúria.*

É claro que a discussão entre os psicólogos S. Kosslyn e Z. Pylyshyn, exemplo de oposição entre iconofilia e iconofobia, decorre no âmbito da abordagem B. Para o iconófilo, as imagens mentais existem: são efeitos de determinadas causas (como a percepção visual verídica, a ingestão de alucinogéneos ou a resolução apenas pensada de problemas como os atrás mencionados) e causas de determinados efeitos (como a apreensão da imagem mental α, a crença de que α ocorreu, e, como as crenças geram crenças, de mais crenças geradas a partir de α). Sendo a multiplicidade-β o "efeito psicológico de α", a diferença que a existência de α faz na vida cognitiva, seja o que fôr α, a multiplicidade-β não será contestada por ninguém (i.e. todos admitem que as pessoas afirmam ter imagens mentais). No entanto, se o iconófobo estiver certo, as causas normais, a descobrir empiricamente, da multiplicidade-β não serão "imagens".

A lição metodológica é a seguinte: neste tipo de estudo os sujeitos descrevem o que se passa consigo e aquilo que afirmam é recolhido e tratado como expressão de crença. No entanto, *os sujeitos têm tanta autoridade sobre a natureza das causas das imagens mentais que dizem ter como têm autoridade sobre a natureza química dos seus genes, i.e. nenhuma.* Por sua vez uma abordagem "fenomenológica" não põe em causa as credenciais epistémicas dos sujeitos que afirmam ter imagens: a autoridade dos sujeitos define um mundo. Simplesmente, as "imagens mentais" não são aqui consideradas como causas normais das multiplicidades-β mas sim como objectos intencionais. A abordagem fenomenológica é independente da investigação das

causas das multiplicidade-β, e pode mesmo permitir-se ignorar resultados experimentais (embora isso não seja recomendável). A "existência" das imagens mentais como construtos lógicos está garantida: nesse sentido as imagens mentais têm precisamente os traços que os sujeitos acreditam que elas têm.

Neste quadro, apenas a disputa entre o iconófilo e o iconófobo científicos têm significado ontológico (num sentido vagamente quineano de ontologia em que se trata de clarificar compromissos com a existência de alguma coisa no âmbito de uma teoria, sendo que o que está aqui em causa é a natureza dos veículos do conteúdo imagético). Na previsão do iconófilo científico, as causas normais do facto de termos imagens são estruturas de informação com características imagéticas, na previsão do iconófobo, as causas normais no cérebro não têm caracteristicas imagéticas. Assim sendo, aquilo que cai por terra por não ter justificação é a intuição que nos inclina a afirmar "Seja o que for que a ciência cognitiva descubra, eu tenho imagens porque sei com toda a certeza que tenho imagens mentais". De resto, a contenda entre iconofilia e iconofobia apenas poderá ser decidida por descobertas empíricas.

3.2.3 *A dor: sentir-se ser e simulação.*

Enquanto traço de mentalidade sensiente, a dor aparentemente resiste ao modelo funcionalista de consciência proposto em BS. A natureza da dor parece ser "resistente a programa" (*program resisting*). O propósito da distinção entre características mentais *program resistant* (sensientes) *e program receptive* (inteligentes) não era, no entanto, declarar a impossibilidade de simulação das primeiras. A distinção foi proposta com a finalidade de argumentar que se os sucessos na simulação computacional se têm restringido a determinadas características mentais (as características *program receptive*) isto acontece não porque existam traços da mentalidade por princípio insusceptíveis de simulação mas porque o desenho e a construção de uma entidade inteligente e de uma entidade sensiente constituem problemas diferentes. No primeiro caso o problema é de programação, no segundo o problema é provavelmente de *hardware*[521]. Ainda assim, para Dennett, esta razão é apenas parte da explicação da recalcitrância da dor a modelos cognitivos. Em *Why You Can't Make a Computer that Feels Pain?* Dennett imagina o desenho de uma entidade artificial que sentisse dor para ilustrar outros aspectos dessa recalcitrância. Dennett pensa que às objecções à possibilidade de dor "artificial", simulada ou sintetizada, subjazem problemas conceptuais, mais do que de problemas de engenharia.

[521] DENNETT 1978l: 191.

Enquanto noção funcional (i.e. por princípio definível independentemente do *hardware* do sistema que a realiza), a inteligência não é menos genuína por ser obtida por meios artificiais. A sensiência, no entanto, e nomeadamente a dor, descrita a partir do ponto de vista humano, difere da inteligência por estar ligada (em nós) a um certo tipo de carne e à perspectiva de dano e de morte. Mais do que isso, o conceito de dor sustenta, explicita ou implicitamente, a definição de "comunidade moral", i.e. o conceito de dor está «ligado às nossas intuições éticas, ao sentido de sofrimento, da obrigação, do mal»[522]. A síntese de uma entidade artificial que sentisse dor não poderia por isso desligar-se da questão do possível sofrimento dessa entidade e das obrigações para com ela. Em suma, o que torna o conceito de dor problemático é o facto de ele pertencer não apenas ao domínio da psicologia cognitiva mas também ao domínio da ética. Uma teoria subpessoal não poderá, por isso, deter o exclusivo de uma teoria da dor. No entanto, esse facto não constitui por si só obstáculo à possibililidade de existência de dor "artificial", numa entidade sintetizada.

O aspecto ético não é o único problema da noção de dor. O problema conceptual da "dor" está relacionado também com a quantidade de fenómenos que o uso comum da palavra cobre, não permitindo uma noção coerente de experiência dolorosa. Para além do caso central da dor que importa, que afecta e que é expressa, podem cair sob o conceito de "dor" fenómenos com características antitéticas às do caso central. Fala-se de "dores" relativamente a "dores" que doem mas não importam (por exemplo pelo efeito de dissociação reactiva da morfina) de "dores" que não são expressas (por paralisia – Dennett dá o exemplo da "anestesia" com *curare*), de "dores" que não são recordadas ou mesmo sequer sentidas, por exemplo devido a analgésicos (é aliás difícil decidir se um analgésico torna alguém insensível à dor ou impede a dor). Além do mais, é sabido que existem várias "profundidades" do estado de inconsciência, e que é difícil saber com segurança quão profundamente inconsciente tem que estar uma pessoa para que não sinta dor. É

[522] A distinção é de K. Gunderson (cf. DENNETT 1978l: 197).

reconhecidamente problemático afirmar ou negar que sob hipnose ou durante o sono é sentida dor. Será então que todas as "dores" unificadas pelo uso do termo comum "dor" são dores, sendo que, no limite e por reunião dessas características, é concebível uma "dor" que não é nem sentida (por anestesia), nem expressa (por impossibilidade de movimento) nem recordada (por amnésia)?

Dennett não pretende, obviamente, regressar à análise conceptual nem tomar o uso da palvra "dor" como guia para a natureza da dor (como Malcolm fizera com o sonho em *Dreaming*). No entanto estes exemplos mostram que os problemas verificacionistas dos sonhos e das imagens se repetem no caso da dor. Como poderíamos saber, por exemplo, que o efeito dos anestésicos comummente utilizados não resulta da conjugação da impossibilidade de expressão corporal com a amnésia? Certamente não podemos, por princípio, sabê-lo em primeira pessoa, mesmo que sejamos nós próprios os anestesiados inexpressivos e amnésicos. A questão da definição de "dor" não tem solução apriorista possível: os traços que pareciam conceptualmente inseparáveis na noção de dor (como localização, modificação de fins, efeito aversivo) não o são. De novo, como acontecera com as imagens mentais e os sonhos, Dennett dissolve o problema recomendando que se desista de encontrar por auto-exame e análise conceptual os traços essenciais da dor. Supor que se pode conhecer de forma incorrigível a natureza da dor através do auto-acesso é quimérico. "Dor" será o que a neurofisiologia descobrir que produz os efeitos normais e não um facto bruto na consciência que dá acesso epistémico a um tipo natural. A conclusão do artigo *Why You Can't Make a Computer that Feels Pain* é que é essa razão, conjuntamente com o facto de para se construir um modelo de A ser necessária uma teoria de A, que subjaz à recalcitrância da dor à realização artificial e não uma (de resto inexplicada) impossibilidade por princípio de simulação da sensiência.

3.3 *Modelos de consciência e natureza das experiências:* Consciousness Explained *(1991) e o* close up *do observador.*

> «These additions are perfectly real, but they are just more text – not made of figment but made of judgment. There is nothing more to phenomenology than that», Daniel Dennett[523]

[523] DENNETT 1991: 366.

É em *Consciousness Explained* (1991) que Dennett desenvolve plenamente a sua teoria filosófica da consciência. A teoria funcionalista que lhe subjaz é apresentada como um produto da ciência cognitiva e conjuga várias aportações da investigação empírica, das quais proviria um consenso emergente. Dois pontos desse consenso são: (i) a ideia de mente como sociedade de especialistas não inteligentes tal como é defendida por exemplo na obra de M. Minsky *The Society of Mind*[524] e (ii) a inexistência de um centro unificador nos sistemas cognitivos. A especialização dos agentes[525], a inexistência de um centro unificador e a sobreposição mais ou menos *ad hoc* de actividades de agentes são ideias caras não apenas a teóricos da IA como M. Minsky como também aos psicólogos evolucionistas. A intuição partilhada é que uma abordagem profícua das capacidades mentais deverá considerar estas como o resultado do funcionamento independente de mecanismos especializados, reunidos como que por *bricolage,* produzindo, nomeadamente no caso biológico, efeitos inesperados.

Uma primeira contribuição específica da teoria dennettiana para além dos pontos consensuais é a ideia de *virtualidade* (e portanto, num certo sentido, de não realidade) *da consciência.* Mais especificamente, a consciência é, de acordo com o modelo de CE, a ilusão do utilizador de uma Máquina Virtual implementada em cérebros nos quais "agentes" em competição realizam as suas tarefas específicas. A terminologia é reveladora: são conceitos funcionais, oriundos da ciência de computadores, que permitem, no âmbito do MEM, estabelecer uma ponte entre a fenomenologia e as neurociências e de algum modo explicar a forma da consciência.

O *close up* do observador levado a cabo em CE tem como finalidade explicitar as consequências filosóficas do consenso emergente na ciência cognitiva. De acordo com Dennett, a primeira consequência desse consenso é a *desmontagem da ideia de consciência como Teatro Cartesiano.* A ideia de "Teatro Cartesiano surge quando

[524] Como diz M. Minsky, «This book assumes that any brain, machine or other thing that has a mind must be composed of smaller things that cannot think at all» (MINSKY 1985: 322). O isolamento e a relativa independência das muitas partes do livro de Minsky (262 partes + apêndices) reflectem estruturalmente o estatuto dos agentes "não inteligentes" que compõem a inteligência (da sociedade) da mente.

[525] "Agentes" na terminologia de Minsky são partes ou processos simples da mente cuja interacção produz processos muito mais complexos. Minsky defende uma particular ideia acerca de agentes que será muito importante no modelo de CE: «There is nothing peculiar about the idea of sensing events inside the brain. Agents are agents – and it is as easy for an agent to be wired to detect a brain-caused event as to detect a world-caused brain event» (MINSKY 1985: 151).

se leva longe demais conceitos perfeitamente operacionais ao nível da psicologia de senso comum, nomeadamente a ideia de um agente inteligente unificado. Dennett considera impossível tomar a unidade da consciência num sistema e a unificação do fluxo fenomenológico ao longo do tempo como dadas: elas terão pelo contrário que ser *explicadas*. Essa necessidade de explicação não é óbvia: de acordo com Dennett, um certo cartesianismo tácito distorce perigosamente a imaginação da maior parte dos teóricos da cognição, mesmo que apareça sob a forma de materialismo. "Materialismo Cartesiano" é de resto o epíteto que Dennett aplica às teorias cognitivas que, contra as suas próprias evidências, continuam a utilizar implicitamente a ideia de um "centro" da experiência associado ao controlo do comportamento[526]. Se um observador pode, ao nível macroscópico, ser considerado (e considerar-se) como um ponto de vista único e um centro, em micro-escala a unidade do observador e do Teatro colapsam e o ponto de vista fragmenta-se em agentes trabalhando isoladamente e produzindo "esboços múltiplos". Em CE Dennett propõe-se retirar todas as consequências que esta situação tem sobre as noções de experiência, controlo, eu e unidade agregadas no conceito de mente.

O modelo de consciência apresentado em CE é chamado Modelo dos Esboços Múltiplos ou das Versões Múltiplas (*Multiple Drafts Model*). Procurando uma teoria da consciência capaz de capturar a natureza nas suas articulações, o modelo revelará algo que se aproxima mais de um híbrido deselegante do que de um simples metafísico ou de um *medium* para o qual que seriam traduzidos os processamentos cognitivos inconscientes. A mensagem central da teoria dennettiana da consciência é precisamente que a consciência não é um *medium*[527]. Pelo contrário, e esta é uma segunda especificidade da teoria apresentada, de acordo com o modelo de CE a consciência não é por princípio algo de diferente do processamento de informação inconsciente pelos vários agentes. Ser um item na consciência, utilizando uma imagem de Dennett, não é comparável com "estar na televisão": a televisão é precisamente um *medium* específico, e o tempo exacto da entrada de um conteúdo nesse

[526] Segundo Dennett, os filósofos apenas incorrem menos vezes na tentação de cartesianismo porque não trabalham em modelos tão específicos de fenómenos cognitivos. Quanto aos restantes teóricos da cognição aplica-se o comentário de António Damásio à ideia de materialismo cartesiano: «This notion which D&K modestly refer to as a prevailing view is far worse than that: it informs virtually all research on mind and brain, explicitly and implicitly» (DAMÁSIO 1992: 208-209).

[527] Cf. DENNETT 1993c, *The Message is There is no Medium*.

medium pode também ser específico. De acordo com o MEM, a consciência é uma espécie de *celebridade cerebral que supõe vitória numa competição.* Assim como não é possível ser famoso durante cinco segundos, não pode existir consciência de um conteúdo C que dure por exemplo alguns milésimos de segundo, sendo o conteúdo depois esquecido. Os conteúdos conscientes são, por definição e de acordo com o MEM, aqueles que perseveram de modo a ter efeitos vários na vida do organismo, nomeadamente o efeito de marcar a memória de modo a virem a ser reportados. Ao sublinhar que a consciência não é um *medium,* Dennett combate a admissão (implícita) de uma matéria mental, de um espaço fenomenal para as qualidades das experiências, os sonhos, as imagens mentais, um espaço fenomenal no qual existiria, retomando um exemplo do autor, uma vaca roxa imaginada que difere da vaca amarela imaginada[528].

Os "esboços múltiplos" da teoria da consciência são eventos de fixação de conteúdo que ocorrem em vários lugares e tempos no cérebro. Tais eventos de fixação de conteúdo são comparados com actos de fala, embora não sejam actos de fala de ninguém[529]. Dennett considera as fixações de conteúdo semelhantes a actos de fala na medida em que quer nuns quer noutros se trata de "efectuação". Fixações de conteúdo são, por metáfora, juízos rapidíssimos de si para si do Intencionador Central (ou seriam, se este existisse...)[530].

[528] DENNETT 1991: 27-28.

[529] São como actos de fala mas sem Actor e sem Fala! É esta a exclamação indignada de Otto, o interlocutor imaginário (que vários críticos consideram ser a "consciência" do autor) cujo papel em CE é mais à frente explicado, perante a noção de actos de fala (cf. DENNETT 1991: 365). Neste pequeno ponto joga-se evidentemente toda a teoria do conteúdo analisada no capítulo anterior. As pressuposições deste movimento estratégico são justificadas em CE no *Appendix for Philosophers* (DENNETT 1991: 457).

[530] Cf. DENNETT 1991: 364 «Suppose there were a Central Meaner. But suppose that instead of sitting in a Cartesian Theater watching the Presentation, the Central Meaner sits in the dark and has presentiments –it just suddenly occurs to him that there is something pink out there, the way it might suddenly occur to you that there's somebody sitting behind you/ What are presentiments exactly? What are they made of?/ Good question, which I must answer evasively first, in caricature. These presentiments are propositions the Central Meaner exclaims to himself in his own special language, Mentalese. So his life consists of a sequence of judgments, which are sentences in Mentalese, expressing one proposition after another, at tremendous speed. Some of these, he decides to publish, in English translation. This theory has the virtue of getting rid of the figment, the projection into phenomenal space, the filling in of all the blanks on the Theater Screen, but still it has a Central Meaner, and the Language of Thought. First, get rid of the Central Meaner by distributing all his judgments around in time and space in the brain – each act of discriminations or discernment or content-fixation happens somewhere, but there is no one Discerner doing all the work. And second, get rid of the Language of Thought; the content of the judgments doesn't have to be expressible in "propositional" form-that's a mistake, a case of misprojecting the categories of language back onto the activities of the brain too enthusiastically».

No entanto o Intencionador Central não existe – essa é uma das principais teses associadas ao MEM. A ideia que substitui o Intencionador Central no MEM é a ideia de "produção por pandemónio", acentrada e não previamente determinada.

Uma outra ideia deflacionária central no modelo vem já de C&C e BS. Segundo o MEM, um sistema cognitivo tem apenas que fazer o trabalho suficiente para "matar a fome epistémica". Isto significa por exemplo que se "ninguém" pergunta (algo), não é necessário que existam respostas preparadas e que "cenários" bastam para criar a ilusão de uma realidade "completa". Uma das imagens utilizadas em CE para ilustrar esta ideia são as visitas de Catarina-a-Grande aos seus domínios, encontrando as aldeias de Potemkine: a ilusão só tem que poder sustentar-se lá, nos pontos para onde o "ilusionado" olha. A esta característica está associada uma outra ilusão além da ilusão do utilizador, a que Dennett se referirá como a "ilusão da imanência".

Não há nada de original, no contexto da ciência cognitiva, na ideia de processamento cognitivo paralelo e distribuído, inconsciente, contrastando com a serialidade do "fluxo da consciência". O que é novo e polémico é afirmar que o MEM é um modelo dos próprios eventos conscientes e não um modelo dos eventos inconscientes que precedem o envio de resultados para a consciência. Assim encarado, o MEM representa a negação da singularidade e da unicidade "naturais" da consciência (o MEM propõe que o fluxo de consciência apenas *parece* ao próprio sujeito consciente único e unificado devido a uma instalação virtual de serialidade no *hardware* cerebral paralelo e nos processos que aí decorrem), a negação da diferença ou divisão entre uma preparação cognitiva pré-consciente ou inconsciente de conteúdos e a consciência ela mesma e em última análise a *negação de uma diferença de género entre consciência e não-consciência.*

O Teatro Cartesiano traz inevitavelmente consigo, além do postulado de uma relação essencial entre consciência, apresentação e unificação, que o liga à noção de *qualia*, pressupostos quanto à relação entre consciência e tempo. Nomeadamente, o Teatro Cartesiano permite supor uma correspondência natural entre o tempo dos eventos cerebrais causalmente "responsáveis" pelo fluxo de conteúdos na consciência e o tempo fenomenológico, i.e. entre sequências temporais de eventos neuronais e a sucessão dos conteúdos fenomenológicos. Mas se, uma vez feito o *close-up*, o (suposto, e

num certo nível real) ponto de vista único se revela distribuído, a defesa dessa correspondência torna-se insustentável: «Se o ponto de vista do observador tem que ser espalhado pelo volume bastante razoável do cérebro do observador, o sentido subjectivo de sequência e simultaneidade do observador tem que ser determinado por outra coisa que não a "ordem de chegada", uma vez que a ordem de chegada está incompletamente definida até o destino relevante ser especificado»[531]. O tempo da consciência não é assim o tempo "real", dos eventos físicos, mas tempo representado. A conclusão tem repercussões importantes na teoria da consciência.

Como se tem vindo a afirmar desde o início deste trabalho, a suposição cartesiana da centralidade e unicidade da aparição não tem que aparecer de forma óbvia nas teorias da cognição, sendo revelada pelo uso de termos banais e inabdicáveis como input e output, aferente e eferente, controlos, comandos, etc. O uso de tais termos supõe implicitamente um ponto de viragem, um *turning point*, que marca a disponibilidade central de um conteúdo e a passagem à consciência e a um papel executivo. Ora, se de facto existisse um *turning point* do processamento inconsciente para a consciência, estaria naturalmente determinada a ordem experienciada, o antes e o depois no tempo da consciência. Se pelo contrário não existe um lugar e um tempo onde tudo é posto conjuntamente, se as fixações de conteúdo constantemente acontecem, muitas delas sem deixarem qualquer rastro na consciência, deixa de ser legítimo pressupor realidades como "disponibilidade central" ou "início da consciência de um conteúdo x".

Uma terceira especificidade do MEM é a forma como ele configura o papel da linguagem na consciência. De acordo com o modelo, os esboços múltiplos são gerados pelos "especialistas básicos" que compõem o cérebro, que «são parte da nossa herança animal (e que) não se desenvolveram para cumprir acções peculiarmente humanas como ler e escrever mas antes para cumprir acções como baixar-se, evitar predadores, reconhecer faces, agarrar, atirar (...) e outras tarefas essen-

[531] DENNETT 1991: 107. Também esta ideia é reportada a M. Minsky (cf. MINSKY 1985: 61). No parágrafo 6.6. (*Momentary Mental Time*) de *The Society of Mind* Minsky faz várias afirmações importantes a respeito do tempo da consciência, nomeadamente as seguintes: (1) se para o "pensador global" o sentido de "agora" é bastante claro, ele é muito menos claro para um agente dentro de uma sociedade, (2) leva tempo para que mudanças numa parte da mente afectem outras partes e por isso há sempre algum atraso, (3) nenhuma agência da mente pode saber tudo o que está a acontecer ao mesmo tempo em todas as outras agências e (4) cada diferente agente da mente vive num mundo de tempo ligeiramente diferente.

ciais» (no entanto eles foram depois) «frequentemente recrutados de forma oportunista para novas tarefas, às quais os seus talentos inatos mais ou menos se adaptam»[532]. Ora, a linguagem natural é grandemente responsável pela unificação da consciência em cérebros humanos constituídos por agentes especialistas conjugados de forma *ad hoc* e produzindo esboços múltiplos. Esta insistência na importância da linguagem para a unificação não é identificável com a tese banal – que por exemplo N. Block[533] atribui a Dennett – segundo a qual a consciência seria uma construcção cultural, apenas existindo consciência a partir do momento em que se encontra culturalmente disponível o conceito de consciência. O que depende da linguagem de acordo com o MEM não é a disponibilidade do conceito de "consciência" mas sim a unificação virtual do fluxo da consciência nos cérebros humanos.

Agentes especialistas, produção por pandemónio, esboços múltiplos, contínua revisão dos esboços, vitória numa competição e unificação através da linguagem são, como se verá, algumas das ideias-chave da teoria da consciência apresentada em CE. Ao longo da obra, Dennett procede à análise de uma enorme quantidade de trabalho empírico deixando para o fim a interpretação filosófica desse trabalho. Esta interpretação opõe-se abertamente às convicções fundamentais dos defensores da consciência fenomenal mencionados no início do presente capítulo. De acordo com a maioria dos críticos, um modelo como o MEM não toca naquela que é para muitos a "verdadeira questão metafísica da consciência", a "questão obrigatória para o materialista", i.e. a questão da locação da consciência fenomenal no mundo físico, que deve ser respondida em termos de alguma forma de superveniência (do mental ao físico, do mental ao comportamental, etc). De acordo com os críticos, Dennett evitaria essa questão. Ora, Dennett considera que o seu modelo constitui de facto uma resposta, embora não tenha certamente o formato usualmente preferido pelos filósofos. Em primeiro lugar, a resposta de Dennett conjuga assumidamente elementos behavioristas, funcionalistas, eliminativistas e instrumentalistas, não considerando que haja necessidade de os distinguir. Em segundo lugar, uma teoria da consciência não consiste, para Dennett, numa listagem das condições necessárias e suficientes para a consciência mas sim na apresentação de um modelo empiricamente plausível, do qual são retiradas as con-

[532] DENNETT 1991: 253.

[533] Cf. BLOCK 1994 b, *What is Dennett's Theory a Theory of?* Este ponto é reportado a J. Jaynes (cf. DENNETT 1998e, *J. Jaynes's Software Archeology*). É sobretudo a unificação (virtual) através a comunicação de si a si que Dennett aprova em Jaynes, bem com a ideia de uma "arqueologia do *software*"

sequências filosóficas. Aliás, numa primeira leitura, CE não parece sequer ser um livro escrito por um filósofo. A teoria cuja apresentação ocupa a quase totalidade da obra é assumidamente "mecânica", como diria Ryle, e só depois, filosófica. O filósofo esconde-se até ao fim (até à Parte III de CE, intitulada *The Philosophical Problems of Consciousness*). No entanto, Dennett pensa que deve ser exactamente essa a ordem metodológica de uma teoria da consciência. O problema da filosofia da mente em geral e da teoria da consciência em particular é pretender responder às grandes questões demasiado cedo. Compreende-se que por exemplo Ned Block considere a teoria da consciência apresentada em CE exasperantemente evasiva (*exasperantingly elusive*)[534].

Antes de passar à exposição do modelo, convém chamar a atenção para dois pontos que permitem uma mais amigável navegação na estrutura "exasperantemente evasiva" da obra. Em primeiro lugar, existe, em CE, um interlocutor ficcional do autor, Otto, que apresenta objecções à teoria que vai sendo desenvolvida, defendendo a realidade irredutível da fenomenologia, i.e. do aparecimento no instante do aparecimento, defendendo que esse aparecimento é isolável de todas as disposições, defendendo que o sujeito controla a adequação do querer-dizer (*mean*) às expressões linguísticas, etc[535]. A defesa da teoria perante as objecções de Otto é importante. Será evidente para muitos leitores que Otto defende as teses correctas, ou pelo menos aponta o verdadeiro *explanandum* de uma teoria da consciência. Em segundo lugar, é necessário ter presente que tudo o que é afirmado em CE acerca de informação e fixação de conteúdos no cérebro pressupõe a teoria do conteúdo discutida no Capítulo 2 deste trabalho. O propósito de CE é precisamente edificar a teoria da consciência sobre a teoria do conteúdo.

3.3.1 *A partir de fora e de cima: SHAKEY, SHRDLU*[536] *e a heterofenomenologia.*

Até mesmo em tempos pós-behavioristas é necessário justificar a utilização, como evidência na teoria da mente, dos proustianos

[534] BLOCK 1993: 181-93.

[535] Cf. DENNETT 1993: 891 «Everything Otto says is drawn, with scant editorial amendment, from heated discussions with philosophers with impeccable credentials». Para a mais longa e articulada troca de argumentos com Otto, cf. DENNETT 1991: 362-368, *Seeing is believing: A Dialogue with Otto.*

[536] Cf. DENNETT 1978h: 115 para um comentário, anterior a CE, ao programa de T. Winograd.

relatos fenomenológicos de que os humanos são capazes. O estatuto dos relatos fenomenológicos preocupa Dennett desde C&C, e constituiu o objecto das análises dos sonhos, das imagens mentais e da dor em BS. Em CE a teorização metodológica foi aperfeiçoada e adquiriu um nome: *heterofenomenologia.*

Colocado em primeira pessoa, o problema que a heterofenomenologia se propõe abordar consiste em saber como tomar descrições fenomenológicas como esta: «Eu tinha aparentemente apenas levantado os olhos do livro que lia e tinha estado primeiro a olhar pela janela sem ver, quando a beleza do que estava à minha volta me distraiu das minhas deambulações teóricas. Luz do sol verde dourada entrava fluindo pela janela nesse dia de início de primavera, e os milhares de ramos (...) entrelaçados da árvore à minha frente eram ainda claramente visíveis por entre uma nuvem de botões verdes que despontavam, formando um elegante padrão maravilhosamente intrincado[537]». Dennett considera necessário um método neutro para tomar este tipo de relatos como evidência. Esse método deve partir do material recolhido (por exemplo: a gravação do som da voz de um humano que fala) e interpretá-lo como actos de fala. Tal procedimento pressupõe encarar a fonte dos actos como um agente racional, a partir da EI. O texto do sujeito não é um acontecimento natural, mas algo já tomado por um intérprete (por dois intérpretes, no caso de ser recolhido, o próprio e quem o recolhe). De facto «Aquilo que existe (...) são eventos de fixação de conteúdo ocorrendo em vários lugares e em vários tempos no cérebro. Estes não são os actos de fala de ninguém» (embora sejam) «de alguma maneira como actos de fala (pois) têm conteúdo e têm o efeito de informar vários processos com este conteúdo»[538].

Relatos de experiências são os dados originários da teoria da mente, até certo ponto tomados como a última palavra. No entanto, o teórico da mente não se permite qualquer certeza a priori de que estes relatos sejam relatos de acontecimentos reais. Segundo Dennett, o heterofenomenólogo aborda os textos dos sujeitos com uma disposição idêntica à do leitor de uma obra de ficção ou à do antropólogo que estuda a crença num deus. A interpretação de um texto de ficção pode conduzir a "verdades" ficcionais (por exemplo verdades acerca da Londres de Sherlock Holmes ou verdades acerca da crença em Feenoman, no exemplo que já fora introduzido no

[537] DENNETT 1991: 26.
[538] DENNETT 1991: 365.

artigo *Two Approaches to Mental Images*). Mas, como Dennett afirmara já em BS, aquilo que a crença em Feenoman estabelece é a realidade da crença em Feenoman, não a realidade de Feenoman.

O mundo heterofenomenológico é uma descrição em terceira pessoa do mundo tal como este aparece e é experienciado em primeira pessoa (T. Nagel comenta aliás, laconicamente, que com o método heterofenomenológico Dennett se limita a repetir a confusão behaviorista acerca da objectividade: embora pretenda poder passar sem a primeira pessoa, o método depende implicitamente do entendimento em primeira pessoa da consciência[539]). Ora, se o mundo nocional era desde logo considerado, no âmbito da teoria do conteúdo, como uma noção sem qualquer referência à causalidade, Dennett agora faz algumas afirmações menos claras acerca do mundo heterofenomenológico. Precisamente, o MEM admite relações entre o mundo heterofenomenológico de um sujeito e acontecimentos no cérebro do mesmo sujeito, i.e. as "fixações de conteúdo" que aí decorrem. Chega mesmo a afirmar que o teórico da consciência procurará descobrir como é que os mundos heterofenomenológicos se mapeiam em eventos no cérebro. É a isto que L. Rudder Baker chama "teste do mapeamento cerebral"[540].

O método heterofenomenológico sanciona descrições totalmente livres do fluxo de consciência. Para manter a neutralidade, parte-se dos casos não polémicos dos mundos de seres plenamente conscientes, os humanos adultos (deixando de lado por exemplo bebés e criaturas não linguísticas). Estes são aliás os sujeitos usuais de psicólogos e neurocientistas em situações experimentais (cujas instrucções são, note-se, fornecidas linguisticamente, e cujos resultados são também reportados linguisticamente). De facto, segundo Dennett, o método heterofenomenológico não é mais do que a explicitação dos princípios que orientam o estudo da mente e da consciência nas várias ciências cognitivas e na antropologia. De qualquer modo, constitui um problema para o método heterofenomenológico saber em que consiste a confirmação (e sobretudo a infirmação) das crenças dos sujeitos na sua própria fenomenologia. O espaço para "confirmação" e "infirmação" da fenomenologia fica aberto pela neutralidade do método, mas é evidentemente mais premente saber como se deverá conceber o tipo de situação em que a fenomenologia do sujeitos difere totalmente dos seus "reais refe-

[539] NAGEL 1995b: 87.
[540] RUDDER BAKER 1994.

rentes". De certo modo, o estatuto de ficção frequentemente atribuído por Dennett ao texto heterofenomenológico significa que não existe "fenomenologia real"[541]. Apetece perguntar ao autor que defendia em C&C que os cérebros não pensam e que apenas as pessoas pensam, o que é que existe se não existe fenomenologia: cérebros-que-não-pensam? De facto, de acordo com algumas formulações do MEM, fica aberta a possibilidade de o texto fenomenológico ser "sobre nada", podendo nesse caso acontecer que pareça haver fenomenologia sem que exista realmente fenomenologia. Esta possibilidade tem que ser encarada, como Dennett diria, *cum granum salis*, pois ela joga com uma ambiguidade na interpretação da afirmação segundo a qual "não há real fenomenologia". De facto, Dennett *não pode querer dizer que não existe fenomenologia*, ele só pode querer dizer que *algo na noção de fenomenologia real (entenda-se: específica, determinada)não faz sentido.*

A possibilidade de não existir uma fenomenologia real em casos em que existe uma "aparente fenomenologia" é exemplificada em CE através do caso do robô Shakey[542], capaz de identificar "visualmente" formas e de levar a cabo instruções. Shakey era uma caixa com rodas e um olho de TV, ligado por rádio a um computador. Shakey "via" e cumpria instruções, introduzidas num terminal, relativas aos objectos que via. As instrucções podiam ser algo como "PUSH THE BOX OFF THE PLATFORM". O interesse de Shakey para Dennett reside na forma como ele é capaz de distinguir visualmente por exemplo uma caixa de uma pirâmide: observadores exteriores, num monitor, podem observar em Shakey um processo de transformação de imagens que Shakey obviamente não vê (aliás Shakey não observa o seu processo de análise perceptual, e o monitor para o qual os observadores olham pode evidentemente ser desligado continuando o processo a decorrer em Shakey). O que os observadores vêem é o seguinte: uma imagem completa de televisão, um arranjo de pixels (dos quais inúmeras combinações podem acontecer) começa por aparecer no monitor, sendo depois rectificada, "purificada" e transformada num desenho com linhas,

[541] Otto pergunta (cf. DENNETT 1991: 365): «But what about the actual phenomenology?». A resposta é: «There is no such thing. Recall our discussion of the interpretation of fiction». Otto insiste (DENNETT 1991:364): «What is the text about in the unstrained sense?». A resposta é: «Nothing. It's fiction. (...) It isn't really about anything».

[542] SHAKEY foi desenvolvido no Stanford Research Institute, Menlo Park, California no fim dos anos 60 por N. Nilsson e B. Raphael entre outros investigadores (cf. DENNETT 1991: 85-95 e BODEN 1977: 280-286, 335 e 348).

que é então analisado e classificado em termos de vértices e arestas, e assim identificado como uma caixa ou uma pirâmide. O que acontece é que Shakey "tem" um programa de semântica de linhas para categorizar objectos, cuja execução supõe que sejam extraídas linhas da imagem, de modo a pôr em relevo bordos classificáveis. O caso de Shakey exemplifica um mecanismo para uma fenomenologia de imagens mentais que não envolve quaisquer "imagens reais". Num certo sentido não existem imagens em Shakey, pois as computações executadas não têm elas próprias cor, tamanho, orientação, etc. Numa experiência de ficção, Dennett "cruza" Shakey com o programa SHRDLU (de T. Winograd)[543]. SHRDLU é um programa que "vive" num mundo imaginário de blocos e que pode manipular esses blocos (imaginários...) em resposta a ordens e responder a questões sobre o que faz e porquê. Com o cruzamento obtem-se uma entidade que não apenas identifica formas como produz relatos das identificações que faz – por mais que sejam, é claro, respostas *ready made*. Se se perguntasse a esse sistema "Como distingues pirâmides de caixas?" ele poderia "responder" a diferentes níveis de profundidade (desde "não sei, é por intuição" a "faço uma busca em sequências de 10000 zeros e uns procurando padrões, i.e. sequências com determinadas características (*estas*).....". Todas as respostas seriam verdadeiras a um dado nível (haveria no entanto um limite para o conhecimento do sistema acerca dos acontecimentos no seu interior). Mas o que é relevante é que esta seria uma entidade da qual se poderia dizer (supondo que desligamos o monitor): ele "diz" que tem imagens mentais, que é essa a sua fenomenologia, mas não tem nada de semelhante. Se quiséssemos até o poderíamos ter instalado de modo a produzir relatos completamente efabulatórios sobre o que se está a passar nele, sem que ele tivesse qualquer noção do estatuto de efabulação da efabulação. Algo de análogo pode ser o nosso caso, daí a necessária neutralidade na fenomenologia.

Em conclusão, o heterofenomenólogo pode apenas, perante um sujeito consciente qualquer, dizer: "Sim, o teu mundo aparece-te assim, parece-te que tu és tu e que esses são os conteúdos da tua experiência, e essas pretensões são invioláveis". Depois, regista um "quadro neutro de como é ser esse sujeito", e para a investigação empírica daquilo que explica a existência de uma tal fenomenolo-

[543] Cf. BODEN 1977: 114-42, 163 e 171-174.

gia. Ora, se as pretensões do sujeito quanto a "como lhe parece ser" são invioláveis, pretensões explicativas já não o são, na medida em que é possível através de investigação empírica descobrir algo de "outro" em relação às entidades do mundo heterofenomenológico assim gerado, ou mesmo descobrir que os objectos que constituem esse mundo "não são feitos de nada". Esta possibilidade é muito importante em CE.

O ponto da transposição de fantasias tecnológicas para a descrição fenomenológica de humanos é mostrar que os sujeitos humanos podem estar a dizer exactamente como lhes parece ser quando fazem descrições fenomenológicas sem que haja qualquer relação simples entre esses relatos e aquilo que é causalmente responsável pela fenomenologia. O que as pessoas dizem acerca do seu fluxo fenomenológico é admitido como evidência, deve ser tomado como dado, mas a evidência só faz prova da maneira como as coisas parecem ser. O problema é que parece não haver, assim, nenhum resíduo irredutível de realidade num parecer que fosse, hipoteticamente, enganoso[544]. No limite, Dennett defenderá que é possível que pareça existir uma fenomenologia, sem que exista qualquer fenomenologia. Mesmo sem chegar a esse extremo é uma suposição básica do MEM que aquilo que parece ser a realidade da consciência para o sujeito consciente pode não ser realidade alguma e que a consciência, longe de ser um pleno, é intervalada e esparsa, sendo conscientes muito menos coisas do que usualmente se pensa.

3.3.2 *O Modelo dos Esboços Múltiplos.*

Procura-se em seguida listar os princípios do MEM em conjunto com os factos de neurociência cognitiva que os fundamentam:

(1) O processamento cognitivo é cumprido no cérebro por processos *paralelos e distribuídos* que têm uma *duração temporal* distinta de zero. Esses processos são responsáveis pelas (ou idênticos às) *fixações de conteúdo*. Tomando como exemplo uma determinada experiência visual, as fixações de conteúdo que lhe correspondem acontecem em diferentes tempos e lugares no cérebro. "Fixações de conteúdo" acontecem quando partes do cérebro entram em estados que respondem diferencialmente a diferentes traços, desde "início do estímulo" a, por exemplo, traços de "cor" e "movimento" no ambiente. Estas fixações de conteúdo localizadas "transferem" os seus efeitos para outras localizações contribuindo para mais discriminações. É um processo deste género que explica as disposições reactivas do sistema global. Dos estados com conteúdo assim fixados e distribuídos, alguns desaparecerão sem rastro, outros deixarão rastro em relatos subsequentes, em estados emocionais, na prontidão semântica, etc. Ao longo do tempo (não negligenciável) do processamento várias interferências sobre os conteúdos fixados podem ocorrer (na terminologia de Dennett, acontecem constantemente *processos editoriais* de alteração, incorporação, substituição, apagamento).

(2) Logo que alguma fixação de conteúdo acontece, ela pode ter *efeitos* (por exemplo, intervir no controlo do comportamento do sistema global). Mas o facto de os conteúdos terem sido fixados e poderem provocar efeitos não significa que eles "tenham entrado" na consciência.

(3) Todas as teorias da percepção admitem que as pessoas não experimentam directamente nem o que acontece nas suas terminações sensoriais nem o processamento interno de informação. Aquilo que elas experienciam é o resultado de fixações de conteúdos, processos editoriais e sondas. O que é específico do MEM é acentuar que as discriminações referidas são feitas *apenas uma vez* e que os conteúdos fixados não têm que ser enviados para um "supervisor central" para uma (re) "apresentação".

(4) Em *qualquer instante* da vida cognitiva de um sistema, existem *esboços múltiplos* em várias fases de edição, sendo uma questão em aberto saber se um conteúdo fixado resultante de informação processada, virá a ser consciente. Isso depende por exemplo da existência de sondas (*probes*). Citando um exemplo de Dennett, se alguém está sentado numa sala a ler, concentrado, e se se lhe pergunta se o relógio acabou de bater as horas, essa pessoa será provavelmente capaz de reconstituir a situação imediatamente passada e dizer que sim, e que foram cinco badaladas[545]. Mas se nada for per-

[544] É esse o propósito do exemplo de efeito de "espalhamento néon", um efeito de indução de cor exemplificado pelo "círculo rosa" da contracapa de CE, cujo estatuto é comentado no diálogo com Otto (DENNETT 1991: 362-364). A discussão gira em torno do seguinte: parece haver um círculo rosa brilhante, mas não há nenhum círculo rosa brilhante (i.e. o "círculo rosa" não existe na figura, o que existe é uma grelha de linhas perpendiculares, sendo as linhas negras e vermelhas, sem qualquer preenchimento rosa em forma de círculo). Não existe rosa na imagem retiniana, além das linhas vermelhas. O cérebro é induzido a distinguir uma região delimitada, com contornos subjectivos, que é "etiquetada" como sendo rosa pelos circuitos especializados em cor (DENNETT 1991: 351-353).

[545] DENNETT 1991 137-138.

guntado, nada será relatado. Supor que nesta situação ou a pessoa esteve ou não esteve consciente das cinco badaladas é uma suposição ilegitimamente absolutista e antropomorfista. Devido aos esboços múltiplos e à necessidade de sondas não é legítimo, como se verá, perguntar *quando ou onde* é que um determinado conteúdo se torna consciente.

(5) Discriminações e fixações de conteúdo são as condições de constituição da "sequência narrativa" da vida mental consciente. Mas a vida mental consciente apenas se assemelha a uma sequência narrativa, i.e. ela não é única como a sequência narrativa, devido à *multiplicidade* de versões que circulam em cada instante. A interferência ou sondagem feita em diferentes ocasiões precipita diferentes narrativas fenomenológicas (e uma intervenção muito adiada pode já não encontrar nada). Não se pode, por isso, falar de algo como a "narrativa canónica da vida mental consciente de um sujeito". No entanto, qualquer narrativa provocada dá lugar a uma "linha do tempo" (i.e. uma sequência subjectiva de eventos do ponto de vista de um observador). Como se verá, esta pode ser divergente relativamente ao tempo dos eventos neuronais (os veículos do conteúdo) que fixam os conteúdos do fluxo fenomenológico.

(6) Um observador é assim constituído por (ou consiste em) uma porção do mundo que compõe narrativas "unificadas" cujo conteúdo explícito está dependente de sondas dirigidas a essa porção do mundo.

Estes princípios traduzem a orientação verificacionista da teoria. O verificacionismo, de acordo com o qual a experiência consciente não tem realidade independentemente de efeitos na acção subsequente, é a única forma de evitar noções bizarras como *qualia* não apercebidos ou fenomenologias sem sujeito.

3.3.3 *O tempo e a consciência. Temporalização*[546].

A principal aplicação e primeiro teste do Modelo dos Esboços Múltiplos em CE é a interpretação de várias anomalias temporais

[546] Paralelamente a DENNETT 1991, a teoria das relações entre tempo e consciência foi apresentada em DENNETT&KINSBOURNE 1992, *Time and the Observer*, um artigo publicado na revista *Behavioral and Brain Sciences*. Tomar-se-á aqui como referência indiferentemente DENNETT 1991 e DENNETT &KINSBOURNE 1992.

no fluxo da consciência (o fenómeno phi com cor[547], o meta-contraste[548], o coelho cutâneo[549], os fenómenos de "referência para trás no tempo" e de "atraso da consciência de intenção" estudados por B. Libet[550] e o carreto de *slides* pré-cognitivo de W. Grey Walter). A

[547] O fenómeno phi com cor é uma versão do fenómeno phi ou percepção de movimento aparente desde há muito (desde o início do século XX) estudado pelos psicólogos. O fenómeno phi é o responsável pela continuidade da consciência visual do movimento (quando se vê por exemplo um filme, que "no mundo" consiste não em movimento contínuo mas na projecção sequencial de um dado número de imagens por segundo). No exemplo mais simples do fenómeno phi, o sujeito percebe a trajectória-de-uma-mesma-luz quando de facto, "no mundo" exterior, existem dois *flashes*, projectados sucessivamente, com um intervalo temporal muito curto e uma distância angular muito pequena. O fenómeno levou o filósofo N. Goodman a desafiar o psicólogo P. Kolers a investigar o que aconteceria se os *flashes* tivessem cores diferentes. O que se verificou foi que os sujeitos continuaram a relatar terem visto o movimento de *uma* luz, que desta vez mudava de cor a meio da trajectória (imaginária!). Em *Ways of World Making*, Nelson Goodman analisa o fenómeno (GOODMAN 1978: 71, *A puzzle about perception*). O fenómeno phi com cor pareceria dever-se a uma estranhíssima pré-cognição do segundo *flash* (por exemplo verde) ou a um adiamento da consciência do primeiro *flash* (vermelho) até o segundo *flash* ter sido percebido, havendo depois uma projecção "para trás no tempo" do que viria a ser percebido. Excluindo a pré-cognição, intuitivamente parece claro, no mínimo, que o cérebro não pode criar a trajectória ilusória até ter recebido o segundo estímulo (senão como "saberia" o cérebro que a segunda luz ia ser exactamente verde e não, por exemplo, roxa?). Portanto, o cérebro tem que "saber" que há uma segunda luz, e que ela é exactamente verde, antes de começar a criar a ilusão, o que significa que tem que "esperar" até poder perceber o todo da situação. Mas se existe uma "espera" como é possível "apagar" o que entretanto foi percebido (uma vez que o intervalo de tempo envolvido nas experiências – 200 milésimos de segundo – é demasiado grande para passar despercebido)?

[548] O fenómeno do meta-contraste acontece numa situação em que um estímulo (por exemplo um círculo) é projectado num écran, seguido, após um pequeno intervalo, por outro estímulo (por exemplo um anel, cujo bordo interior corresponde ao bordo exterior do círculo) que o "oculta". Apenas o segundo estímulo (o anel) é recordado e reportado pelo sujeito como tendo sido experienciado, o círculo aparentemente desaparece da memória do sujeito (cf. DENNETT 1991: 141-144). O exemplo é também utilizado em DENNETT 1996g, *Consciousness: More Like Fame Than Television* e DENNETT&KINSBOURNE 1992.

[549] Cf. DENNETT 1991: 142-144. O "coelho cutâneo" é o efeito criado na percepção de sujeitos por grupos de leves pancadas sucessivas no braço (pulso, cotovelo, antebraço), com intervalos temporais e distâncias regulares entre elas. As pancadas são percebidas como sendo provocadas pelos passos de um animal que sobe pelo braço acima. O efeito foi pela primeira vez estudado em 1972 pelos psicólogos F. Geldard e C. Sherrick.

[550] Ambas as situações são criadas e analisadas pelo neurocientista Benjamin Libet. Num primeiro caso, Libet (cf. LIBET 1979 e LIBET 1981) pretendeu comparar a indução de experiência sensorial por via normal, por exemplo estimulação numa mão, com a indução de uma experiência análoga por estimulação directa do córtex somatossensorial. No primeiro caso, a distância de transmissão nervosa é longa, no segundo muito mais curta. Apesar disso, Libet chegou à conclusão de que embora em ambos os casos decorra um tempo determinado desde o início da estimulação até à "adequação neuronal", no caso normal a experiência é imediatamente referida "para trás no tempo" e em alguns casos é sentida como tendo acontecido "antes" da experiência produzida por estimulação cortical. Em alguns escritos, Libet parece considerar que os casos constituem um desafio ao materialismo e que são um sintoma da dissociação entre tempo mental e tempo físico (cf. DENNETT 1991: 153-162). No segundo caso, o caso do atraso da consciência de intenção, Libet procurou determinar o momento em que o agente julga que iniciou um determinado movimento intencional, comparando-o com a "iniciação cerebral". Libet pretendia avaliar o papel da vontade consciente na acção

aplicação do MEM às anomalias temporais foi provavelmente o aspecto mais discutido do modelo[551].

O problema do tempo é importante e especial na teoria da consciência antes de mais porque *é difícil distinguir o tempo como conteúdo da consciência* (o tempo representado) *do aspecto temporal dos veículos da representação* (i.e. do acontecimento temporalmente situado dos veículos de conteúdos temporais). A questão é obviamente relevante quando se trata de decidir o que é "real" no fluxo da consciência. Por outro lado, divergências entre tempo apercebido e tempo dos eventos cerebrais, impossibilitando a identificação entre eventos mentais e eventos neuronais, parecem deixar aberta a porta ao dualismo. Dennett, que não nega que em alguns aspectos a sua teoria da consciência identifica a experiência consciente com eventos portadores de informação no cérebro, sabe que certas «propriedades dos items heterofenomenológicos podem ser consideradas essenciais – tais como a posição que os items têm na sequência temporal subjectiva, e nesse caso eles não poderiam ser identificados com os eventos cerebrais disponíveis, que podem acontecer numa diferente sequência, sem violar a Lei de Leibniz»[552].

Acrescentando à reconhecida disparidade entre tempo fenomenológico e tempo dos eventos cerebrais o princípio segundo o qual os eventos neuronais causalmente responsáveis pela fixação dos conteúdos da consciência são processos paralelos e distribuídos, está gerado o problema. Todos os casos práticos para a análise do problema do tempo em CE dizem respeito a anomalias temporais na fenomenologia dos sujeitos e mesmo (aparentes) manipulações do tempo pela mente consciente. Essas anomalias temporais conduziram inclusivamente alguns autores a afirmações dualistas — J. Eccles e K. Popper[553] por exemplo consideraram que os casos de B. Libet constituíam uma prova das capacidades supra-naturais da mente auto-consciente — ou à proclamação da necessidade de uma revolução na física[554]. Dennett, pelo contrário, considera as anomalias são apenas aparentes, geradas por erros conceptuais. De modo a pôr em evidência esses erros conceptuais, Dennett desenvolve uma interpretação dos casos referidos que prolonga a análise das fenomenologias feita em BS.

A interpretação dos casos supõe antes de mais que se abdique do mito de um momento absoluto da consciência. Este, acompanhado pela correspondente convicção de que existe sempre uma resposta determinada à pergunta acerca do instante em que aconteceu uma dada experiência, gera por sua vez duas explicações das alterações das sequências dos conteúdos na consciência que para Dennett não têm qualquer razão de ser. Dennett chama a estas explicações *revisões "orwellianas"* (i.e. revisões posteriores ao momento de uma experiência real, que é, assim, esquecida[555]) e *revisões "estalinistas"* ("revisões" pré-experienciais, i.e. correspondentes a um processo editorial anterior ao momento da consciência[556]). No quadro do MEM a distinção entre revisões pré-experienciais e pós-experienciais revela ser injustificada. O facto de se considerar que as duas alternativas exaurem as possibilidades de interpretação dos casos de "anomalia temporal" depende exclusivamente da suposição de um momento absoluto da consciência. Os responsáveis pelas revisões orwellianas e estalinistas seriam evidentemente mecanismos no cérebro, o primeiro apagando memórias de experiências reais, o segundo atrasando a consciência, de modo a inserir revisões nos conteúdos antes do momento da experiência consciente. Ambos os mecanismos pressupõem a existência de qualquer coisa a que se pode chamar "o *realmente* experienciado", independentemente de o experienciador reter ou não uma memória dessa experiência real.

voluntária e a relação desta com a iniciação cerebral inconsciente da acção (cf. LIBET 1985, *Unconscious cerebral iniciative and the role of conscious will in voluntary action*). Para isso registou a actividade cortical correspondente à iniciação da acção voluntária e a actividade cortical correspondente à experiência subjectiva de iniciação da acção e comparou-as (i.e. comparou os tempos). Libet concluí, de forma polémica, que a iniciação inconsciente de um acto voluntário espontâneo precede a "voluntariedade consciente". Nestas circunstâncias os eventos conscientes não teriam papel causal algum na iniciação da acção voluntária. A experiência subjectiva de iniciação da acção acontece algumas centenas de milésimos de segundo depois da actividade cortical, embora antecedendo a performance motora. Haveria portanto um "atraso" da consciência subjectiva de intenção relativamente à iniciação da acção. A consciência poderia assim apenas "sancionar" ou "vetar" algo já iniciado (cf. para interpretações diferentes destes resultados DENNETT 1991: 162-166 e PENROSE 1994: 386-388).

[551] E também o aspecto melhor recebido de toda a teoria da consciência, talvez porque algumas das conclusões apresentadas acerca da representação de conteúdos temporais sejam menos radicais do que aquilo que Dennett pretende fazer crer.

[552] DENNETT 1991: 459. A Lei de Leibniz é o princípio segundo o qual coisas idênticas têm exactamente as mesmas propriedades.

[553] POPPER &ECCLES 1977.

[554] Cf. PENROSE 1989 e PENROSE 1994.

[555] A expressão é uma alusão ao trabalho do Ministério de Verdade do romance *1984* de G. Orwell, que consiste em reescrever a história. Isto supõe uma experiência real de um acontecimento, que é posteriormente apagada da memória.

[556] A expressão alude a uma situação em que o que é realmente percebido é já o resultado de encenação (como no caso dos julgamentos encenados do tempo da URSS de Estaline, em que pessoas faziam *realmente* confissões falsas, que ficavam registadas como relatando, obviamente, factos "verdadeiros").

Todas as situações atrás descritas (o fenómeno phi, o meta-contraste, o coelho cutâneo, etc) são susceptíveis de uma leitura orwelliana e de uma leitura estalinista. O argumento que Dennett contrapõe à possibilidade de dupla interpretação é o seguinte. Se existisse um "lugar" único da consciência de um particular conteúdo poder-se-ia justificadamente pretender conhecer o momento da consciência desse conteúdo. No entanto, e como de acordo com o MEM não existe esse lugar único, não tem sentido procurar um tal momento. O pressuposto do momento absoluto da consciência e a consideração da alternativa entre interpretação orweliana e interpretação estalinista das anomalias temporais como exaustiva são sintomas de um indefensável *materialismo cartesiano*.

O caso da mulher que passa é um dos exemplos introdutórios da aplicação dos princípios do MEM em CE. Alguém, que está a pensar numa mulher de cabelo curto e de óculos, vê uma mulher passar a correr. Quando lhe perguntam como era a mulher, responde sinceramente que ela tinha cabelo curto e óculos. O observador exterior sabe que a mulher que passou não tinha cabelo curto e óculos. Como se justifica então o relato sincero do sujeito? Será que ele percebeu e esqueceu a mulher que "viu" ou será que ele nem chegou a percebê-la, pois a interferência de uma recordação anterior no processamento deu-se antes da experiência? O ponto de Dennett é que esta alternativa entre uma revisão orwelliana e uma revisão estalinista, que tem, sem dúvida, sentido por exemplo numa macro-escala de acontecimentos históricos não tem sentido na micro-escala em causa. De novo, a boa interpretação do caso segue os moldes da interpretação do fenómeno phi, o exemplo elementar de todos casos referidos. Recapitulando: dada uma anomalia temporal, coloca-se uma aparente alternativa entre uma explicação por revisão estalinista e uma explicação por revisão orwelliana. À primeira vista é plausível considerar que a alternativa se coloca assim para o próprio sujeito da experiência, quando quer saber o que se passou consigo, e mesmo para cientistas que usassem mecanismos de *scanning* dos veículos de conteúdo no cérebro. No entanto, em ambos os casos, na tentativa de auto-conhecimento pelo sujeito e na inspecção do seu cérebro por meios tecnológicos, ambas as interpretações, orwelliana e estalinista, só se configuram pressupondo uma Grande Divisão entre consciência e processamento ainda inconsciente. Ora, esta divisão é inadmissível de acordo com os princípios do MEM. A distinção entre esquecimentos

imediatos por um lado e preenchimentos e projecções para trás no tempo[557] por outro só faria sentido se se concedesse a distinção absoluta entre consciência e não-consciência. O MEM, que não concede a referida distinção, permite uma outra explicação do caso. O que se passa no fenómeno phi, no caso da mulher que passa e nas anomalias temporais em geral, é simplesmente que *o tempo subjectivo* (o conteúdo temporal pensado) *não corresponde ao tempo do processamento* (que diz respeito aos veículos da representação e à situação destes no espaço-tempo físico). Assim, para a criação retrospectiva *de conteúdo*, que de facto acontece nestes casos, o cérebro não tem que fazer qualquer preenchimento ao nível dos veículos, mas apenas que "concluir" em termos de conteúdo. Não existem portanto preenchimentos e projecções ao nível dos veículos, apenas um "saber-que", que cumpre essa função.

Dennett e Kinsbourne utilizam uma história ilustrativa dos princípos de *design* através do quais o cérebro lida com este tipo de disparidades entre tempo fenomenológico e tempo físico[558]. Eles afirmam que perguntar quando é que alguém se tornou consciente de um conteúdo A é como perguntar, no contexto da história que se segue, quando é que o Império Britânico se tornou consciente da trégua. Em 8 de Janeiro de 1815 houve em New Orleans uma batalha na qual morreram desnecessariamente (uma vez que já tinha sido assinada a trégua na Europa) mais de mil soldados ingleses. Imagine-se a comunicação das notícias no Império Britânico antes da rádio e do telégrafo: a notícia da trégua parte da Europa para a América, para a Índia e para África. Enquanto a notícia segue, a batalha de New Orleans é travada e as notícias da derrota são por sua vez enviadas para a Europa, para a Índia, etc. O comandante do exército em Calcutá recebe *primeiro* a notícia da derrota de New Orleans, *depois* a notícia da trégua assinada na Europa. O comandante do exército em Calcutá só não pensa que a batalha foi travada antes de o tratado de trégua ter sido assinado devido à prática de datar as cartas que transportam a informação. Ora, datar uma carta é precisamente incluir uma representação de informação temporal (i.e. um conteúdo temporal) num veículo de conteúdo que tem ele próprio uma dada situação espaço-temporal totalmente diferente.

[557] "Projecção para trás no tempo" é uma expressão usada quer por B. Libet quer por N. Goodman. São estas as interpretações que N. Goodman faz do fenómeno phi com cor (cf Dennett 1991: 127), e Dennett ensaia a transposição para o caso da mulher que passa.

[558] DENNETT & KINSBOURNE 1992. A história não é verídica em todos os pormenores.

O conteúdo temporal distingue-se do tempo (por exemplo do tempo de chegada a um ponto A) do veículo de conteúdo. No caso da história apresentada, a data da chegada da carta às mãos do comandante é totalmente irrelevante no que respeita à informação temporal que transporta. Dennett sugere que algo de semelhante se passa no cérebro.

É a distinção entre tempo da representação e tempo representado (que não é senão uma instância da distinção entre veículos de conteúdo e conteúdos representados) que subjaz às (aparentes) anomalias temporais nos casos referidos. Eventos neuronais são eventos em tempo real, a sequência temporal na consciência é uma questão de conteúdo. De resto, a representação de conteúdo temporal pelo cérebro não utiliza sequer necessariamente tempo no cérebro. Assim sendo, o cérebro pode inserir elementos retrospectivos (de conteúdo) sem qualquer causação "retrospectiva" ou projecção para trás (*backwards*, no termo de B. Libet e de N. Goodman) no tempo. A conclusão é que as propriedades temporais dos estados responsáveis pela discriminação, os estados dos veículos de conteúdo, cuja situação espaço-temporal pode ser determinada no cérebro, não determinam as propriedades temporais da experiência subjectiva (i.e. a "temporalização"). Assim, a fixação da situação temporal um dado conteúdo temporal não é função de uma ordem de chegada dos veículos de conteúdo a algum ponto A (no caso, no cérebro).

Voltando aos casos do meta-contraste e do coelho cutâneo, eles são exemplos de "torções espaço-temporais" explicáveis por meio da distinção entre conteúdo e veículo de conteúdo numa situação em que existem esboços múltiplos e vitória de uma interpretação. O fenómeno do meta-contraste é usualmente interpretado de forma estalinista (i.e. considera-se que o segundo estímulo, o único que os sujeitos afirmam ver, impede a percepção do primeiro). O que o MEM sugere é que o cérebro, tendo que lidar a cada instante com muitos acontecimentos, e portanto fazendo suposições simplificadoras sobre os esboços múltiplos, avança uma *conclusão conservadora em termos de conteúdo*, de saber-que, segundo a qual só existe um anel, "convertendo" assim o contorno exterior do círculo no contorno interior do anel. Não há razão para afirmar que o primeiro estímulo (o círculo) chegou a ser percebido conscientemente pelo facto de esboços relativos a ele terem chegado a circular (eles poderiam até ter efeitos inconscientes). Esses esboços "saíram de circulação", i.e. foram vencidos e substituídos pela versão revista (no caso, pela conclusão conservadora "só existe o anel"). Também o fenómeno do "coelho cutâneo" é explicável mediante esta ideia de "suposição simplificadora". No caso do coelho cutâneo a suposição simplificadora assume uma distribuição regular das pancadas no braço (os passos do "animal"), assim sendo vencidas interpretações parciais anteriores (esboços que chegaram a circular) relativos a grupos de pancadas em lugares particulares. Evidentemente, só depois de todas as pancadas acontecerem é que o cérebro pode adoptar a versão final. No entanto, quando o faz as outras versões são pura e simplesmente vencidas (embora possam deixar rastros, efeitos laterais).

A interpretação dos problemáticos casos de Benjamin Libet de referência para trás no tempo (*backwards referral in time*) e de atraso da consciência subjectiva de intenção é uma prova particularmente decisiva para as capacidade do MEM. Numa das suas experiências, Libet pretendeu comparar os tempo envolvidos respectivamente na indução de experiência sensorial por via normal (por exemplo, estimulação numa mão com um breve choque) e na indução directa de uma experiência análoga por estimulação do córtex somatossensorial. Os sujeitos são pacientes de neurocirurgia despertos e capazes de fazer relatos da estimulação. Repare-se que se no primeiro caso (a estimulação normal) a distância para a transmissão nervosa é longa e no segundo (a estimulação cortical directa) ela é muito mais curta. Apesar disso, perguntando aos pacientes qual sensação foi sentida primeiro e analisando as suas respostas, Libet chegou às seguintes conclusões:

(1) Em ambos os casos decorre um tempo determinado desde o início da estimulação até à "adequação neuronal" (na linguagem de B. Libet a adequação neuronal acontece no momento em que existe experiência consciente do estímulo). Há portanto um atraso (*delay*) da consciência relativamente à estimulação que Libet estima em cerca de 500 milésimos de segundo.

(2) Seria natural, dada a diferença das distâncias, que a adequação neuronal subsequente à estimulação cortical directa acontecesse mais rapidamente do que a adequação neuronal subsequente à estimulação normal. Não foi isso que (de acordo com as interpretações que Libet faz dos relatos dos sujeitos) se verificou nas experiências de Libet. A adequação neuronal foi atingida mais cedo no caso da estimulação normal. Numa outra experiência, em que o córtex somatossensorial esquerdo foi estimulado antes da mão

esquerda (de uma forma que, esperava-se, originaria a sequência subjectiva "sensação na mão direita seguida por sensação na mão esquerda") a sequência reportada pelo sujeito é "sensação na mão esquerda seguida por sensação na mão direita".

(3) Para a experiência normalmente induzida ser sentida como tendo acontecido "antes" da experiência produzida por estimulação cortical, ela teve que ser referida "para trás no tempo" (*backwards in time*), de uma forma que é, segundo Libet, "automática".

Na terminologia de Dennett, o modelo de Libet é "estalinista": os processo editoriais de revisão acontecem *antes* do momento da adequação neuronal. Parte da excitação associada a estes casos deve-se ao facto de Libet tender por vezes a considerá-los como provas da dissociação existente entre tempo mental e tempo físico e base para argumentos dualistas. Isto foi aproveitado por alguns autores como ponto de partida para especulações acerca de capacidades neurofisiologicamente inexplicáveis da mente auto-consciente. As interpretações de Libet foram contestadas por várias razões[559] ou simplesmente afastadas cepticamente. Dennett pensa que é antes de mais importante pôr em relevo as suposições implícitas que lhes subjazem. Antes de mais, nada haveria de anormal nos casos de Libet se não estivessem a ser considerados os relatos verbais que os sujeitos fazem dos seus juízos temporais. Libet chega às suas conclusões procurando estabelecer correspondências entre a sequência de eventos no cérebro e sequências subjectivas relatadas pelos sujeitos. Todo o caso pode portanto ser rebatido sobre a correcta interpretação dos relatos dos sujeitos. Ora os relatos dos sujeitos, considerados enquanto enunciações significativas e não enquanto sons, acontecimentos físicos, são construtos teóricos resultantes de uma interpretação dupla, a do próprio sujeito e a do

[559] Por exemplo Patricia Churchland (CHURCHLAND 1981 a e CHURCHLAND 1981b) repetiu as experiências de Libet, dando diferentes instruções ao sujeitos, pedindo-lhes nomeadamente para dizer "go!" quando sentissem – i.e. estivessem conscientes de – o estímulo normalmente induzido. De acordo com Churchland as disparidades de Libet não se verificam. Libet replicou que o "go!" poderia ser inconscientemente iniciado, e que, de qualquer modo, o desenho da experiência de Churchland era forçado, pois os sujeitos deveriam ter o tempo que quisessem para examinar introspectivamente a evidência (cf. LIBET 1981 e DENNETT 1991: 160 para um comentário). Libet evoca um privilégio de escolha do momento da sondagem pelos sujeitos, mas não tem argumentos que sustentem essa pretensão. No entanto, segundo Dennett, também Patricia Churchland não distingue, nas suas interpretações, o tempo da representação do tempo representado embora note acertadamente que nada distingue estas ilusões temporais das espaciais. Isto significa que quer Libet quer Churchland pressupõem aquilo que segundo Dennett e de acordo com o MEM não podem legitimamente pressupor: *um momento absoluto da consciência de um conteúdo*. Só assim lhes é possível discutir acerca de exactamente quando as sensações foram sentidas.

experimentador, e não observações factuais exactas ou registos de parâmetros físicos (ao contrário dos registos da actividade cortical com os quais vão ser comparados). De resto, o próprio Libet admite que a referência subjectiva para trás no tempo tem lugar na "esfera mental", não sendo aparente a nível cerebral. O problema é que Libet trata os *relatos dos sujeitos* como *observações empíricas*. Ora, os relatos dos sujeitos podem ser considerados observações empíricas do experimentador, mas aquilo que os relatos relatam não (o conteúdo dos relatos, pertence, de acordo com o MEM, ao mundo heterofenomenológico dos sujeitos). Indiciando saber que os casos lidam com *conteúdo* temporal e não com tempo real, Libet estabelece uma analogia com os casos de "projecção espacial", i.e. os casos em que o sujeito vê o espaço-do-mundo fora de si e não o espaço-no-cérebro com que vê. Nestes casos, o facto de a configuração espacial e a localização dos eventos neuronais em causa serem muito diferentes da configuração espacial e localização pensadas não põe em causa o materialismo. Como o fenómeno phi, também as experiências de Libet lidam com a representação de conteúdos de tempo no cérebro. Todo o problema de Libet surge de uma comparação directa entre conteúdo (relatado) e veículos, acompanhada da pressuposição da existência de um momento absoluto da consciência. Libet procura o tempo da experiência subjectiva de um estímulo, por oposição à resposta comportamental que admite poder ser anterior. Nos casos da referência para trás no tempo de Libet trata-se portanto, segundo Dennett, de um problema quanto à interpretação de justaposições menos usuais do tempo fenomenológico e do tempo físico, e não de algo de fisicamente extraordinário ou impossível (como seria impossível por exemplo um caso em que o tempo entre o estímulo-de-cão e a enunciação "cão" seria menos do que o tempo fisicamente requerido para o conteúdo se estabelecer e mover através do sistema). As experiências de Libet são, assim, uma tentativa interessante mas inconclusiva de estabelecer alguma coisa acerca da forma como o cérebro representa a ordem temporal. Fica em aberto uma questão mecânica relativa ao modo como o cérebro manipula representações de tempo. No entanto, conceber a questão em jogo como uma questão acerca dos mecanismos da temporalização no cérebro, sem pressupor um tempo absoluto da consciência de conteúdos, é um primeiro passo no bom caminho.

A interpretação feita por Libet das experiências acerca da consciência de intenção pressupõe também ela que a questão acerca do

momento absoluto da consciência é uma questão de facto. Para Dennett o interesse destas experiências é o facto de elas lidarem com a *voluntariedade mental.* As experiências, como foi atrás exposto, comparam a consciência subjectiva de intenção com a iniciação cortical da acção e concluem que existe um atraso da consciência subjectiva relativamente à iniciação cerebral, medida com eléctrodos, que de acordo com Libet registariam os eventos neuronais que determinam a iniciação da performance da acção. De acordo com as experiências de Libet, a decisão consciente do sujeito, a sua consciência subjectiva de decisão, acontece *post factum*, o que evidentemente nega o estatuto de *decisão* consciente, e, em geral, o papel executivo puro da consciência.

Mais uma vez, Dennett sublinha o quanto os resultados de Libet (que pretende comparar o instante t em que os sujeitos pensam que decidem com o instante t–1 dos eventos neuronais que determinam a iniciação da acção) dependem de *juízos de simultaneidade* dos próprios sujeitos. Nas referidas experiências, Libet pede aos sujeitos que tomem uma decisão (flectir a mão, por exemplo) e que nesse instante registem a posição de uma marca num dispositivo que serve de relógio. Assim, os sujeitos deverão poder reportar qual era a posição da marca no instante da sua decisão. Sem a relação estabelecida pelos relatos dos sujeitos, os quais envolvem representações mentais de tempo, não haveria nada a interpretar. No entanto, e para não falar na pseudo-aleatoriedade dos actos voluntários estudados, o problema é que apesar de Libet começar por distinguir o que é representado (o conteúdo) do momento em que o conteúdo é representado (uma característica do veículo do conteúdo), ele alcança a conclusão acima citada (segundo a qual a iniciação inconsciente precede a decisão consciente) partindo de premissas relativas ao que é representado (o conteúdo) e chegando a conclusões acerca do veículo da representação (o momento da representação). O problema subjacente a todos estes casos é, desde logo, o de saber o que significa "o tempo de ocorrência" de uma representação interna (no caso, marcada pela posição no relógio), uma vez que ela não é "una" em termos de veículos, nem espacial nem temporalmente. E aliás a tarefa pedida ao sujeito (determinar onde estava a marca quando foi tomada a decisão) requer ela própria uma decisão voluntária. Em geral, Dennett considera estas experiências de Libet pouco esclarecedoras devido ao carácter pouco natural do seu desenho: juízos de temporalidade como aqueles que são pedidos na

situação de Libet não têm um papel no controlo do comportamento normal.

Uma experiência também ela relativa à consciência de intenção e de iniciação de uma acção e que, segundo Dennett, permite evidenciar de uma forma mais simples as conclusões que se impõem, é a experiência de aparente pré-cognição que Gray Walter realizou com um carreto de *slides*[560], na qual o avanço deste é provocado pela amplificação do sinal de um eléctrodo implantado no córtex motor do sujeito e não pelo movimento do dedo do sujeito, ao contrário do que este crê. O efeito é de uma *aparente antecipação do acto à decisão subjectiva.* A surpresa dos sujeitos deve-se, segundo Dennett, à normal monitorização que acompanha a iniciação de tarefas: o cérebro espera *feed-back* das acções iniciadas, que neste caso chega cedo demais. A diferença (muito importante) relativamente aos casos de Libet é que neste caso, o juízo temporal do sujeito (acerca de quando foi iniciado o quê) não é artificialmente requerido, nem é obviamente deliberado. A exploração de experiências como esta pode produzir conclusões acerca do normal atraso esperado do *feed-back* visual relativo a uma acção iniciada, mas não conclusões acerca da situação temporal real da decisão: simplesmente não é isso que está em causa.

Em geral, a razão para os fenómenos de torção do tempo representado relativamente ao tempo dos veículos do conteúdo é o facto de os cérebros "viverem" pressionados pelo tempo na sua tarefa de controlar um corpo num ambiente de condições em constante variação – um cérebro tem que extrair informação relevante em tempo útil, para a usar nomeadamente em antecipações, os canais de tratamento de informação são vários, a comunicação entre agentes não é a melhor e por melhor que fosse ocuparia sempre tempo, há muita coisa para perceber, muitas acções para iniciar a tempo, etc. O problema empírico consiste em saber como são possíveis estes fenómenos de temporalização, i.e. de fixação de uma sequência temporal pensada. A sugestão de Dennett é que o cérebro procede em casos como os referidos por adaptações sensíveis ao conteúdo (*content-sensitive settings*): as inferências temporais envolvidas nos arranjos e simultaneizações poderiam ser conseguidas comparando correspondências de conteúdo de baixo nível de vários arranjos de dados.

[560] Cf. DENNETT 1991: 167-168.

O objectivo geral da análise que Dennett faz destes casos de (supostas) anomalias temporais é reiterar o princípio básico do MEM segundo o qual é impossível falar da consciência como uma questão de tudo ou nada, apenas sendo possível fazer sentido do (suposto) *espaço fenomenal* dos acontecimentos pensados ou relatados pelos sujeitos como um *espaço lógico*, cujas propriedades são constituídas pelas crenças dos sujeitos. O grande problema após o primeiro teste ao MEM que as "anomalias temporais" representam é evidentemente o resto constituído pela ideia de *sondas que precipitam narrativas*. A ideia de *self-probe* ou auto-sonda e a sua relação com a fixação dos conteúdos da vida mental própria determinada e explícita é essencial na teoria dennettiana da consciência. O problema é que ela parece reintroduzir a *auto-iniciação*. Para além disso, e ainda no que respeita ao tempo, persiste a intuição forte de que «as nossas experiências de eventos ocorrem na mesma ordem na qual as experienciamos a ocorrer. Se alguém pensa o pensamento "Um, dois, três, quatro, cinco" o seu pensamento de "um" ocorre antes do seu pensamento de "dois" e assim por diante»[561]. Dennett admite a verdade do princípio ao nível macroscópico, o nível do observador (mesmo que se trate do auto-observador e da sua fenomenologia). Tudo o que o MEM pretende estabelecer é que *ao nível microscópico o princípio não tem sentido*. No entanto, como N. Block nota[562], tudo nas análises dennetianas das anomalias temporais depende da suposição implícita de que a consciência fenomenal não existe. De facto, nada no processamento paralelo distribuído nos impede de considerar que alguns eventos mentais no sistema são fenomenalmente conscientes enquanto outros não o são. Um defensor da consciência fenomenal não é obrigado ao compromisso com um lugar único no cérebro que corresponda à "passagem" de conteúdos para a consciência nem com memórias infalíveis. Aliás, e quanto ao facto de o registo e a memória serem no MEM criteriais para a consciência, de modo que se algo não fica registado não é considerado real, poder-se-ia objectar que inúmeras coisas reais desaparecem sem rastro, i.e duram muito pouco (por exemplo certas partículas estudadas pelos físicos) não sendo por isso menos reais. Porque não poderia ser esse o caso da consciência fenomenal de determinados conteúdos? Pensar o contrário faz depender a consciência fenomenal da forma como nós podemos pensar os

[561] DENNETT 1991: 168.
[562] BLOCK 1997c.

conteúdos da consciência, faz depender a realidade de alguma coisa dos limites da nossa própria forma de apercebermos a realidade. É claro que neste caso, se a consciência fenomenal existe e se ela não é uma questão de juízos, deixa de ser arbitrária a fixação do momento da consciência como Dennett pretende, i.e. deixa de ser "impertinente" a diferença entre revisões orwellianas e estalinistas. Ned Block[563] faz notar a propósito que Dennett e Kinsbourne confundem, nas suas críticas aos autores que procuraram interpretar as anomalias temporais, um compromisso com a existência de consciência fenomenal com uma confusão entre tempo representado e tempo da representação. Segundo Block, por exemplo Patricia Churchland não faz confusão nenhuma na discussão dos casos de Libet: o que ela faz (de uma forma aparentemente inadmissível para Dennett, mas totalmente correcta para Block) é simplesmente *admitir a realidade da consciência fenomenal.*

3.3.4 *De novo a partir de dentro e de baixo. A evolução da consciência do ponto de vista da terceira pessoa e evitando considerar a representação do que quer que seja. Fronteiras, razões, sensiência e futuro. Evolução no cérebro. Máquinas Virtuais instaladas.*

Como se viu, o MEM tem como primeira aplicação a interpretação de aparentes anomalias temporais na fenomenologia de sujeitos. No entanto, o que em última análise justifica o modelo são teorias da evolução (do cérebro e no cérebro). De resto, grande parte dos conceitos básicos do modelo de consciência proposto por Dennett – conceitos como sociedade de agentes, Máquina Virtual, ilusão do utilizador, etc – provêm do pensamento sobre máquinas computacionais artificiais[564]. A diferença entre uma perspectiva que toma relatos de sujeitos sobre aquilo em que acreditam e outra perspectiva em que se parte de processos físicos e se pergunta como poderiam eles dar origem a estados conscientes é a diferença entre *um ponto de vista a partir de cima e de fora e um ponto de vista a partir de dentro e de baixo*. Desde sempre Dennett pretendeu conjugar ambos os pontos de vista na teoria da mente.

[563] BLOCK 1997c: 177.
[564] O que torna especialmente curioso procurar descobrir como supera Dennett em CE os defeitos simétricos que imputa a duas disciplinas da cognição: à IA Dennett atribui um pendor filosófico para a abstracção, à neurociência uma ausência de teoria para as arquitecturas cognitivas de nível elevado.

Na análise feita a partir de dentro e de baixo da evolução da consciência não está em causa o mundo-texto dos sujeitos mas sim a passagem de entidades físicas não-conscientes a entidades conscientes, por um processo evolutivo[565]. A história da evolução da consciência apresentada por Dennett em CE não é original, antes resultando de contribuições de inúmeros teóricos da mente, desde A. Turing e J. Von Neumann a M. Minsky, A. Newell, D. Hofstadter, G. Edelman, W. Calvin, J.Holland e W. Levelt[566] e sendo um aperfeiçoamento dos princípios esboçados em C&C. A ideia chave é a necessidade, que terá sido experimentada pelos sistemas cognitivos, de uma melhor comunicação interna de informação, a qual terá conduzido à criação de *hábitos de auto-exortação*, que são os "substitutos mecânicos" do olhar interno frequentemente identificado com a consciência. Em geral Dennett considera que o *design* para a consciência resulta da sucessão de três processos a que chama (1) o aparecimento das fronteiras e das razões, (2) a produção de futuro e (3) a evolução no cérebro e o Efeito Baldwin.

O início da narrativa remonta ao ponto que não existiam razões mas apenas causas. Não existem "razões" enquanto não existem entidades com interesses, e estes aparecerão a partir do interesse de replicadores na auto-replicação. A partir do momento em que existem replicadores "interessados" na replicação aparece a divisão daquilo que existe em "bom" e "mau", relativamente ao interesse do replicador. Com o interesse na auto-preservação as fronteiras tornam-se importantes e passam a existir processos que têm como função fazer perseverar interesses e fronteiras. Passa assim a haver "razões" (i.e. fins-para-um-interior), e não apenas causas de movimentos. Essas razões não são no entanto "apreciadas": elas são razões em entidades, e não razões para as entidades, que seriam representadas explicitamente por elas para elas. Dennett chama-lhes *free-floating rationales*.

Os factos primordiais da história evolutiva da consciência são assim: (1) a existência de "razões" para reconhecer, (2) a existência de "pontos de vista" para avaliar essas razões, (3) a existência de "agentes" que distinguem um interior de um exterior, (4) a efectuação dos reconhecimentos por uma pluralidade de rotinas cegas e mecânicas, (5) a inexistência, na fronteira que define o interior, de um "Quartel General", i.e. de um centro, (6) a não importância da origem dos dispositivos efectuadores dos reconhecimentos (o que funciona como x, é um x), (7) a existência de elementos com múltiplas funções no seio da economia de um mesmo organismo. A sensiência e a produção de futuro são as linhas de acordo com as quais Dennett analisa a evolução de entidades a partir destes factos primordiais. Nos organismos sensientes mais simples, a "sensiência" é simplesmente o facto de o organismo estar instalado (*wired*) para tocar e recolher ou avançar em resposta ao que é bom ou mau no ambiente. Distinguir aquilo que é "bom" daquilo que é "mau" para a entidade é a razão de ser da sensiência. No entanto, a efectuação da distinção não tem que passar por algum centro da entidade. Esta ligação no sentido mais directo possível (o sentido de "tocar") da sensiência do organismo às fontes do bem e do mal será segundo Dennett a base dos "traços mais terríveis e deliciosos da consciência". A sensiência resultante da evolução biológica é portanto concomitante de uma distinção entre bom e mau para uma entidade cuja existência individuada corresponde a uma divisão entre interior e exterior.

O segundo passo em direcção à consciência é a produção de futuro, nomeadamente a efectuação do controlo do comportamento de entidades por sistemas nervosos progressivamente mais complexos. A chave do controlo é a capacidade de aperceber e antecipar traços relevantes no ambiente. Os cérebros, resultado distante deste processo, são essencialmente máquinas de antecipação. Um organismo ligado da forma directa referida quanto à sensiência não tem muito futuro (não no sentido de ser incapaz de sobreviver, mas no sentido de não ter muito futuro a fazer parte do seu presente). Noutras palavras, o organismo sensiente elementar não faz *planeamento*. No entanto, alguns outros organismos desenvolveram talentos antecipativos adaptados a regularidades ambientes e incorporaram-nos no seu *design*. Estes talentos antecipativos podem ser por exemplo, e de forma ainda verificável nos humanos, baixar-se para evitar o que vem contra nós ou reacções automáticas de alerta perante eixos verticais de simetria que significariam "está alguém a olhar para mim" (um predador, uma presa, um rival, um parceiro sexual, de qualquer modo uma coisa relevante no ambiente). Esses mecanismos são crus e grosseiros: eles não são feitos para serem

[565] Uma das inspirações metodológicas de Dennett para a abordagem do problema da consciência a partir de dentro e de baixo são, de resto, os "ensaios de psicologia sintética", de V.Braitenberg. Os ensaios de psicologia sintética envolvem uma série de mecanismos, que vão desde entidades claramente não vivas até entidades aparentemente biológicas e psicológicas (cf. BRAITENBERG 1984).

[566] Cf. por exemplo MINSKY 1985, NEWELL 1992, EDELMAN 1987, EDELMAN 1989, HOFSTADTER 1995 e LEVELT 1989.

precisos e verdadeiros mas para serem rápidos e económicos. São por isso despoletados muito facilmente, dando frequentemente lugar a falsos alarmes e pondo, além do mais, em conjunto (i.e. categorizando como "o mesmo") uma *motley crew*, na expressão de Dennett, i.e. um conjunto heteróclito de elementos (e não elementos de categorias naturais bem delimitadas que seriam assim reconhecidas). Um funcionamento grosseiro é suficientemente bom se um número suficiente de vezes, na situação para que foram feitos os mecanismos, for gerada a acção apropriada.

A atenção de psicólogos evolucionistas e filósofos tem sido atraída, na análise destas questões, pela ligação dos estados de consciência de organismos a este tipo de alarmes filogenéticamente antigos[567], conduzindo à hipótese segundo a qual *não poderia haver estados de sensiência emocionalmente neutros*. A natureza não faria máquinas epistémicas, máquinas de pensar a partir de fora, "como se não fosse nada com o organismo" que pensa, pois não teria razão para isso. Criaturas biológicas não poderiam subitamente tornar-se simplesmente «recolhedoras desinteressadas de informação» (pensantes-observadores) porque «os repórteres nelas são os alarmes e os excitadores dos seus antepassados, nunca enviando uma mensagem intacta, mas sempre com algum vestígio de torção editorial positiva ou negativa»[568]. A razão de ser da evolução dos sistemas sensoriais, e posteriormente dos sistemas nervosos, que estabelecem as ligações entre o interior e o exterior de entidades, não é a detecção de categorias naturais, mas a sobrevivência dos organismos. Estes sistemas servem para manter os organismos vivos: eles têm um propósito narcisista e não um propósito epistémico. Antes de qualquer controlo "futurizante" existir, a maior parte da actividade das entidades era controlada em "piloto-automático": os vários controlos especializados funcionavam separadamente sem mobilização de todos os recursos da entidade. No entanto, as respostas de antecipação – que podiam ser alarmes – traduziram-se na possibilidade de mobilizar todos os recursos da entidade para lidar com uma emergência, abrindo uma *arena centralizada temporária para o controlo*. De acordo com Dennett, estes episódios breves de vigilância aumentada por interrupção da actividade em curso do organismo são provavelmente os precursores

[567] Cf. HUMPHREY 1995.
[568] DENNETT 1991: 181.

evolutivos dos estados conscientes[569]. Teriam começado por ser reacções a sinais de alarme, interrupções ligadas a situações de emergência nas quais seria conveniente a mobilização geral na entidade. Dada a sua utilidade para uma actualização geral do estado do organismo, os animais começaram a entrar nesses modos mais frequentemente. Esta vigilância tornada usual permitiu a evolução de um novo tipo de comportamento em certas criaturas, os "informívoros" (na expressão do psicólogo George Miller)[570], que exploram regularmente, e nos quais os estados de vigilância se tornam um hábito. Eles adquirem informação pela informação, já não necessariamente apenas de forma despoletada pelo exterior e à medida que é necessária, mas de forma internamente iniciada. Mesmo assim, e dadas as origens do comportamento, o *emotional overtone* desta "vigilância para a informação" manter-se-á.

Até aqui está em causa a história evolutiva dos sistemas nervosos, de acordo com os princípios gerais da evolução por selecção natural. O passo seguinte da evolução da consciência diz respeito ao funcionamento da selecção natural nos interiores não inteiramente fixados (*hard-wired*) dos próprios sistemas nervosos. Começa então a evolução no cérebro, terceira etapa para a evolução da consciência. A plasticidade que a evolução no cérebro supõe teria evoluído de forma aproximadamente contemporânea do processo acima descrito. Apareceram no entanto nesta via dois *media* para uma evolução muitíssimo mais rápida do que a evolução dos organismos: os cérebros e os memes (o termo "meme" é de R. Dawkins[571], que o utiliza para nomear replicadores que não são as moléculas de ADN, mas que estão sujeitos a um processo de evolução considerado análogo). O darwinismo neuronal – na expressão de G.Edelman[572] – aqui em causa tem sido explorado teoricamente por vários autores, como J.P.Changeux, W.Calvin, J.Holland com os seus algoritmos genéticos, bem como pelo próprio G. Edelman. Independentemente da particular teoria, o que é importante para Dennett é conceber os princípios que regem os mecanismo neuro-

[569] DENNETT 1991: 180. Estas são as "respostas de orientação" ou interrupções da actividade corrente. A reacção à detecção de eixos de simetria verticais seria um exemplo. O psicólogo O Neumann sugere que as respostas de orientação são o análogo biológico das situações com o lidar de uma situação de emergência numa embarcação com uma ordem de "Toda a gente para o convés!" (DENNETT 1991: 180).
[570] DENNETT 1991: 181.
[571] DAWKINS [1976].
[572] Cf. EDELMAN 1987.

nais da aprendizagem assim possibilitada. Ora, a capacidade de antecipar depende de uma aposta humeana na regularidade dos processos no ambiente que são importantes para os organismos. As instalações de antecipação "referem-se" a essas regularidades. Há no entanto processos que interessam a organismos e que não são processos regulares. Os sistemas que melhor lidarão com esses processos serão organismos capazes de se re-desenharem, i.e. organismos capazes de aprendizagem. Dennett chama a esta possibilidade fixação pós-natal do *design*. A fixação pós-natal do *design* dá-se em virtude de um processo idêntico ao processo "pré-natal" de fixação de *design*: i.e. em virtude de um processo de evolução por selecção, que neste caso acontece no interior. Os selectores têm que ser já fixados, e a eles se apresentarão candidatos à selecção (estruturas cerebrais que controlam comportamentos). Estes processos evolucionistas são hoje objecto de modelizações em computador (de resto, de acordo com Dennett, tais modelos, mais do que simulações de processos cerebrais são realizações num suporte diferente de processos fundamentalmente idênticos [573]).

Cérebros plásticos e capazes de se reorganizarem adaptativamente constituem um novo *medium* para a evolução, que dá lugar a uma aceleração desta. "*Efeito Baldwin*" é o nome para o fenómeno de aceleração: populações de organismos cujo sistema nervoso é dotado de plasticidade detêm uma vantagem sobre os seus parentes "rígidos". Dennett exemplifica o processo com o "Bom Truque", um qualquer talento comportamental útil. Imagine-se uma população na qual existe uma grande variação na instalação (*wiring*) dos cérebros à nascença. Apenas uma das instalações dá ao organismo capacidade para o Bom Truque comportamental, que o favorece. Todas as outras são indiferentes (a situação representa-se por meio de paisagens de aptidão – *fitness landscapes*). No entanto os indivíduos, mesmo os que não nasceram com essa instalação, podem mover-se no espaço do *design* graças à plasticidade e todos tendem a aprender o Bom Truque. Sem a plasticidade, "*a miss is as good as a mile*", é indiferente falhar por pouco ou por muito. Com a plasticidade, a geração seguinte terá mais indivíduos mais próximos do alvo (a instalação para o Bom Truque). O processo continuará até à

[573] Cf. acerca de modelos, MACHUCO ROSA 2002 (Parte II, Sistemas Acentrados, nomeadamente os sub-capítulos: 1. Sistemas Dinâmicos, 2 Fenómenos críticos e redes de *spins*, 3. Redes neuronais artificiais, 4. A Nova Robótica e 4.3.1. O conceito de *hardware* evolutivo).

população ter fixado geneticamente o Bom Truque, que pode assim ser passado de forma relativamente rápida às gerações futuras.

Até mesmo no cérebro de um animal como a rã, por exemplo, a evolução interna é muito mais rápida do que a evolução por selecção natural de organismos: as "gerações" podem durar segundos e não dias, meses ou anos. É no entanto num tipo de controlo poderoso conseguido com ajustamentos da ordem dos milésimos de segundos que os humanos são melhores auto-controladores do que quaisquer outros organismos que também se auto-controlam, mas cuja atenção focalizada é curta e cuja capacidade de planeamento não alcança um longo prazo. Ora, as chaves dessa habilidade dos humanos são, segundo Dennett, a partilha de *software*, a auto-estimulação (nomeadamente a auto-estimulação linguística), e, em geral, as ligações virtuais assim conseguidas e em particular o estabelecimento de um centro de comando do sistema.

Para Dennett, as *ligações virtuais* são a pedra de toque da abordagem da consciência feita a partir de baixo e de dentro e têm um papel muito importante na transformação do problema geral que se coloca aos agentes (o problema "O que é que hei-de fazer a seguir?") no meta-problema "O que é que hei-de pensar a seguir?". Este problema está relacionado com a necessidade de controlo de alto nível do fluxo de informação no sistema de controlo do organismo. Para conceber o controlo, a função executiva, Dennett vai buscar a Oliver Selfridge, o precursor de modelos descentralizados dos processos cognitivos, a ideia de "Pandemónio": «Como o nome dado por Selfridge é tão apto, vou utilizá-lo de forma generalizada no livro para todos os seus descendentes»[574]. Para se chegar à concepção correcta da evolução da consciência, a ideia de *pandemónio* terá que ser conjugada com a ideia de um *controlo centralizado virtual.*

Um terceiro *medium* da evolução intema é constituído pelos memes, coisas não vivas mas que satisfazem a definição de evolução através dos parâmetros da variação, replicação e aptidão (*fitness*) diferencial. A evolução dos memes rege-se ainda pelo princípio segundo o qual replicação não é necessariamente boa para qualquer coisa mas sim "cega". Os três processos referidos (evolução genética, evolução neuronal dada a plasticidade fenotípica e evolução memética) têm velocidades crescentes. A evolução cultural é em

[574] DENNETT 1991: 189. O. Selfridge é um dos primeiros exploradores de redes neuronais e elaborou o modelo do pandemónio para o reconhecimento perceptivo. Cf. BECHTEL & GRAHAM 1998: 489-491 e 771. Cf. ainda BECHTEL, ABRAHAMSEN & GRAHAM 1998:11-13.

grande medida responsável pela instalação da Máquina Virtual da consciência. Este é um ponto que muitos críticos de Dennett consideram inconcebível: Dennett concede um papel tão importante à linguagem no seu modelo de consciência que, sabendo-se que esta é relativamente recente[575] e que o espantoso crescimento do cérebro humano estava concluído antes da existência de linguagem, fica aberta a hipótese de terem existido humanos neurofisiologicamente semelhantes a nós e não conscientes. A hipótese liga-se à recepção particularmente calorosa que Dennett faz da ideia do psicólogo Julian Jaynes acerca da origem da consciência[576] . Se os nossos cérebros não são fundamentalmente diferentes, do ponto de vista anatómico, dos cérebros dos humanos de há 10.000 anos, por hipótese, e se no entanto nós somos mais inteligentes, a nossa maior inteligência deve ser uma questão de *software,* e de partilha de *software* e não de *hardware.* A arquitectura funcional criada por infestações de memes em cérebros humanos não será portanto capturável neurofisiologicamente, mas apenas a um nível de descrição mais abstracto, análogo ao nível do *software* nos computadores. O problema da consciência torna-se então no problema de saber como é que a consciência humana pode ser realizada através da operação de Máquinas Virtuais em cérebros humanos. O que é preciso compreender é o processo de geração e selecção de padrões neuronais e, segundo Dennett, só se conseguirá fazer sentido desse processo a um nível bastante mais abstracto do que o nível neurofisiológico, já que aquilo que é necessário analisar são *os efeitos da linguagem e da auto-estimulação linguística nos cérebros humanos.* Para Dennett, o momento histórico matricial teria sido o momento em que um pedido de informação ou de ajuda (num estádio anterior da linguagem usualmente utilizado em direcção a outros humanos, de modo a despoletar os efeitos desejados, de modo talvez análogo ao que se passa hoje com as vocalizações de certos animais) foi ouvido pelo próprio, na ausência de outros. Este teria sido o início da exploração da auto-estimulação linguística, ligada ao pensamento voluntário, que é custoso e lento relativamente a processamentos cognitivos inconscientes, mas é uma inovação essencial em criaturas em cujos cérebros o acesso interno é menos do que óptimo. Em criaturas assim, são muitas as situações nas quais especialistas cerebrais determinados podem possuir uma informação determinada sem estarem ligados aos especialistas que necessitam dessa informação. A auto-estimulação linguística "serve para" sulcar, nestas circunstâncias, vias de comunicação interna não instaladas (uma capacidade ainda notória nos humanos, por exemplo em casos de recuperação após lesões cerebrais).

[575] Dennett considera por exemplo (DENNETT 1991: 190) que as especializações inatas chomskyanas são aquisições recentes e aproveitamentos de circuitos anteriores, juntamente com o Efeito Baldwin.

[576] JAYNES 1976. Cf. DENNETT 1998e, *J.Jaynes Software Archeology.*

É no contexto definido por este quadro evolucionista que Dennett introduz a seguinte definição de consciência: «A consciência humana é um enorme complexo de memes (ou melhor, de efeitos de memes em cérebros) e a melhor maneira de compreendê-la é vê-la como a operação de uma Máquina Virtual "à maneira de Von Neumann", implementada na arquitectura paralela de um cérebro, que não foi projectado para actividades tais. Os poderes desta Máquina Virtual favorecem vastamente os poderes do *harware* orgânico subjacente no qual corre, mas ao mesmo tempo algumas das suas características mais curiosas, e especialmente limitações, podem ser explicadas como produtos laterais (*byproducts*) dos *kludges*[577] que tornam possível esta curiosa mas eficaz reutilização de um orgão para um novo propósito»[578]. Aparentemente não se poderia estar mais longe da filosofia. J. Searle considera esta teoria uma versão da IA Forte que passa totalmente ao lado do problema filosófico da consciência[579].

Vários dos conceitos utilizados na definição anterior só ficaram disponíveis para ajudar a pensar sobre o pensamento graças às máquinas computacionais: por exemplo a Máquina Virtual como estrutura temporária[580] e a serialidade (o efeito "gargalo de garrafa" da arquitectura de Von Neumann, *design* do qual todos os computadores digitais são descendentes directos)[581]. A ideia de Máquina Virtual, ou máquina que se comporta como outra máquina remonta à ideia de Máquina de Turing. Para Dennett, no entanto, a ideia de Máquina de Turing é demasiado rigorosa para ser um modelo

[577] O termo *kludges* é um termo utilizado por *hackers* para referir maneiras *ad hoc* de pôr uma coisa a funcionar melhor.

[578] DENNETT 1991: 210.

[579] Cf. SEARLE 1997: 97-115.

[580] Não é apenas em cérebros que existe plasticidade, mas também em computadores. Estes também não estão completamente "fixados" desde o início, podem implementar programas, receber dados, armazenar novo *software* na memória. São no entanto tipos diferentes de plasticidade.

[581] DENNETT 1991: 214. Dennett comenta que os computadores deveriam ter sido chamados *mentes* electrónicas e não *cérebros* electrónicos, pois os computadores comuns são seriais como as mentes conscientes e não paralelos como os cérebros.

realista para a vida mental humana. Ela é "substituída" em CE pela ideia de *Máquina Joyceana*. O conceito de Máquina Joyceana não é obviamente um conceito técnico como o conceito de Máquina de Turing, mas uma forma de chamar a atenção para as marcas da implementação de Máquinas Virtuais em mecanismos resultantes de evolução biológica. A ideia original de Turing (que afinal fazia lógica matemática quando criou o conceito de Máquina de Turing e procurava formalizar a noção de algoritmo e não pensar uma máquina real e muito menos um cérebro) é biologicamente irrealista. Utilizada fora do seu contexto, como modelo psicológico, nomeadamente pelo funcionalismo, produz uma imagem intelectualista da mentalidade. O fluxo de consciência dos humanos não corresponde ao esquema de Turing: não é um processo serial num espaço de trabalho restringido, tratando uma célula de cada vez, um processo para o qual dados e instruções são trazidos a partir de uma memória inerte e absolutamente confiável (a "fita") sendo então cumpridas operações. O irrealismo da Máquina de Turing como modelo psicológico não obsta a que existam nela dois pontos importantes e definitivos: (1) a ideia de que passos mecânicos podem constituir todos os processos inteligentes e (2) a ideia de que uma Máquina Universal se pode tornar virtualmente outra máquina. As Máquinas Virtuais são precisamente «conjuntos temporários de regularidades altamente estruturadas impostas ao *hardware* subjacente por um programa; uma receita estruturada de centenas de milhares de instruções que dão ao *hardware* um enorme e entrelaçado conjunto de hábitos e disposições reactivas»[582]. Embora a fenomenologia humana não se identifique com a proposta de Turing, Dennett propõe que a consciência é uma Máquina virtual mas uma máquina virtual joyceana, que não está desde o início instalada no sistema e que por isso mesmo vem a "aperceber" o seu próprio funcionamento. De que forma se passa isso?

Nós temos acesso a um fluxo fenomenológico que parece serial. O problema que se coloca é conceber a forma como uma máquina paralela (o cérebro) pode vir a implementar uma Máquina Virtual serial (a consciência). A diferença entre arquitecturas seriais e paralelas não é necessariamente muito relevante do ponto de vista teórico quando se trata de encarar o problema, já que a ideia de Máquina de Turing Universal serve precisamente para estabelecer a pos-

[582] DENNETT 1991: 216.

sibilidade do encaixe de máquinas virtuais sucessivas, por hipótese com arquitecturas diferentes, cada uma das quais poderá ser reconhecida por meio de um determinado *interface*. A essa "apresentação" chama-se a *ilusão do utilizador*. Olhando para o detalhe das instruções que constituem os programas ou máquinas virtuais, é possível "não ver nada". No entanto, a um nível mais abstracto, surgem entidades virtuais perfeitamente determinadas.

O problema consiste, como se disse, em saber como será possível obter uma "máquina de uma coisa de cada vez" (uma máquina serial) a partir de uma "máquina de muitas coisas ao mesmo tempo" (uma máquina paralela). A proposta de Dennett é a seguinte: do mesmo modo que é possível simular estruturas paralelas em máquinas seriais, o que foi reconhecido desde o início da era dos computadores[583], é possível simular uma máquina serial, uma Máquina de Von Neumann, numa estrutura paralela. De facto, «as mentes conscientes são máquinas virtuais mais ou menos seriais ineficientemente implementadas no *hardware* paralelo com que a natureza nos dotou»[584]. O que seriam então os "programas" que se instalam nas redes neuronais (cerebrais)? Dada a plasticidade, poderão ser instalados micro-hábitos (conexões entre neurónios), que constituirão macro-hábitos para o *hardware* e regularidades no comportamento da entidade. O problema seguinte é saber como são instalados esses programas. Se nas máquinas artificiais um programa pode ser instalado fazendo um *download* do programa para a memória (e assim o computador ganha imediatamente "novos hábitos cognitivos"[585]), num cérebro é necessário treino, nomeadamente auto-estimulação repetitiva, para a instalação de novos hábitos cognitivos. Se a CPU de um computador é rígida no tratamento de instruções, formuladas numa linguagem da máquina que é fixa, não há no entanto razões, no que toca a cérebros, para pensar que dois cérebros diferentes terão o mesmo sistema de interconexões (este é de novo o problema da Linguagem de Pensamento, o problema de Fodor, tratado no capítulo 2). Assim, se dois cérebros

[583] Cf. McCULLOCH & PITTS 1943, e o Perceptron de F. Rosenblatt (BECHTEL, ABRAHAMSEN & GRAHAM 1998: 32-33).

[584] O problema da velocidade (acerca do qual a prova de Turing nada diz) é importante para Dennett aqui e também na refutação do Quarto Chinês, cf DENNETT 1987l, *Fast Thinking*.

[585] Que a neurociência computacional futura venha a poder fazer o mesmo em humanos é uma hipótese muito presente na ficção científica cinematográfica. Recorde-se de novo filmes como *Existenz* (Cronenberg) ou *Matrix* (Irmãos Wachowsky). Neste último é por exemplo feito o *download* de um programa para "saber guiar um helicóptero" na personagem Trinity.

vão partilhar *software*, e a aposta da teoria da consciência de Dennett é afirmar que vão, não poderão fazê-lo através de um processo tão simples como o *download* de um mesmo programa nas memórias de ambos. Aliás, a própria separação estrutural entre memória e CPU também não tem análogo em cérebros. É o próprio Dennett quem aponta todas estas diferenças, considerando no entanto que a analogia funcionalista entre consciência e *software*, mesmo imperfeita, continua incontornável por várias razões e nomeadamente porque: (1) a consciência é demasiado recente em termos evolutivos para ser uma "instalação" inata; (2) efeitos "culturais", exteriormente disponibilizados, são muito importantes na constituição de um fluxo de consciência de tipo humano; (3) os traços funcionalmente importantes para a consciência são invisíveis ao escrutínio neuroanatómico já que resultam de inúmeras micro-instalações na plasticidade do cérebro; (4) a ideia de ilusão do utilizador de uma Máquina Virtual é boa para conceber a consciência: o nosso acesso consciente ao que se passa no nosso cérebro[586] é nulo.

3.3.5 *A ilusão do utilizador da Máquina Virtual joyceana.*

A ideia segundo a qual a consciência é uma ilusão do utilizador de uma Máquina Virtual instalada num cérebro, o cérebro do corpo a que chamamos nosso, é um ponto do MEM difícil de assimilar. É o ponto no qual cada um de nós é obrigado a perguntar se ser consciente poderá de facto consistir numa *ilusão*. Quando se fala de ilusão do utilizador relativamente a um dado programa, "cá fora" existe alguém que reconhece o interface num computador, i.e. existe um *utilizador* propriamente dito, uma pessoa. O problema com a transposição da ideia é, obviamente, saber para benefício de quem se daria a ilusão do utilizador. Ora, se o lugar de origem da ideia de ilusão do utilizador de uma Máquina Virtual é a interacção de uma pessoa com um computador (o particular programa ao qual corresponde uma determinada ilusão do utilizador, recorde-se, é a Máquina Virtual por oposição à máquina física) o que poderá significar a ideia de que a consciência corresponde à ilusão do utilizador de uma Máquina Virtual?

Antes de mais, a ilusão do utilizador de uma Máquina Virtual alude ao facto de o utilizador não poder dizer como é, a nível de

[586] Cérebro? Ou outra coisa qualquer? Como o saberíamos?

hardware, a particular Máquina Virtual com que lida, como está implementada ou a quantos níveis de distância do *hardware* se encontra. A aplicação da ideia a nós próprios (ou a qualquer ser consciente) produz o seguinte quadro: afirmar que pensar que eu sou consciente é uma ilusão do utilizador de uma Máquina Virtual é afirmar que eu não sei, ao ser consciente, como está essa máquina implementada no cérebro nem a quantos níveis de "distância" (por exemplo do nível dos neurónios – ou outro *hardware* qualquer). O problema maior é que de certo modo, sendo essa ilusão (com as suas características de controlo, de serialidade, etc) uma ilusão, ela "in-existe", e com ela toda a minha vida consciente (e eu com ela). O que podem tais ideias significar?

Para a ideia de Máquina Virtual (e a correspondente ideia de interface) ser legítima, o desdobramento entre utilizador e "apresentação" que existe na interacção que acontece "cá fora" não pode deixar de estar presente. Dennett admite-o[587]. O que Dennett nega é que o utilizador tenha que ser um observador consciente, um agente cognitivo global; ele propõe que o "observador" está incorporado no próprio sistema, sendo constituído por «pedaços de interface cujos vários formatos continuam a constranger os tipos de questões que podem ser respondidas constrangendo assim os conteúdos que podem ser expressos»[588]. A ideia é portanto que o interface é constituído pelo formato das interacções entre partes do sistema.

Afirmou-se atrás que os humanos partilham *software* por aprendizagem, imitação e reforço. Cada um instala hábitos mentais organizados e pré-testados. Além do mais, o cérebro faz *load-off* de muitos dos dados necessários à prática desses hábitos mentais em *buffers* do ambiente (o que significa que o lugar e os materiais do pensamento humano não são totalmente interiores: são muitos os humanos que não conseguem, por exemplo, executar cálculos complicados nem pensamentos sofisticados sem o apoio de símbolos escritos e diagramatizações). Escrita é um bom nome (de resto já utilizado...) para essa exteriorização e despossessão do mental. Mas o que no âmbito do MEM se sugere é que essa exteriorização foi "interiorizada" de uma forma que foi importante para a consciência humana poder existir. Nomeadamente, um dos hábitos que os humanos adquiriram foi o hábito de falarem consigo próprios, i.e. de utiliza-

[587] DENNETT 1991: 312 «How could there be a User Illusion without this separation? There couldn't be».

[588] DENNETT 1991: 312.

rem línguas naturais para a estruturação e controlo das suas vidas mentais. É certo que capacidades como a auto-exortação à recordação parecem essenciais no modo de vida humano e impossíveis sem a consciência, sem uma capacidade geral de supervisão e de controlo de sistemas cognitivos que permite sondar, encontrar defeitos no sistema e corrigi-los, reunir ou centralizar recursos separados. Ora, segundo Dennett, subjacente a todas estas habilidades está uma capacidade básica de tratar o meta-problema "o que pensar a seguir?" com apoio na linguagem. Como se afirmou atrás, este problema é o problema sucessor do problema "o que fazer a seguir?" que se coloca aos organismos. O recrutamento geral em situação de emergência (nas respostas de orientação vistas atrás) não é suficiente para o considerar resolvido, pode até dar lugar a excesso e caos. É necessário um controlo central e uma recuperação de conhecimentos bem dirigida. Dennett chama a este problema o problema dos pássaros de Platão[589], transposto para o cérebro. As respostas de orientação atrás referidas precisam de "quem" as organize. É preciso um controlo global dos especialistas separados: «Aprender a raciocinar é aprender estratégias de recuperação de conhecimentos para a superfície mental»[590] e é aqui que aparece a necessidade de hábitos mentais, cuja vantagem consiste em moldar "passagens de pensamento"[591]. A veia ryleana de Dennett revela-se de novo no tratamento desta questão. Nomeadamente, G. Ryle sugeriu (e antes de Ryle, um grande número de filósofos) que pensar é falar consigo próprio[592]. Trata-se então de auto-estimulação. Técnicas de auto-estimulação são essenciais para o pensamento sofisticado, nomeadamente para o raciocínio explícito para o raciocinador. Por exemplo, a nossa memória não é uma memória RAM, não vamos nela onde queremos e quando queremos, de forma fiável. Daí, a necessidade de maneiras *ad hoc* de pôr as coisas a funcionar, hábitos individuais de auto-estimulação como ensaios, mnemónicas, ritmos, sons, gestos, coçar a cabeça, franzir o sobrolho, etc. Trata-se de sustentadores do processo de pensamento, marcas, ancoragens talvez não tão aleatórias quanto parecem. São afinal as marcas da adopção idissincrática da Máquina Virtual e da imposição de uma disciplina no cérebro individual em causa. Sendo estas as condições da instalação de uma Máquina Virtual num cérebro humano, justifica-se que Dennett chame a esta máquina uma Máquina "Joyceana".

[589] Cf. PLATÃO, *Teeteto*, 197-198 a, onde o problema é saber se o conhecimento é qualquer coisa que se pode possuir sem tê-lo junto a nós. Porque não basta ter os pássaros, é preciso que o pássaro certo venha quando é chamado...

[590] DENNETT 1991: 223.

[591] Como Dennett nota, até mesmo um humano contemporâneo, um pensador sofisticado não consegue manter de forma ilimitada a atenção presa ao problema que trabalha. No entanto, somos capazes de lidar com problemas que demoram muito mais do que esse tempo curto a serem resolvidos. Daí a necessidade de estratégias como recapitulações e descrições para nós próprios, ensaios após ensaios. O efeito é lançar para a memória, sulcar, efectuar uma *Bahnung* propositada, utilizando o termo de Freud.

[592] RYLE 1979.

A grande questão a colocar é obviamente a seguinte: porque é que os mecanismos descritos haveriam de ser conscientes? E a resposta de Dennett é que a máquina não estava já pronta, instalada de forma maximamente eficiente desde o início, e que o funcionamento da Máquina Joyceana é tão perceptível como qualquer das coisas no mundo que ela foi projectada para perceber. Como nota também por exemplo M. Minsky, perceber "dentro" não é, nestas circunstâncias, mais difícil do que (ou diferente por natureza de) perceber "fora".

É com este "jogo de espelhos" que a parte construtiva da teoria da consciência de CE culmina. É, assim, um erro pensar que o modelo de CE, apesar de pretender eliminar o Teatro Cartesiano, elimina completamente a *percepção interior*: a ideia de uma Máquina Virtual instalada no cérebro e que apercebe o seu próprio funcionamento é uma ideia acerca de percepção interior. A metade construtiva da teoria da consciência de CE é, assim, além de um *modelo funcionalista*, uma *teoria da percepção interior*, que supõe um desdobramento entre audiência e "palco", mesmo que o teatro seja um teatro impreciso, não envolvendo distinção absoluta entre consciência e não-consciência. Só a percepção interior permite compreender a ideia de ilusão do utilizador. Esta por sua vez "dá" como características da consciência determinados traços (como o controlo, a serialidade e a continuidade) que não se encontram nos processos neuronais subjacentes.

3.3.5.1 *Ainda assim um interior: querer-dizer*[593]*, pandemónio e actos de fala.*

«Whatever we may want to say, we probably won't say just that»,
Marvin Minsky[594]

[593] Utilizar-se-á esta forma verbal como tradução do verbo inglês *to mean*.

[594] MINSKY 1985: 236.

A grande possibilitadora do tipo humano de mentes (e para Dennett do único tipo de consciência no pleno sentido do termo que conhecemos) é a linguagem natural. Nietzsche defendeu na *Gaia Ciência*[595] que os humanos podem pensar e agir sem ter consciência e que *a consciência se desenvolveu sob a pressão da necessidade de comunicação*. O desenvolvimento da linguagem e da consciência andariam assim a par. Em CE Dennett defende algo de análogo. De acordo com o que foi dito atrás, é *software* no cérebro que cria aquilo a que Dennett chama um milagre político interno, i.e que propicia o aparecimento de um comandante dos especialistas ou agentes sub-pessoais. De um ponto de vista exterior e global sabemos que controlo e decisão acontecem nos e pelos sistemas cognitivos. O problema é conceber a forma como Máquinas Virtuais instaladas em cérebros criam *centros* que à partida não estão lá e aos quais está associado o *controlo executivo do comportamento*. Um dos aspectos do encerramento do Teatro Cartesiano é o fim do Intencionador Central, o centro do qual o controlo das intenções na acção, inclusive as intenções semânticas, proviria. Qual é então, em geral, a fonte do controlo? Qual é, em particular, a fonte dos actos de fala? Esta questão constitui o objecto da secção *How Words Do Things With Us* de CE.

Recorde-se as identificações feitas por Shakey, referidas no ponto 3.3.1 do presente capítulo. Elas eram "comunicadas ao exterior" de uma forma semelhante à forma como um PC "utiliza" a linguagem quando nos pergunta solicitamente "tem a certeza que quer pôr isso no lixo?". Otto, o crítico do MEM, diria que é ilegítimo conceber o uso humano de linguagem à imagem dos relatos de Shakey ou das "perguntas" "endereçadas" pelo PC ao seu utilizador: quando humanos falam, eles querem-dizer (*mean*) aquilo que dizem. Quando humanos formulam frases, eles moldam-nas de modo a exprimir com acuidade aquilo que querem-dizer. Isto seria conseguido prestando atenção à experiência consciente própria e ajuízando acerca das palavras que mais justiça fazem a essa experiência. Nada disso se passa com Shakey ou com o PC: eles não têm interior nenhum, o input e o output estão ligados da maneira errada, não "fica" um observador no meio que possa querer-dizer antes de os relatos serem produzidos, não existe um pensamento determinado à espera de ser posto em palavras. Para Otto, Shakey é um *zombie* e da mesma maneira que constitui um erro conceber as imagens mentais à imagem das "imagens" de Shakey, é um erro conceber o querer-dizer à imagem da utilização de linguagem por Shakey. Pelo contrário, a primeira coisa que o uso de linguagem por humanos revela é que quando um humano fala, ele sabe o que quer-dizer, trata-se de pôr em público a vida consciente, que está antes ou atrás, e cuja expressão pode aliás ser mais ou menos perfeitamente conseguida. Como sempre, a ideia de Dennett é que o parecer-ser pode e deve ser tomado como dado mas não sustenta qualquer pretensão teórica acerca do que se passa. A produção da fala é um problema sério para a linguística, que aliás, de forma reveladora, o trata menos do que trata a análise da fala. Encarar a produção da fala supõe que se lide com a questão da iniciação e do controlo, e, logo, com o problema do Intencionador Central.

[595] NIETZSCHE [1882].

Em CE Dennett trabalha sobre o modelo do psicolinguista holandês Willem Levelt[596]. O modelo de Levelt prevê um lugar para o "Conceptualizador", que cria o que o sistema vai dizer e decide que isso seja dito. De acordo com o modelo de Levelt, a fala não é produzida palavra a palavra: existe uma certa previsão global do que vai ser dito (que existe uma certa previsão é visível por exemplo no fenómeno da prosódia, i.e. na entoação e distribuição desta pela frase). Vários erros de fala como *spoonerismos*[597] são sugestivos do modo como é feita a selecção de palavras. Aliás, erros de fala só podem ser considerados erros se puderem de alguma forma ser considerados diferentes daquilo que o falante queria-dizer. Se tais fenómenos ocorrem é porque existe uma determinação do querer-dizer, apesar das interferências possíveis. Essa determinação é tarefa do "conceptualizador", admitido por Levelt como uma reificação provisória que precisa de ser explicada. O conceptualizador decide pela performance de um acto de fala, envia uma ordem para a burocracia sob o seu controlo, a qual inclui as partes do modelo que são o formulador e o articulador, bem como as Relações Públicas que escolhem as palavras apropriadas, as tornam gramaticais, combinam o tom de voz e executam o acto.

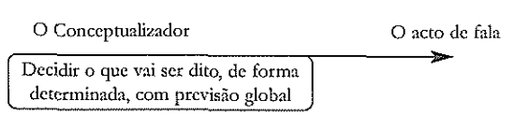


FIG. 4 O Conceptualizador

[596] LEVELT 1989.

[597] "Spoonerismos" são trocas de ordem nos componentes das palavras produzindo palavras inexistentes mas que deixam perfeitamente perceber a "intenção" que lhes deu origem.

É evidentemente problemático dizer em que linguagem é dada a ordem do conceptualizador. Para Levelt, a ordem (apenas para actos de fala e não para todas as actividades cognitivas) é dada numa Linguagem do Pensamento. Este modelo de originação da fala corresponde à ideia de processamento através de uma hierarquia de rotinas, agentes e sub-agentes. Em alternativa Dennett propõe que a produção dos actos de fala deve ser concebida a partir de um processo de pandemónio[598] e que o eventual emparelhamento de conteúdos com expressões é o resultado final de competições, do fazer e desfazer de coligações. Tudo começa com modos de fazer ruído vocal, sem razão. Este ruído excita agentes, (na metáfora de Dennett pedaços de palavras a dizer "deixa-me ser eu!") que "procurarão" modulá-lo. O resultado é incompreensível (mas é já, por exemplo, inglês-incompreensível). Outros agentes moldam em palavras o fluxo de inglês-incompreensível. Várias frases-candidatas emergem, uma será pronunciada, a vencedora, mas logo a seguir a disputa será retomada e o falante ficará a ponderar sobre aquilo que deveria ter dito. No entanto, a pessoa que produz a enunciação asseverará que disse aquilo que queria-dizer. De acordo com o modelo do pandemónio dizer aquilo que se quer-dizer não envolve a existência de um Intencionador Central. A "apropriação" de uma frase pelo falante acontece num cenário de geração por processamento paralelo e de trabalho de agentes anónimos: essa é a fonte do conteúdo, não uma ordem ditada *ex nihilo* e *ex cathedra* pelo Intencionador Central. A ideia de Dennett quanto à produção da fala retoma a história de abertura de CE, o "jogo de sociedade chamado psicanálise"[599]. O jogo tem como resultado uma história, e no entanto não existe um autor da história, apenas um processo do qual a história emerge sob o questionamento incessante de questionadores. A ideia segundo a qual o pensador de um pensamento não parte para a expressão desse pensamento com uma intenção determinada e o Intencionador Central deve ser substituído por um

[598] O termo "demónio" é utilizado no contexto da teoria cognitiva com o significado de agente de reconhecimento, intervindo em circunstâncias determinadas.

[599] DENNETT 1991: 10-16. O objectivo do jogo é reconstruir um sonho que supostamente existiu. Diz-se a um jogador que quando ele saiu da sala, alguém contou o sonho aos restantes participantes. Ele deve então fazer perguntas acerca do sonho. Obtém depois respostas (sim ou não) aleatórias para as suas questões, respostas que são geradas por uma regra como "Responder sim se a última letra da última palavra da pergunta fizer parte da primeira metade do alfabeto e não em caso contrário". O jogador vai assim "reconstruindo" uma história que não existiu e tenta através dela identificar o sonhador de tal sonho.

processo evolutivo de geração provem da investigação em Inteligência Artificial, nomeadamente do trabalho de téoricos como M. Minsky[600] e D. Hofstadter[601]. Como diz M. Minsky na citação em epígrafe, o que quer que queiramos-dizer, provavelmente não queremos-dizer exactamente isso, i.e. não existe uma determinação prévia concebida do querer-dizer.

Um argumento a favor do modelo de produção por pandemónio é o facto de intenções não concordantes de realizar actos de fala poderem (co) existir num mesmo indivíduo. Como prova disso mesmo, Dennett refere experiências realizadas pelo psicólogo inglês Anthony Marcel nas quais são dadas instruções a um paciente com visão cega para realizar um determinado acto de fala (dizer quando vê um *flash* de luz) através de três acções determinadas (dizer "sim", premir um botão, piscar afirmativamente)[602]. Ora, frequentemente, as três acções do paciente não vão no mesmo sentido. Várias patologias neuropsicológicas parecem também confirmar uma tal "ramificação" de intenções: na afasia de jargão[603] os pacientes (cuja inteligência é normal) produzem performances verbais gramaticais mas totalmente descontroladas, enquanto que pacientes com síndroma de Korsakoff[604] (que acompanha casos graves de alcoolismo) produzem uma efabulação de experiências, com toda aparência de um discurso normal, acreditada pelo efabulador, mas que nada tem de verídico. Estes são casos de anosognosia, uma vez que os pacientes não têm qualquer noção das anomalias, e casos de discurso produzindo-se "sem dono", praticamente independente da pessoa que o pronuncia, que não seriam de esperar em sistemas controlados e hierárquicos de produção de fala. Pelo contrário, milhares de demónios de palavras podem precisamente fazer com que pareça existir um Conceptualizador num processo em que os fins procuram materiais e os materiais "procuram oportunidades". De acordo com o modelo proposto, tanto pode ser o conteúdo-a-ser-expresso a ser ajustado como os candidatos à expressão verbal. Note-se que o modelo do Pandemónio torna insustentável a linha do apercebimento proposta em C&C. Esta era acompanhada pela ideia segundo a qual erros em relatos da experiência própria

[600] MINSKY 1985.

[601] HOFSTDTER 1995 (embora a data deste livro seja posterior à data de CE, o trabalho relatado não o é).

[602] DENNETT 1991: 248.

[603] DENNETT 1991: 249.

[604] DENNETT 1991: 249.

teriam que situar-se de um lado ou outro da linha, que seria logica e funcionalmente clara (mesmo que fosse neurologicamente indiscernível). Dennett considera agora que a ideia é insustentável quando se trata da produção de fala (e do controlo da acção em geral): embora deva haver um *standard* para a correcção de erros na produção da fala, aquilo que funciona como *standard* não tem que manter-se idêntico nem sequer durante a produção de um único acto de fala. A conclusão geral da abordagem da produção da fala resume-se através de uma ideia sempre reiterada de Dennett: frequentemente só descobrimos o que queremos dizer dizendo-o. Isto contradiz a existência de uma intenção prévia ao acto de fala e afasta a ideia do Componente Controlo claramente demarcado do modelo de BS. *O modelo da produção por Pandemónio é evidentemente generalizável às acções intencionais em geral.* É, assim, evocando a confluência de trabalhos empíricos sobre processamento paralelo e distribuído e sobre agentes especialistas semi-independentes que Dennett defende a inexistência de um fluxo de consciência único e definitivo[605]. A singularidade da consciência é substituída por «canais múltiplos nos quais circuitos especialistas em pandemónios paralelos fazem as suas várias coisas, criando esboços múltiplos à medida que avançam. A maioria destes esboços de narrativa tem vida curta na modulação da actividade corrente do organismo mas alguns são promovidos a outros papéis funcionais, em rapida sucessão, pela Máquina Virtual do cérebro (...) O resultado só não é uma confusão porque as tendências impostas em toda esta actividade são elas próprias resultantes de *design*, algum inato e partilhado com outros animais»[606]. O modelo proposto pode ser esquematizado em traços largos da seguinte maneira:

[605] O MEM como exploração desta ideia conjuga várias aportações: a ideia segundo a qual a consciência é conseguida por uma sociedade de especialistas distribuídos, equipada com uma memória de trabalho, chamada espaço de trabalho global, cujos conteúdos podem ser emitidos para o sistema como um todo é, como se viu no início do presente capítulo, a ideia de B.Baars. Baars sugere ainda que é a própria lentidão da actividade mental consciente que sustenta a hipótese de que o cérebro não seria para ser serial. A ideia de esboços múltiplos é comum em IA bem como as estruturas de dados projectadas para representar sequências temporais. Na filosofia Gareth Evans é responsável pelo desenvolvimento de ideias paralelas (EVANS 1982). Na neurobiologia W.Calvin pretendeu o mesmo com a máquina darwinista. A ideia de consciência como Máquina Virtual há muito tempo que circula, segundo Dennett. A ideia de especialistas básicos que constituem a mente tem uma importância óbvia no estudo das mentes animais, que são por vezes definidas como "*quilts*" de IRMs (*innate release mechanisms*) e FAPSs (*fixed action patterns*). É partilhada por exemplo por M. Gazzaniga que considera que o estudo dos défices neuropsicológicos revela que a mente é um feixe de agências semi-independentes, por Fodor, com a sua defesa da modularidade (FODOR 1983), por Minsky com os agentes em *The Society of Mind* (MINSKY 1985), etc.

[606] DENNETT 1991: 253-254.

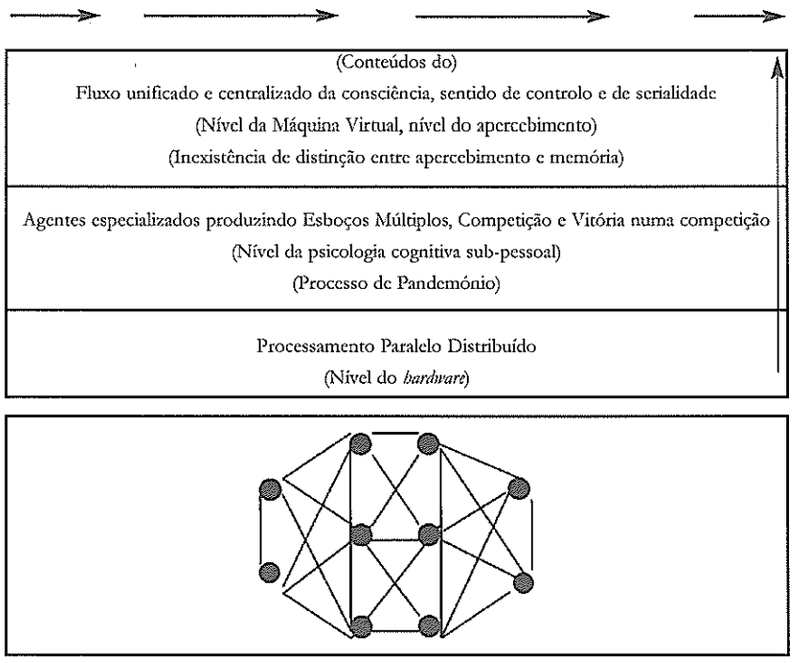


FIG 5. O modelo funcionalista de CE

3.3.5.2 *O Eu*[607].

Tweedledum: You're only one of the things in his dream!.
You know very well you are not real.
I am real! said Alice and began to cry[608].
Lewis Carroll, *Through the Looking Glass*

Afirmou-se no início deste capitulo que os críticos de Dennett o acusam de rebater a consciência sobre noções funcionais tais como o acesso, o eu e estados mentais de segunda ordem. Ora um componente essencial do modelo de consciência apresentado em CE é precisamente uma teoria do Eu. A ideia básica é que a existência de

[607] Esta questão será retomada no próximo capítulo, especificada então relativamente à identidade pessoal. Cf. MINSKY 1985, Capítulo 4 (*The Self*), Capítulo 5 (*Individuality*), Capítulo 6 (*Insight and Introspection*), e Capítulo 22 (*Expression*).

[608] CARROLL 1995: 116.

um Eu num sistema cognitivo físico resulta de uma passagem de agentes especialistas que "falam entre si" ao nível sub-pessoal para uma entidade que "fala consigo própria", que se toma a si própria como unificada e autora das acções de um corpo. De acordo com Dennett, uma tal passagem envolve a instituição de um centro de controlo virtual a partir de uma (auto) representação. Essa auto-representação serve, na metáfora que Dennett vai buscar à física, como "centro de gravidade" da narrativa biográfica produzida linguisticamente pelo sistema.

O Eu é posterior ao *self*, o sentido-de-si existente mesmo em sistemas biológicos não linguísticos, que acompanha a fronteira entre si e não-si instaurada por uma entidade viva. Saber que coisa no mundo se é, é indispensável a todo o agente ou proto-agente e permitir um tal "conhecimento" é a função do *self*. Os *selves* correspondem portanto a limites-de-si reconhecidos e mantidos por uma organização, embora não sejam coisas nem lugares. Os limites-de-si são no entanto plásticos, extensíveis e retraíveis nas criaturas biológicas e o Eu humano é precisamente uma extensão linguística e narrativa do *self*. A narrativa biográfica constitui, tanto como os *selves* estendidos dos outros animais (por exemplo as carapaças dos caranguejos ou as barragens dos castores), uma auto-protecção. Ela surge "como se" saísse de uma só fonte, que *seria* o Eu. No entanto, a proveniência e o correspondente sentido de autoria são de alguma forma ilusórios: o Eu é produto e não fonte da narrativa. Mais precisamente, o Eu é um "centro de gravidade narrativo", constituído na e pela narrativa, ela própria produzida de acordo com o modelo do pandemónio atrás referido. Na situação em que uma pessoa diz por exemplo "Eu sou o João" ou "Eu sou a Alice" esses efeitos não são causados por um autor com intenções. Dennett utiliza a imagem ilustrativa de uma máquina que escreve romances. Imagine-se que um romance está a ser escrito pela referida máquina e que nesse romance existe uma personagem que diz "Call me Gilbert" (como acontece no *Moby Dick* de H. Melville, quando Ishmael diz "Call me Ishmael"). A pergunta a fazer aqui é: "quem fala?", ou "o que é isso que fala?". Dennett sugere que quem fala é Gilbert, uma personagem criada através de uma narrativa ficcional involuntária a partir de actividades não inteligentes e não mentais. Ora, o que acontece com os humanos é exactamente análogo: também as narrativas biográficas dos humanos, cuja personagem central é o Eu, são feitas existir por processos não cons-

cientes e não inteligentes. Quanto à metáfora do centro de gravidade narrativo, ela significa o seguinte: o centro de gravidade de um corpo é um ponto imaginário, um exemplo daquilo a que H. Reichenbach chama um *abstractum*, não uma coisa real a mais para além das partículas que constituem o corpo. Como *abstractum*, ele não está sujeito a constrangimentos físicos, por exemplo constrangimentos à deslocação: o seu movimento pode por exemplo ser descontínuo. Mas ao contrário do centro de gravidade de um objecto, cuja única propriedade é precisamente uma posição espaço-temporal, um Eu não tem uma posição espaço-temporal, a não ser grosseiramente definida. O que o Eu e o centro de gravidade têm em comum é o facto de serem ficção úteis e operantes e também de certa forma *indeterminados*, como qualquer entidade ficcional. Se se perguntar relativamente a uma personagem ficcional "naquele dia 19 de Outubro em que o crime aconteceu, antes de partir para a investigação, Sherlock Holmes tomou ou não tomou o pequeno almoço?" e se nada acerca disto está mencionado no romance, não há, nem tem que haver resposta para tal questão; ora uma indeterminação análoga é característica da identidade pessoal humana[609]. É a indeterminação que distingue os *abstracta* de outro tipo de objectos teóricos, a que Reichenbach chamou *illata*, entidades inferidas, como átomos ou neutrinos. A indeterminação advem do facto de os *abstracta* só possuírem exactamente as propriedades que a teoria em que figuram lhes atribui. Além disso, e mostrá-lo é o propósito da imagem da "máquina que escreve romances", uma ficção não tem que ser criada por uma entidade não ficcional, agindo intencionalmente, pode ser criada por um dispositivo que não sabe o que faz, como um cérebro humano.

Os seres humanos são assim auto-intérpretes automáticos, produtores e produtos de uma narrativa em constante reescrita, que pode inclusivamente conter inconsistências. No caso normal, da produção narrativa emerge um único Eu. No entanto não é possível afirmar que deve necessariamente existir um único Eu por ser humano. Segundo Dennett, não é mais bizarra a existência de vários Eus do que a existência de um só Eu[610]e, de resto, nem a mais

[609] Cf. por exemplo PARFIT 1984. Esta ideia só pode evidentemente ser compreendida no âmbito de uma *visão reducionista da identidade pessoal*, segundo a qual segundo a qual não existe alguma entidade prévia à efectuação da identidade pessoal por algum tipo de continuidade (psicológica, por exemplo).

[610] DENNETT & HUMPHREY 1989.

unificada das pessoas é totalmente auto-possuída e una. É certo que as várias desunificações do Eu que foram sendo exploradas pelos teóricos da mente (desde o cérebro dividido[611] até à Desordem de Personalidade Múltipla[612], para não falar dos casos mais clássicos de esquizofrenia ou paranóia, de *akrasia* e de auto-engano) colocam, como Nagel notou, o problema de decidir o que fazer com as funções morais e metafísicas associadas ao conceito tradicional de um Eu unificado associado a cada pessoa, uma vez que elas parecem ficar de algum modo estilhaçadas e impedidas por estas considerações funcionais[613]. Mas do ponto de vista cognitivo o que importa é que o Eu a quem é atribuída a autoria da vida mental e o controlo das intenções é apenas uma representação de unidade global do sistema, resultante da vitória numa competição e não uma "unidade prévia" a essa competição ou de uma outra ordem. A questão do Eu enquadra-se do seguinte modo no modelo acima proposto (o esquema focaliza o nível sub-pessoal e o nível do apercebimento):

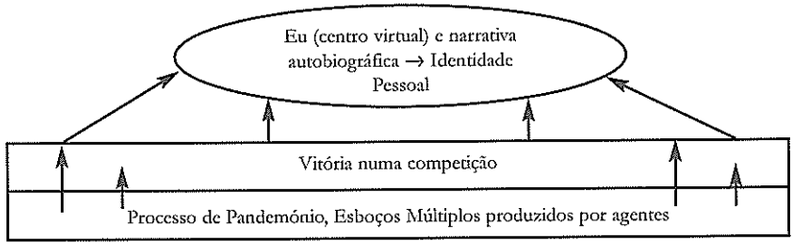


FIG. 6 O Eu

[611] Chama-se cérebro dividido (*split-brain*) a fenómenos provocados pelo corte do corpo caloso (comissurotomia), uma prática dirigida nomeadamente ao controlo da epilepsia. Esses fenómenos consistem naquilo que parece ser uma consciência (dos campos visual e auditivo) "separada" nas metades esquerda e direita do cérebro. Esta separação é provocada artificialmente, "segregando" o input para os dois hemisférios cerebrais. Os indivíduos têm de resto um controlo normal do seu comportamento.

[612] Esta condição, chamada *Dissociative Identity Disorder* desde 1994, está prevista no DSM (Diagnostic and Statistical Manual da American Psychiatric Association) que orienta os clínicos e consiste basicamente na "ocupação" ou "dominação" de um mesmo corpo físico, de forma alternada, por personalidades distintas. A personalidade dominante num particular período de tempo determina o comportamento. A etiologia desta desordem de personalidade é normalmente um acontecimento traumático na infância (por exemplo abuso sexual).

[613] NAGEL 1979d, *Brain Bisection and the Unity of Consciousness*. Nagel considerou o problema relativamente aos casos de cérebro dividido (*split brain*).

3.3.6 *A última palavra (filosófica) quanto ao Teatro Cartesiano.*

A última palavra de Dennett quanto à consciência – a desqualificação dos *qualia*, o encerramento do Teatro Cartesiano e a incoerência da noção de *zombie* – decorre naturalmente do modelo apresentado em CE. Este pretende ser suficiente para justificar o estatuto de todos os fenómenos mentais e nomeadamente daqueles que aparentemente seriam fenomenais ou que decorreriam de uma originação interior com apresentação e apreciação, como as imagens mentais, a experiência de cores, os actos de fala e a introspecção.

3.3.6.1 *Mostrar ou dizer.*

Dennett inicia a Parte III de *Conciousness Explained* – a parte finalmente filosófica da obra, em que o modelo apresentado é submetido à panóplia de testes dos filósofos (como os *zombies*, o Quarto Chinês, Mary e o morcego) – retomando a questão das imagens mentais. Estas constituem aparentemente um obstáculo ao encerramento do Teatro Cartesiano, pois não parecem poder dispensar a apresentação. Como se viu atrás, o modelo de S. Kosslyn admite um *medium* imagético, entidades quase-pictóricas e uma apresentação visual ou mostração interna dessas entidades quase-pictóricas. Uma tal mostração seria análoga àquela que acontece num écran e aconteceria aliás, segundo Kosslyn, de uma forma semelhante ao que acontece num vulgar computador: a informação pode estar armazenada na memória com outro formato, no entanto, quando é recuperada, a sua mostração é imagética. É sobre mostrações imagéticas que as transformações são executadas e é às mostrações imagéticas que as soluções dos problemas propostos aos sujeitos se reportam.

Dennett sublinha antes de mais que a alusão à mostração num tubo de raios catódicos (TRC) é *metafórica* (o que aliás Kosslyn admite) e relembrando Shakey[614] começa a introduzir modificações no quadro de Kosslyn. Existem sistemas que de facto manipulam imagens reais: por exemplo sistemas gráficos de computadores como sistemas CAD (*computer-aided design*) realmente manipulam imagens para resolver problemas de engenharia e de arquitectura. O problema é saber se se passará o mesmo no cérebro humano. Na memória do computador os objectos são objectos tridimensionais

[614] Cf. 3.3.1.

virtuais, definidos por coordenadas, e o ponto de vista do observador é mais um ponto definido por coordenadas no mesmo espaço. Dados acerca de propriedades de cor, opacidade, textura são igualmente armazenados como informação codificada. A rotação de um objecto como um objecto de Shepard[615] neste sistema consiste num cálculo de coordenadas e as conclusões acerca de superfícies visíveis e de ocultações são conclusões geométricas. Estes cálculos podem inclusivamente servir para criar movimento no écran. No entanto, tais sistemas que manipulam objectos tridimensionais virtuais são segundo Dennett *algo de novo na natureza, não reproduzem o que se passa no cérebro humano.* Vários problemas espaciais, aparentemente não mais complicados (em termos do cálculo envolvido) do que os casos estudados pelos psicólogos, provam-no, na medida em que são irresolúveis mentalmente.

O "argumento" de Dennett é de novo uma experiência mental. Dennett propõe que se imagine um sistema CAD para engenheiros cegos, a que chama CADBLIND. O sistema resolveria questões acerca de imagens e produziria output num formato não visual (como por exemplo linguagem Braille ou voz sintetizada). Quando um engenheiro cego se confrontasse com um problema tal como "a superfície A do sólido S será ou não visível para o observador O após a rotação R?" ele colocaria a questão ao sistema e esperaria pela resposta. Como se pode prever, o objectivo de Dennett é provar que há um sentido em que o que se passa no interior do cérebro para haver imagens é um "dizer" e não um mostrar, precisamente como Z. Pylyshyn defendeu contra S. Kosslyn.

Num imaginário Mark CADBLIND I existiria um sistema CAD, um TRC em frente do qual estaria um *Vorsetzer*[616] (no caso um sistema de visão por computador com uma câmara de televisão dirigida ao TRC) e dedos robóticos para manipular o sistema CAD. Ao contrário de Shakey, este sistema realmente estaria perante uma imagem real num écran. Quando um problema lhe é colocado, Mark CADBLIND I gera imagens reais, nomeadamente visíveis para a câmara do *Vorsetzer*. Mas Dennett imagina em seguida um Mark CADBLIND II mais sofisticado, sem câmara e sem écran. Para que a mesma questão seja colocada, resolvida e utilizada pelo engenheiro cego não tem que existir agora qualquer mostração real. Pode existir simplesmente um mapa de bits, representando explicitamente cada ponto, enviado ao *Vorsetzer* através de um cabo, ou mesmo apenas uma codificação geral etiquetando regiões no mapa, sem necessidade de representar ponto a ponto as características da "imagem". Propriedades assim especificadas não têm que ser mostradas de novo a observador algum. É evidente que o ponto de Dennett é que os sistemas CADBLIND manipulam imagens e resolvem problemas acerca de imagens sem qualquer *apresentação.*

[615] Cf. o ponto 3.2.2. deste capítulo.

[616] Traduzido literalmente *Vorsetzer* significa "o que está sentado em frente a".

Dennett não nega a partir daqui as capacidade imagéticas dos humanos estudadas por psicólogos como Kosslyn. Reconhece mesmo que essas capacidades imagéticas são benéficas na resolução de problemas (por exemplo a diagramatização, em contexto teórico ou outro, faz com que padrões saltem aos olhos e os mecanismos perceptivos humanos estão desenhados para reconhecer padrões) e nem sequer nega que um sistema de imagens com mostração (como o Mark CADBLIND I) possa existir. As pessoas não são sistemas CADBLIND. No entanto a mera possibilidade destes sistemas invalida a objecção colocada a uma teoria funcionalista da consciência como Máquina Virtual pela existência de imagens mentais que aparentemente exigem a mostração num *medium* imagético interno. De facto, também as imagens mentais das pessoas podem ser fenómenos suficientemente reais sem que tenha que existir apresentação interior num Teatro Cartesiano. O que a indeterminação e a abstracção das imagens mentais (e também a própria experiência visual) provam é que o cérebro *diz sem mostrar.* Tudo o que é possível afirmar acerca de imagens mentais pode ser afirmado em termos de funções, disposições reactivas e juízos.

3.3.6.2 Querer-dizer, reportar e exprimir.

Um outro aspecto da centralidade desconstruído em CE além da apresentação é o *intencionar*, o aspecto de comando supostamente envolvido no querer-dizer linguístico e no querer-fazer motor. À primeira vista, se existe intenção no querer-dizer (*mean*), deveria existir um Intencionador Central. Ora, basicamente, como se viu atrás, Dennett argumenta que *o que existe no interior dos humanos é um processo acentrado e não um "decisor" a cujo acto de fala corresponderia a intenção ou ordem.* Assim sendo, uma das incumbências do modelo de CE é oferecer um esclarecimento epistemológico do estatuto daquilo que é dito acerca da vida interior própria. É

com essa finalidade que Dennett retoma em CE a distinção entre *reportar* e *exprimir* desenvolvida por David Rosenthal na sua teoria da consciência[617].

Antes de mais, Rosenthal distingue consciência de estado (*state consciousness*) de consciência de criatura (*creature consciusness*). Esta última consiste em estar desperto e sensiente e não é o que é necessário explicar. A consciência de estado pressupõe a consciência de criatura e é o que é necessário explicar, pois é o que faz com que os estados mentais conscientes difiram dos estados não conscientes. Se todos os estados mentais fossem conscientes não poderia haver explicação da consciência de estado, no entanto não é esse o caso. A ideia de Rosenthal é que é possível explicar o que é para um estado mental ser consciente através de estados mentais que não são eles próprios conscientes. Estes estados mentais não conscientes serão depois explicados em termos não mentais. Note-se que Rosenthal explica assim a propriedade de "ser consciente" de um estado mental como uma propriedade relacional (e não "intrínseca") desse estado.

Rosenthal parte de uma diferença relativa à consciência de estado, uma diferença entre consciência transitiva e ser-intransitivamente-consciente. Que um estado seja intransitivamente consciente não é uma propriedade intrínseca desse estado: estados mentais são intransitivamente conscientes se uma criatura for transitivamente consciente deles. A consciência transitiva, por sua vez, não é um tipo de consciência de estado mas um tipo de consciência de criatura. Daqui se obtém a definição com a qual Rosenthal se propõe resolver o problema de partida quanto à consciência de estado: um estado mental é consciente se fôr acompanhado por um pensamento de ordem superior (*higher order thought*, ou HOT) acerca dele. Como estes pensamentos de ordem superior não são usualmente pensamentos conscientes, não nos apercebemos daquilo em que consiste a consciência de estado.

É neste âmbito que Rosenthal introduz a distinção entre *reportar* e *exprimir* estados mentais que lhe servirá para caracterizar o que se passa quando uma criatura descreve a sua vida mental. Basicamente, quando se trata de *exprimir* estados mentais trata-se de dizer os estados mentais sem nos referirmos a eles. Quando se trata de *reportar* estados mentais trata-se de fazer relatos acerca dos (nossos) estados mentais, referindo-os. Ora, é o reportar que tem uma ligação com a consciência: quando alguém *reporta* um estado mental próprio, o seu acto de fala *expressa* o "pensamento de ordem superior" (*higher order thought*) de que está nesse estado mental. Exprime-se estados mentais de muitas maneiras, até com comportamentos. No entanto só é possível reportar através de actos de fala. O que é importante para Rosenthal é encontrar a pedra de toque dos estados (intransitivamente) conscientes e essa pedra de toque consiste em ser acompanhado por pensamentos de ordem superior (que não têm que ser conscientes) para que o seu objecto seja consciente[618].

Segundo Dennett, a distinção de Rosenthal mostra claramente que quando alguém diz aquilo que pensa, essa pessoa não exprime directamente as suas crenças de ordem superior (i.e. as crenças sobre crenças). O que acontece é antes que as suas crenças (que poderiam jazer não convocadas há muito tempo) geram um pensamento episódico de ordem superior e é esse o pensamento que é expresso. Não se presta atenção ou focaliza o pensamento que se exprime e sim aquilo acerca de que esse pensamento é. O pensamento expresso pode portanto ser inconsciente, e é esse aliás o estatuto, de acordo com Rosenthal, da maioria dos pensamentos de ordem superior, exceptuando os pensamentos introspectivos (que não constituem a maioria dos pensamentos de ordem superior).

Embora a teoria da consciência de Dennett seja por vezes posta em paralelo com teorias da auto-monitorização como a de Rosenthal, Dennett crê que Rosenthal sugere apenas uma útil análise conceptual, não uma teoria. As distinções de Rosenthal são introduzidas em CE com propósitos bem específicos: a teoria de Rosenthal é, conjuntamente com a interpretação da visão cega que a seguir se analisará, um dos apoios fundamentais para o ataque à noção de *zombie* que é essencial à teoria dennettiana da consciência. De facto, a teoria de Rosenthal permite pensar que a consciência é uma questão relativa à ocorrência de pensamentos de ordem superior. Para Rosenthal um estado é consciente se é acompanhado por um pensamento relativo a estar-se nesse estado, o qual pode ser inconsciente. O ângulo a partir do qual Rosenthal aborda o problema da consciência permite assim pôr totalmente de lado a intuição comum acerca do relatos da vida interior de acordo com a qual

[617] D. Rosenthal é um dos proponentes das chamadas *higher-order monitoring theories of consciousness*. Cf. BLOCK FLANAGAN & GÜZELDERE 1997, Parte X. Cf. também ROSENTHAL 1993 e ROSENTHAL 1997.

[618] São os pensamentos introspectivos, estados de terceira ordem sobre estados de segunda ordem, que serão assim.

estes pressuporiam apreensões internas prévias. A partir de Rosenthal, Dennett sugere que se acrescente ao sistema até agora proposto – um sistema capaz de ter imagens mentais e de querer-dizer mas que pode apesar de tudo ser um *zombie* — a capacidade de executar actos de fala que exprimem estados de consciência, e a capacidade de ter pensamentos de ordem superior, no sentido de Rosenthal. Dennett chama a este sistema um zimbo: «Um zimbo é um *zombie* que em resultado de auto-monitorização tem estados informacionais internos de ordem superior (embora inconscientes) que são acerca de outros estados informacionais de ordem inferior»[619]. Os zimbos pensam que são conscientes e que têm *qualia*. Infelizmente estão errados mas nem eles próprios nem ninguém poderá alguma vez vir a descobrir esse facto. Se o adversário, o defensor do *zombie*, admite esta noção de zimbo, é a própria noção de *zombie* que cai por terra: o defensor do *zombie* teria que demonstrar que a diferença que propõe entre *zombies* e não-*zombies* não existiria igualmente entre estes zimbos, o que seria, de acordo com Dennett, impossível. É a sofisticação dos mecanismos do zimbo – a sua arquitectura cognitiva – a responsável por aquilo a que se chama "consciência". A teoria da consciência de Rosenthal serve como um elemento para a prova de que um tal *zombie* é simplesmente um ser consciente.

Neste momento tem-se a seguinte situação: ou a análise de Rosenthal se sustenta e a ideia de *zombie* cai, ou a ideia se *zombie* se sustenta e nesse caso é preciso mostrar que a análise de Rosenthal falha. É fácil imaginar em que sentido se inclina Dennett. Ele pensa que a análise de Rosenthal se sustenta. No entanto, Dennett considera aquilo que Rosenthal faz é apenas uma análise lógica do conceito comum de consciência e o facto de não existirem pensamentos de ordem superior *salientes e independentes* constitui um problema. Evocando de novo o modelo da produção por Pandemónio, o que de facto acontece, segundo Dennett, quando alguém relata o que se passa consigo numa "situação de Rosenthal" «não é entrar num estado de auto-observação de ordem superior, criar um pensamento de ordem superior, de modo a poder reportar o pensamento de ordem inferior exprimindo o pensamento de ordem superior. O que acontece é antes que o estado de segunda ordem (...) vem a ser criado pelo próprio processo de enquadrar o relato (*frame the report*).

[619] DENNETT 1991: 310.

Nós não apreendemos primeiro a nossa experiência no Teatro Cartesiano e depois, com base no conhecimento adquirido, possuímos a habilidade de enquadrar relatos para exprimir; o nosso sermos capazes de dizer como é, é a base para as crenças de ordem superior»[620]. É, em suma, a "emergência" da expressão que fixa o conteúdo do pensamento de ordem superior, não tem que existir um pensamento episódico a mais.

De tudo isto retenha-se especialmente que, de acordo com a teoria da consciência defendida em CE, a *complexidade* funcional dos *zombies*-zimbos, nomeadamente a sua capacidade de estados informacionais de ordem elevada, é o que está em causa na consciência. Nada poderá ser comportamentalmente idêntico a um humano se não for funcionalmente complexo desta forma. E se algo for funcionalmente complexo desta forma será necessariamente consciente.

3.3.6.3 *Quebrar a barreira da testemunha: uma interpretação da visão cega.*

A visão cega constitui à primeira vista uma objecção importante à convicção de Dennett segundo a qual a consciência não é isolável da cognição em geral. Essa convicção é, como se tem vindo a notar, a principal razão pela qual, para Dennett, os *zombies* são *logicamente impossíveis*. No entanto, sujeitos com visão cega seriam "*zombies* parciais" reais (a fortiori possíveis): neles a função cognitiva está presente e a consciência ausente. Surpreendentemente, Dennett defende em CE que longe de sustentar o conceito de *zombie*, o fenómeno da visão cega o destrói. Em conjunto com a ideia de pensamentos de ordem mais elevada, a interpretação da visão cega constitui o núcleo do caso erguido em CE contra a noção de *zombie*. A visão cega consiste, recorde-se, na situação em que pacientes são capazes de "adivinhar" estímulos visuais (como *flashes* de luz, formas geométricas, movimentos) que declaram taxativamente não ver. Nnguém nega que os sujeitos recebam informação visual, até porque o seu sistema visual, exceptuando o cortical, está intacto. Ninguém nega também que os sujeitos reportem sinceramente que não apercebem o que quer que seja. O que acontece é que os sujei-

[620] DENNETT 1991: 315. Dennett considera estar a esclarecer com estas análises o que anteriormente afirmara em C&C acerca da análise anscombiana da intenção e o que afirmara acerca do componente RP no contexto do modelo de BS.

tos precisam de "pistas" (*cues*) para darem as respostas que os experimentadores procuram. Sem pistas nada acontecerá. Esta é uma diferença muito grande relativamente à visão normal, na qual as pessoas não precisam de "pistas": um paciente com visão cega que espontaneamente "adivinhasse", i.e. que não precisasse de pistas, geraria grandes dúvidas acerca do seu estado. Uma pessoa com visão normal é capaz de utilizar a informação fornecida pelos seus próprios olhos sem pistas, enquanto que a pessoa com visão cega é caracterizada precisamente por necessitar de pistas para utilizar a informação. Ela necessita de ser forçada a adivinhar.

O estatuto das pistas neste tipo de situações é complicado. Não teria por exemplo sentido instruir um sujeito da seguinte maneira: «quando estiver consciente da luz, pressione o botão da esquerda, quando a luz se acender *mas não estiver consciente*, pressione o botão da direita»[621]. O que interessa Dennett, no entanto, é que o estatuto do controlo "inconsciente" pode ser alterado. Dennett sugere que se imagine um sujeito com visão cega treinado para dar pistas a si próprio (*self-prompting*). Ele seria assim capaz de reconhecer quando é o momento propício para adivinhar. Segundo Dennett, apenas empiricamente se poderá decidir se isso é possível ou não. O problema de Dennett é saber se o sujeito se tornaria assim consciente dos estímulos em causa. A arriscada sugestão, central nesta interpretação, é que o sujeito se tornaria de facto "consciente" dos estímulos. Otto, o crítico, contrapõe que desenvolver um talento-conscientemente-controlado-de-adivinhar-quando-adivinhar não é idêntico a estar directamente consciente do evento sobre o qual se "adivinha": no primeiro caso falta alguma coisa para existir consciência visual (de acordo com Otto faltam os *qualia*, evidentemente).

Mas Dennett, como se sabe, não concede à partida uma tal diferença. Pelo contrário, ele defende que aquilo a que se chama consciência fenomenal ou *qualia*, cuja ausência se evoca, é apenas riqueza de informação disponível e acessibilidade pronta dessa informação situação. A diferença entre o paciente com visão cega "normal" e o paciente que aprende a adivinhar quando adivinhar imaginado por Dennett é uma diferença de grau, e é apenas nisso que consiste a diferença entre a (suposta) consciência fenomenal e outros tipos de apercebimento. Dennett exemplifica esta proposta com uma situação, o jogo de "esconder o dedal" (*hide the thimble*). De acordo

[621] DENNETT 1991: 328.

com as regras, o dedal tem que ser escondido à vista, simplesmente quando alguém entra no quarto e o procura não o vê. Ou melhor, não-o-vê-embora-o-veja. Tudo o que a distinção entre a primeira situação (ver sem ver) e a segunda (ver) envolve é uma intensificação da consciência (*consciousness raising*) e Dennett sugere que seria exactamente isso o que se passaria no caso da visão cega "auto-controlada". A "consciência" do indivíduo com visão cega não vem do nada: os estados informacionais em causa estão já presentes, no "fundo" de apercebimento (como no caso do jogo de esconder o dedal). A ideia é que uma ligação mais forte poderia ser construída no paciente com visão cega, uma ligação que intensificasse a consciência, como se intensifica a consciência de um provador de vinhos que aprende a distinguir vinhos ou a consciência da criança que de súbito vê o dedal.

Otto objecta que ainda que se isso se verificasse essa consciência não seria consciência *visual*, pois os *qualia* visuais estariam ausentes. Dennett introduz então o caso da visão dérmica[622], o caso de experiência com pacientes cegos em que estes são estimulados numa área (no ventre ou nas costas) por uma "imagem" (um arranjo espacial de pontos). Também aqui, como na questão do imaginado treino do sujeito com visão cega, a questão é saber se se trata realmente de um "ver", dada a estranheza do método de obtenção das "imagens". O facto é que os sujeitos vêm a ser capazes de reconhecer imagens, isto é, vêm a ser capazes de interpretar como imagens de objectos os arranjos de toques na pele. No entanto, a questão quanto ao carácter "visual" da experiência permanece. Ora, de acordo com Dennett, a única razão para não se falar aqui de "ver" no sentido usual do termo é a pobreza relativa da informação. O facto de a visão não ser nestes casos acompanhada, nomeadamente, por efeitos por exemplo emocionais, é explicado por Dennett em termos da diferença de quantidade no fluxo de informação[623]. Também na visão cega se trata apenas de menor quantidade de informação e a sugestão de Dennett é que é precisamente a diferença de quantidade que explica o facto de a informação ser

[622] DENNETT 1991: 339. Dennett fala de *prosthetic vision* (visão protésica). A expressão visão dérmica é usada por N. Humphrey em HUMPHREY 1995:85. P. Bach-y-Rita, o criador do aparelho de "substituição sensorial" afirma que os pacientes adquirem percepção visual (e não táctil) já que tal percepção os informa acerca das qualidades espaciais das coisas no ambiente.

[623] Em DENNETT 1991: 341 Dennett cita o caso dos dois rapazes cegos aos quais foram mostradas por meio do aparelho fotografias da revista *Playboy* tendo ficado muitíssimo desapontados com a ausência de reacção, embora pudessem aperceber o conteúdo das fotografias.

sentida tão diferentemente (nomeadamente não ser *sentida como* experiência visual).

3.3.6.4 *Ver é saber?*

De modo a reforçar os princípios que regem a análise da experiência – neste momento poderíamos considerar que tais princípios são não apenas, em geral, o *funcionalismo* mas também o *intelectualismo* e o *princípio da presença virtual* – e continuando a ter como referência a visão, Dennett chama a atenção para o funcionamento do ponto cego na visão normal. O ponto cego não é apercebido pelo "vidente", i.e. este não apercebe uma descontinuidade nas imagens. O facto é usualmente explicado fazendo apelo a um *preenchimento*. Para Dennet, como se sabe, a ideia de preenchimento é homuncular. Ele sugere então que o que acontece é que a região na qual "cai" o ponto cego não corresponde a uma falta no campo visual porque a região está já etiquetada. A região é representada (embora não individuadamente) sem que esteja actualmente a ser recebida informação[624].

Ora, assim como não tem que existir preenchimento actual do "ponto cego" numa imagem, nem todas as maneiras de representar cores necessitam de representar actualmente e individuadamente a cor que representam. Fora dos cérebros humanos, existem aliás maneiras diferentes de representar cores: por exemplo um sistema CAD pode representar cores associando um número de cor a cada pixel. A eficiência da representação da cor por números assenta no facto de não terem que ser preenchidos explicitamente os valores para cada pixel. O "preenchimento" pode ser puramente informacional e em massa. Dennett não pretende defender que o cérebro humano utiliza um sistema de cores-por-números para representar cores. No entanto "números" são evidentemente magnitudes quaisquer. Falando do cérebro falar-se-á por exemplo de um sistema de vectores para a codificação de cores[625], e as magnitudes em causa serão qualquer variação física (por exemplo as frequências de disparos neuronais) suficiente para a codificação. O ponto de Dennett é que uma codificação semelhante não requer uma descodificação que "traga de volta" a cor "apresentada". Assim como no exemplo do círculo cor-de-rosa na contracapa de CE o cor-de-rosa não existe[626] (i.e. não há cor-de-rosa na imagem retiniana nem preenchimento num *medium* do valor de um dado parâmetro), pode haver uma etiqueta única para toda uma região num mapa[627].

A questão da representação da cor no cérebro é obviamente uma questão empírica aberta. No entanto Dennett está interessado em aplicar à percepção das cores o princípio interpretativo da presença virtual que tem como exemplo central o "papel de parede coberto de Marilyns" já referido no Capítulo 1. Recorde-se o exemplo: alguém entra num quarto e sabe que é o caso que o papel de parede está coberto de imagens de Marylin (e não de manchas indistintas que poderiam, mediante verificação posterior, vir a ser ou não imagens de Marilyn). A pessoa sabe que o papel está coberto de imagens de Marilyn e no entanto é impossível que todas as imagens tenham sido "vistas" (a visão parafoveal é pobre e não é possível visar mais do que uma ou duas Marilyns de cada vez). O que acontece é que toda a região é simplesmente etiquetada como "mais do mesmo", no caso, "mais Marilyns". Ver é assim "saber" e a presença da maior parte das imagens de Marylin é virtual. A "riqueza" está no mundo ele próprio. Não é no entanto isso que (a)parece ao vidente. A este parece-lhe ver *todas* as Marylins. A ideia de Dennett é que ele de facto as vê, não porque existam centenas de Marilyns representadas uma a uma no cérebro mas porque o cérebro representa que "há centenas" (i.e. representa a informação, o conteúdo, e não as centenas unidade a unidade).

O desmantelamento do programa de protecção de testemunhas culmina assim numa teoria da presença virtual, em que a etiquetação cognitiva, o "saber", substitui o preenchimento real. A etiquetação é virtual pelo menos em dois sentidos: pelo facto de ser um saber-que e pelo facto de não ter que estar já lá. Se existir evidência contraditória, haverá reajustamento. No entanto, e nisso o caso

[624] É precisamente um caso de representação de presença *sem presença de representação específica* (i.e. sem que seja recebida informação específica).

[625] Cf. por exemplo CHURCHLAND 1995.

[626] Seria conveniente a presença neste local da imagem de que se fala. No entanto, e de acordo com os princípios de Dennett (e portanto cometendo de certo modo uma petição de princípio...) considerar-se-á que é possível descrevê-la verbalmente. Trata-se de uma grelha de linhas perpendiculares (13 por 13), em geral negras, excepto numa zona circular, definida sobre a grelha, em que as linhas são vermelhas. Os quadrados ou células definidos pelas linhas perpendiculares são totalmente brancos (embora não seja isso que se *vê*). O que se vê é aparentemente um tom cor-de-rosa que preenche a zona circular.

[627] As interpretações de Dennett não são de todo consensuais. Cf. CHURCHLAND & RAMACHANDRAN 1993, para uma crítica ao total afastamento da noção de preenchimento.

do ponto cego é exemplar, não ter nenhuma evidência não é o mesmo que ter evidência contraditória, que perturbe e despolete o reajustamento. Concretamente, como no cérebro não existem precedentes de recepção de informação proveniente do ponto cego, também não existem "agentes ansiosos por essa informação". Logo, embora a informação não chegue, não há "reclamação" e a sua "falta" não é sentida. Como diz Dennett, o ponto cego é um caso de negligência[628] sem problemas, um caso em que todas as pessoas com visão normal sofrem de anosognosia, i.e. não se apercebem do seu próprio défice. O que Dennett pretende com os princípios da presença virtual e do intelectualismo é generalizar uma tal "negligência" à nossa normal consciência de conteúdos. Em suma, nos supostos casos de preenchimento, o cérebro não preenche, antes ignora. De novo Dennett reencontra um princípio sugerido por M. Minsky em *The Society of Mind*[629], de acordo com o qual "Nada pode parecer x a não ser que seja representado como x". O MEM tem absoluta necessidade da ideia de negligência, que o princípio de Minsky reforça. É ela que sustenta a substituição da Testemunha pelos especialistas: os especialistas ou agentes procuram apenas o seu próprio objecto e nada mais. Evidentemente isto parece estar a deixar de fora (relegando-as por exemplo para estados mentais imateriais) as Marilyns que são vistas mas que não existem no cérebro. Elas realmente (a) parecem e não parece ser possível negá-las a não ser que esteja a ser negado, contra todas as evidências, aquilo que existe. A última palavra de Dennett é que se trata de presença virtual, de um "saber" disponível no caso de alguma "pergunta" ser feita. Assim, não é verdade, ao contrário do que objecta Otto, que o que não está lá no cérebro tenha que estar lá, algures, na mente consciente, quando uma tal presença faz diferença na experiência. De facto, nós, os "utilizadores" dos nossos próprios cérebros, não somos capazes de distinguir por introspecção entre "o que sempre esteve lá, na memória, armazenado sob a forma correcta" e "o que é disponibilizado para apresentação pelos agentes" apenas quando uma "pergunta" é feita. Pensar que seríamos capazes de tal distinção é esquecer que a nossa experiência presente se dá de acordo com aquilo a que M. Minsky chama a *ilusão da imanência*[630].

[628] Nos sindromas neuropsicológicos de negligência as pessoas deixam por exemplo de "apoderar/considerar" como suas partes do (seu) próprio corpo.

[629] MINSKY 1985: 287.

[630] «The Immanence Illusion: Whenever you can answer a question without a noticeable delay, it seems as though the answer were already active in your mind» (MINSKY 1985: 155).

3.3.6.5 *A Cor, de novo (e uns certos gozos e desgostos ligados aos* qualia*)*[631].

«But, Dan, *qualia* are what make life worth living!», Wilfrid Sellars[632]

«If all I want when I drink fine wine is information about its chemical properties, why don't I just read the label?», Sidney Shoemaker[633]

Para além de procurar desconstruir as ideias de mostração e de querer-dizer comandado por um Intencionador Central, o modelo de CE procura ainda desconstruir com uma teoria funcionalista e evolucionista da cor a ideia de espaço fenomenal, a que estaria associada uma distinção entre *aparecimento e sentimento por um lado* e *disposições reactivas por outro*. As cores são exemplos por excelência dos *qualia*. A ciência física mostra que as cores não estão "lá fora" no mundo. No entanto, como nós experienciamos o mundo colorido, as cores têm, aparentemente, que estar "cá dentro na mente" como qualidades fenomenais, propriedades intrínsecas de experiências conscientes (não se vai afirmar que elas não existem!).

A questão da cor tinha sido abordada brevemente por Dennett em C&C. Fora então proposto que as cores eram propriedades funcionais, por oposição a propriedades físicas primárias e a propriedades fenomenais. Na altura, a ideia era definida da seguinte maneira: «uma coisa é vermelha se e só se quando é vista em condições normais por humanos normais lhes parece vermelha, o que significa: há utilizadores dos termos para cores demonstravelmente não excêntricos e eles dizem sinceramente que a coisa parece vermelha. Que eles digam isso não depende de alguma qualidade interna mas da sua percepção do objecto, do facto de aperceberem-1 que a coisa é vermelha»[634]. Isto significa que se todos os seres humanos fossem cegos à distinção entre vermelho e verde «seria conhecimento comum que os rubis e as esmeraldas eram ambos "gred" – pois afinal pareciam a observadores normais exactamente como as outras coisas "gred": carros de bombeiros, relvados (...), maçãs ver-

[631] Para fundamentar a discussão que se segue, cf. HARDIN 1988 e HARDIN 1993, CLARK 1998 e CHURCHLAND 1995. Para críticas à teoria da cor defendida por Dennett cf. TOLLIVER 1994 e WHITE 1994.

[632] Em conversa pessoal, citada em DENNETT 1991: 383.

[633] Em conversa pessoal, citada em DENNETT 1991: 383.

[634] DENNETT 1969: 146.

des e maçãs maduras»[635]. Pessoas que vissem a distinção entre verde e vermelho não poderiam chegar ao pé dessas outras pessoas e afirmar que o seu sistema de visão era "mais verdadeiro". Em CE Dennett continua a elaborar uma teoria funcionalista, apoiada em bases evolucionistas.

De modo a tornar os "qualófilos" mais sóbrios Dennett faz notar que por exemplo os insectos têm visão da cor e ninguém mostra especial vontade de lhes atribuir consciência[636]. É portanto razoável pensar que visão de cores sem consciência é não apenas possível como actual. O debate filosófico em torno da natureza da cor estrutura-se usualmente como um debate entre subjectivistas e objectivistas[637]. Saber se as cores são objectivas é saber se é possível afirmar que elas existem como existem a massa e o movimento, i.e. como propriedades de objectos no mundo. Como não é legítimo afirmar simplesmente que as cores não existem no mundo mas apenas para o observador (não é legítimo pois este faz parte do mundo) trata-se antes de considerar, quando se discute a objectividade das cores, se elas serão as *causas* do facto de os objectos

[635] DENNETT 1991: 379.

[636] DENNETT 1991: 377. Pelo contrário vários mamíferos aos quais todos nos sentimos tentados a atribuir consciência (como cães e gatos) não têm uma visão de cores tão rica com a nossa (são dicromatas), enquanto pássaros, peixes, répteis e os aludidos insectos (por exemplo abelhas), têm sistemas visuais tricromáticos semelhantes aos humanos e alguns têm mesmo uma visão de cores mais rica do que a dos humanos (por exemplo pentacromática). A classificação é feita em função dos tipos diferentes de pigmentos nos cones (células fotossensíveis da retina, que determinam a gama de radição electromagnética à qual o sistema é sensível). Entre os mamíferos, apenas os primatas possuem visão de cores, embora com muitas diferenças entre si. É um problema teórico para a biologia evolucionista saber por que razão têm visão de cor exactamente as espécies que a têm.

[637] A exposição que se segue baseia-se sobretudo em HARDIN 1988, HARDIN 1993, TOLLIVER 1994 e HUMPHREY 1992, para além de DENNETT 1991, Capítulo 12, *Qualia Disqualified*. Dennett defende em CE uma teoria objectivista, funcionalista e evolucionista da cor. Humphrey argumenta que os nossos gostos e desgostos em relação a cores reflectem a nossa história evolutiva. J. Tolliver (TOLLIVER 1994) está particularmente preocupado em defender que o conteúdo da experiência de cor é informacional e não representacional. Hardin parte da investigação contemporânea sobre cores e visão de cores para a defesa de uma posição filosófica *subjectivista* acerca da natureza da cor, investigando possibilidades alternativas de reducionismo fisicalista acerca de cores (cores como comprimentos de onda da luz, cores como disposições dos objectos para afectarem observadores normais em condições standard de observação, etc). Acaba por considerar preferível pensar nas cores como propriedades definíveis em termos de sistemas de estados neuronais de observadores e não por relação com a enorme variação no mundo exterior que origina estes estados: «colored objects are illusions, but not unfounded illusions. We are normally in chromatic perceptual states, and these are neural states» (HARDIN 1988: 111). Como Hardin acentua, qualquer ideia de ilusão de cor (uma ideia sem a qual não se exporá nada acerca de cores e mistura de cores) é imediatamente complicada: ela supõe uma falha na correspondência entre a cor que o objecto realmente tem e a cor que ele parece ter. O problema é evidentemente procurar determinar o que significa isso, para um fisicalista.

aparecerem coloridos ao observador, se existirá uma relação sistemática entre as propriedades dos objectos que causam percepção de cor e o conteúdo da experiência de cor do observador. A maneira lockeana de considerar as qualidades secundárias – como poderes ou propriedades disposicionais que as coisas do mundo têm de provocar determinandos efeitos nas mentes de observadores normais – acompanha ainda hoje usualmente o objectivismo. O problema é que é muito difícil dizer exactamente que propriedades no mundo "provocam" as cores nas mentes de observadores normais. Não parece possível por exemplo dizer de uma maneira não trivial o que é que todas (e apenas as) causas de aparências verdes têm em comum que constituiria o seu serem "objectivamente (vistas/visíveis como) verdes" por observadores normais[638]. De facto, ligando a cor à radiação electromagnética com vários comprimentos de onda sendo reflectida e absorvida pelas superfícies, o que se verifica é que não existe uma propriedade simples P de superfícies tal que apenas as superfícies com essa propriedade P sejam (vistas como) vermelhas, ou azuis, ou verdes, etc. Superfícies com propriedades reflexivas diferentes podem ser vistas como sendo da mesma cor, e a mesma superfície sob diferentes condições de iluminação pode ser vista como sendo de cores diferentes[639]. A relação entre os comprimentos de onda da luz absorvida e reflectida pelas superfícies e as cores vistas pelos observadores é portanto no mínimo indirecta.

Se não existe uma ancoragem objectiva das cores na maneira como as coisas são, se elas estão pelo contrário ligadas à experiência de cor, as cores serão subjectivas. De um ponto de vista biológico, o que se verifica é que diferentes sistemas de visão de cor evoluíram de forma independente, dando lugar a espaços de cor diferentes em diferentes criaturas A visão de cores não proporciona às criaturas que a possuem acesso a "géneros naturais". Mas se, assim, "ser vemelho" consiste num (a)parecer vermelho a um determinado tipo de seres, será circular explicar que algo (a) parece vermelho porque é (independentemente do facto de ser percebido) vermelho. Isto parece apoiar a intuição subjectivista segundo a qual as cores estão ligadas às experiências de cor. Por exemplo vermelho é a pro-

[638] Quando se começa a olhar de perto, as coisas "desaparecem todas". Este princípio de CE também se aplica a noções inabdicáveis na colorimetria, tais como as noções de observador normal e de condições standard de observação.

[639] Cf. HARDIN 1988 para muitos exemplos.

priedade que acontece quando objectos (a)parecem vermelhos. O problema regressa então sob a seguinte forma: não há uma fenomenologia determinada da cor à qual ligar as cores[640]. Se a intuição subjectivista fosse verdadeira, as cores deveriam depender de experiências de cor estáveis. Ora não se encontra uma tal estabilidade na experiência de cor.

A posição *objectivista*, *relacional* e *evolucionista* de Dennett acerca das cores parte da seguinte ideia: algumas coisas na natureza precisavam de ser vistas, outras precisavam de ver e assim as cores foram feitas para serem vistas por aqueles que foram feitos para as ver. O que a ciência física e biológica mostra é que as propriedades que os objectos têm de reflectir a luz fazem com que criaturas entrem em «estados discriminativos, distribuídos nos seus cérebros, e subjazendo a um conjunto de disposições inatas e hábitos aprendidos de complexidade variável (...) estes estados discriminativos dos cérebros de observadores têm várias propriedades primárias (...) e em virtude destas propriedades primárias têm várias propriedades secundárias meramente disposicionais. Em criaturas humanas com linguagem (...) estes estados discriminativos (...) eventualmente dispõem as criaturas a exprimir juízos verbais aludindo à "cor" de diversas coisas. Quando alguém diz "eu sei que o anel não é realmente cor-de-rosa, mas realmente parece cor-de-rosa" a primeira cláusula exprime um juízo acerca de alguma coisa no mundo e a segunda cláusula exprime um juízo de segunda ordem acerca de um estado discriminativo acerca de alguma coisa no mundo. A semântica de tais asserções torna claro o que as cores supostamente são: propriedades reflexivas das superfícies de objectos (...) e é isto que elas são de facto – embora dizer exactamente quais propriedades reflexivas são sejam baste complicado (*tricky*)[641]».

[640] Cf. os problemas levantados pelo metamerismo e a definição de cor como vista por um observador normal em condições standard em HARDIN 1993.

[641] DENNETT 1991: 372-373. O problema, quando se trata de dizer exactamente quais propriedades são, é que «colors are not "natural kinds" precisely because they are the product of biological evolution, which has a tolerance for sloppy boundaries when making categories that would horrify any philosopher bent on good clean definitions. If some creature's life depended on lumping together the moon, blue cheese and bicycles, you can be pretty sure that Mother Nature would find a way for it to "see" those as "intuitively just the same kind of thing"» (DENNETT 1991: 381). Esta razão explica aliás em parte por que os *qualia* são tão "inefáveis e indefiníveis": não estão a ser tomadas propriedades básicas ou simples, mas apenas a ser feitas discriminações úteis ou que foram úteis para os antepassados do sistema. Em suma, é difícil dizer *o que é* (uma cor, um odor, um sabor), mas podemos dizer com toda a clareza por que é que é difícil dizer (a razão é a apontada).

Em suma, embora os sistemas visuais não detectem propriedades simples no mundo (como propriedades de superfícies tais que apenas as superfícies que possuem essa propriedade são "vermelhas"), as cores são propriedades objectivas, i.e. correspondem a estados discriminativos internos específicos. Nada obriga no entanto a afirmar que tais estados discriminativos têm propriedades intrínsecas e Dennett pensa que não há qualquer diferença qualitativa entre a maneira como um dispositivo como o CADBLIND Mark I realiza uma tarefa de comparação de cores e a maneira como os humanos a realizam: em ambos os casos existem estados discriminativos específicos e não existem quaisquer *qualia*.

A verdadeira chave da teoria das cores, sublinhe-se de novo, reside na resposta à questão "por que é que existem cores?". E a resposta ecologicamente relevante é a seguinte: as cores existem para codificar-por-cores. A cor e a visão de cor foram feitas uma para a outra: a visão de cor é "acerca de" reconhecimento de padrões e as cores foram feitas para serem vistas por sistemas que podem vê-las. Assim, as cores estão ligadas a classes específicas de observadores, não fazendo sentido supor cores num mundo sem observadores. Não haveria razão para a existência de cor e visão de cor "isoladamente". O facto de poder não haver algo de especificamente comum ao que é visto como a mesma cor por dados sistemas, o facto de não se tratar, na visão de cor, da detecção de géneros naturais, é mais um sinal da tolerância aos falsos positivos em sistemas cognitivos resultantes de evolução. Noutras palavras, há razões reais para a preferência e o horror por cores, mas não há razões para a existência de "cores reais": «É um erro pensar que primeiro existiam cores – rochas coloridas, água colorida, céu colorido, ferrugem alaranjada e cobalto azul brilhante – e depois a Mãe Natureza[642] chegou e tirou vantagem destas propriedades utilizando-as para codificar coisas por cores. O que aconteceu foi antes que existiam várias propriedades reflexivas de superfícies, propriedades reactivas de fotopigmentos, etc, e a Mãe Natureza desenvolveu a partir desta matéria prima sistemas de codificação por cor e sistemas de visão de cor mutuamente ajustados»[643].

Logo que um sistema de visão da cor está instalado num organismo (feito para ver coisas que precisam de ser vistas) esse sistema serve para ver muitas outras coisas, que não precisavam de ser

[642] I.e. a evolução por selecção natural.

[643] DENNETT 1991: 378.

vistas (nem do "ponto de vista" do vidente nem da coisa vista). Esse "bónus lateral" provavelmente dá conta, de acordo com Dennett, do estatuto de quase tudo o que a visão de cor humana actualmente nos dá, num ambiente em que a significação natural das cores praticamente desapareceu. No entanto continua a ser possível a significação artificial de acordo com os mesmos princípios, como o mostram vários exemplos[644], embora seja certo que a mesmidade de cor (entre uma maçã vermelha e um vestido vermelho, por exemplo) é na maioria das vezes uma coincidência e não já algo de ecologicamente significativo.

Resta esclarecer um último ponto quanto à objectividade das cores. Embora os factos acerca de qualidades secundárias estejam por natureza ligados a um grupo específico de observadores, há maneiras fortes e fracas de conceber essa relação. Na terminologia de Dennett, pode-se dizer que qualidades secundárias são ou *lovely* ou *suspect*. Supostamente alguém poderia ser *lovely* sem nunca ter sido observado pelo tipo de observador que o julgaria *lovely* (as disposições estão lá embora a qualidade vá depender da observação). No entanto, ninguém poderia ser *suspect* sem qualquer relação com alguém que o considerasse *suspect* (i.e. a qualidade nunca poderia ser definida sem já ter existido uma determinada relação com um observador). Ora as cores são, segundo Dennett, qualidades *lovely* (i.e. subjectivas de forma forte, no sentido não "subjectivista"): a intuição de que uma esmeralda não observada é "já" verde não tem que ser negada.

O facto de as cores e os cheiros serem feitos para serem vistos e cheirados introduz a questão do gosto e desgosto sentidos pelas criaturas que os experimentam. Gostos e desgostos não acontecem por acaso: eles têm razões evolutivas. Os detectores de qualidades não são "repórteres desinteressados" mas sinalizadores. O que queremos não é ser informados mas «ser informados da nossa maneira favorita (...) e as nossas preferências baseiam-se em inclinações (*biases*) que ainda estão instaladas nos nossos sistemas nervosos embora a sua significação ecológica possa ter desaparecido há milhões de anos»[645]. O que não faz sentido é pensar que uma propriedade intrínseca explicaria estas reacções.

[644] Por exemplo as cores dos cabos de aparelhos electrónicos ou as cores nos corredores de edifícios complexos como grandes hospitais que tornam possível distinguir, reconhecer e seguir pela cor (DENNETT 1991: 377).

[645] DENNETT 1991: 384. O psicólogo N. Humphrey relata experiência com cores em macacos e humanos em HUMPHREY 1995.

Otto rejeita totalmente que a maneira como as coisas aparecem seja desta forma identificável com disposições reactivas idiossincráticas do sistema nervoso de criaturas confrontadas com um padrão de estimulação. No entanto, segundo Dennett, é exactamente isso e apenas isso que é referido quando se afirma "estes são os meus *qualia*". O que interessa aqui, uma vez que as cores são talvez o exemplo mais frequentemente utilizado nas experiências mentais acerca de *qualia*, é ver de que modo esta teoria das cores se comporta no confronto com tais experiências. A questão é importante pois é em intuições acerca de *qualia* (ausentes em *zombies*, invertidos em casos de espectro invertido, não possuídos por Mary, etc) que os adversários de Dennett apoiam as suas teorias da consciência. Ora, tomando como exemplo as fantasias filosóficas de inversão dos *qualia* – um dos tipos de argumentos que suportam intuições opostas às de Dennett – o que se vê é que tais cenários pura e simplesmente pressupõem que as "disposições reactivas" e a "maneira como as coisas parecem ser" são separáveis e tudo o que foi aqui afirmado acerca de cores mostra que Dennett se atem ao princípio central de *Quining Qualia* segundo o qual não é possível uma comparação dos *qualia*, já que a comparação (seja entre sujeitos, seja intrasubjectiva) pressuporia (injustificadamente) a referida separabilidade. Tomando um caso intrapessoal de espectro invertido, teríamos por exemplo uma pessoa que acorda um dia com os *qualia* invertidos (a erva ficou vermelha, o céu amarelo, etc). Como mais ninguém nota nada de anormal (seria importante que a diferença fosse súbita e não gradual para ser notada pela própria pessoa), o problema deve estar na própria pessoa, que conclui que algum neurocientista maligno deve ter "trocado os cabos". Ora, Dennett pretende que se de facto todas as disposições reactivas tivessem sido trocadas, o comportamento relacionado com cores e a discriminação por cores também teria sido alterado: não é possível imaginar que apenas se interfere com os *qualia* e que a pessoa apercebe a situação da maneira descrita. Dennett ilustra a sua posição com o exemplo do carro pintado da exacta tonalidade de azul do carro em que alguém uma vez teve um acidente. Feitas as trocas, no início não se reagiria ao azul-invertido (amarelo, suponha-se) mas, uma vez estando tudo no sítio, voltariam as disposições reactivas e, nomeadamente, perante a pergunta "o carro do acidente de que cor era?" dir-se-ia "azul como isto" sem hesitar.

As experiências imaginárias com *qualia* invertidos costumam supor duas "inversões", uma das quais restauraria o estado inicial

(por exemplo por habituação). A finalidade de tais experiências imaginárias é descrever um caso no qual seja óbvio que os *qualia* estariam invertidos quando as disposições reactivas estivessem normalizadas e de facto muitas pessoas tendem a pensar que são concebíveis casos assim. Dennett pensa que a "intuição" existe porque essas pessoas pensam que a adaptação ocorreu *pós-experiencialmente*. No entanto para Dennett não é possível isolar propriedades apresentadas na consciência das reacções múltiplas do cérebro às discriminações destas. Enquanto exemplo típico de consciência, as cores mostram que o está em causa é o que é apercebido no apercebimento, o que é discriminado em virtude das funções do sistema, e não alguma coisa para além disso.

3.3.6.6 *Mary e os* zombies. *O Quarto Chinês. O morcego.*

A teoria dennetiana da consciência recai em última análise sobre os problemas que consistem em (1) *saber como se sabe que se é consciente* e (2) *saber o que se sabe quando se sabe que se é consciente*. Segundo Dennett a resposta só pode ser "continuista", i.e. ela deve ser dada sem admitir à partida uma distinção de natureza entre conteúdo e consciência. É com base nestes pressupostos que Dennett pretende afastar o epifenomenismo, a possibilidade de tudo o que foi dito nada dizer ainda sobre consciência, a qual poderia vir ou não a juntar-se aos funcionamentos descritos de uma forma que não faria diferença causal mas constituiria uma outra realidade. Embora o caso erguido contra o *zombie* tenha uma função análoga, é especialmente sobre a avaliação do Argumento do Conhecimento de F. Jackson que recai em CE a função crucial de afastar o epifenomenismo. Este seria a grande alternativa ao modelo de consciência proposto em CE e apenas demonstrando que ele é absurdo se consolida a proposta.

Dennett reconhece que o Argumento do Conhecimento de F. Jackson convoca irresistíveis intuições acerca de *qualia* servindo-se precisamente da experiência de cor. Considera no entanto que tais intuições – que levam a pensar que mesmo conhecendo todos os factos físicos Mary aprenderia alguma coisa no momento em que pela primeira vez experienciasse a cor – por mais irresistíveis que sejam, estão erradas. De acordo com Dennett, Mary não aprenderia nada. Antes de mais, a conclusão do argumento

de Jackson segundo a qual Mary aprenderia qualquer coisa saindo do quarto (nomeadamente, como-é-experienciar-cores) decorre de um mau entendimento das premissas. Na situação de Jackson, Mary conhece supostamente todos os factos físicos, nomeadamente neurofisiológicos, sobre o que acontece quando ocorre visão de cor, mas nunca experienciou cores. Ora Dennett declara antes de mais que a ideia segundo a qual Mary tem toda a informação física, conhece todos os factos, acerca da causas e efeitos da visão de cor é pura e simplesmente inimaginável. Aquilo que se imagina para configurar a situação de Jackson é outra coisa, por exemplo todo o conhecimento *actual* acerca da neurofisiologia da cor.

É certo que, numa imaginação razoável do caso, sair do quarto mostraria a Mary algo de novo. No entanto, precisamente, numa imaginação razoável do caso, Mary nunca poderia "ter toda a informação física". Por outro lado, Jackson evita explorar o que Mary saberia fazer e reconhecer quanto a objectos coloridos, dado todo o conhecimento que ela é suposta possuir. Se se explorasse tal aspecto concluir-se-ia por exemplo que perante objectos Mary diria talvez "Isto é azul e é diferente disto, que é vermelho". Ela fá-lo-ia a partir do seu conhecimento neurofisiológico, aplicado às suas próprias disposições reactivas perante tais objectos em situações concretas. Ter todo o conhecimento neurofisiológico significaria necessariamente conhecer os efeitos neuronais que o confronto com a propriedade "ser azul" provocaria.

A grande aceitação que a conclusão do caso de Mary encontra é para Dennett um clássico caso do síndroma filosófico que consiste em "tomar uma falha da imaginação por um *insight* acerca de necessidade". Confrontando Mary com o modelo de CE, Dennett defende que Mary não aprenderia nada de novo ao sair do quarto: ela poderia utilizar o seu conhecimento neurofisiológico para reconhecer que um objecto vermelho é vermelho quando o visse notando os efeitos da visão do objecto nas suas próprias disposições reactivas.

Como se sabe, o caso de Mary permite a Jackson concluir pela *falsidade do fisicalismo* (definido como a doutrina segundo a qual conhecer todos os factos físicos é conhecer tudo o que há para conhecer) e a favor do *epifenomenismo* (ao conhecer toda a informação física Mary não conhece tudo o que há para conhecer porque há mais qualquer coisa, i.e. existem *qualia* epifenomenais). Ora segundo

Dennett o epifenomenismo é uma doutrina absurda[646]. É de resto por considerar o epifenomenismo uma doutrina absurda e por considerar que ele está intimamente ligado com a figura do *zombie* que Dennett considera de forma tão evasiva a importância deste em experiências mentais acerca da consciência. Importa portanto, de modo a esclarecer a proposta metafísica de CE, compreender exactamente o que é o epifenomenismo e porque é que este é (ou não é) assim tão absurdo.

Em geral o epifenomenismo é a concepção do mental de acordo com a qual este não tem por si poderes causais e portanto não tem efeitos físicos, é "um efeito sem efeitos", o que proíbe desde logo pelo menos uma interacção de género cartesiano, e especificamente qualquer eficácia causal do mental sobre o físico. A isto pode acrescentar-se a convicção de que todos os efeitos são físicos, inclusive os acontecimentos mentais. De acordo com autores como Jackson, o epifenomenismo é também a tese segundo a qual factos sobre a experiência consciente não sobrevêm aos factos físicos. Para Dennett há dois conceitos de "epifenomenal" a considerar, um filosófico e um psicológico. O conceito filosófico de epifenomenal nomeia aquilo que não tem efeitos físicos no mundo, que é causalmente irrelevante e portanto irrelevante para a explicação; o conceito psicológico de epifenomenal nomeia um efeito lateral ou propriedade não funcional. Para Dennett o primeiro é um conceito ilegítimo: não pode haver qualquer evidência do que quer que seja de epifenomenal no sentido filosófico já que o que não tem efeitos é indetectável. Quanto a epifenómenos no sentido psicológico, características que têm efeitos mas não fazem diferença funcional, a vida mental está cheia deles (como atrás se viu por exemplo com as disposições reactivas face às cores e a persistência daquilo que foi ecologicamente pertinente, mas não é mais). Simplesmente, tais epifenómenos não são problemáticos, ao contrário do que supostamente acontece com os *qualia* epifenomenais de Jackson. A confusão entre os dois sentidos é, aliás, o problema não apenas de Mary mas também dos *zombies*. A confusão entre os dois

[646] Isto é algo que Dennett admite praticamente sem argumentos. No entanto, o epifenomenismo pode ser coerentemente defendido e imaginado, embora seja para muitos autores uma concepção "deselegante" (cf. CHALMERS 1997: 402 e CHALMERS 1996: Capítulos 4 (*Naturalistic Dualism*) e 5 (*The Paradox of Phenomenal Judgment*)). Chalmers defende de resto que posições usualmente postas de lado sem mais considerações como o pampsiquismo, o epifenomenalismo e o interaccionismo não são assim tão absurdas quando está em causa o lugar da consciência no mundo de acordo com uma perspectiva fisicalista.

sentidos (filosófico e psicológico) de "epifenomenal" está inscrita no próprio termo *zombie*. Para ser possível afirmar que um *zombie* é indistinguível de uma entidade consciente é necessário supor à partida que a consciência é epifenomenal no sentido filosófico, o sentido de epifenomenal que não tem sentido para Dennett. É aliás no âmbito desta crítica que Dennett faz aquela que se poderia considerar a afirmação emblemática da sua teoria da consciência: «os *zombies* são possíveis? Eles não são apenas possíveis, eles são reais. Nós somos todos *zombies*»[647]. Noutras palavras, é impossível distinguir um *zombie* de um ser consciente sem pura e simplesmente pressupor a epifenomenalidade (no sentido filosófico) da consciência. Se o *zombie* for imaginado como funcionalmente complexo e capaz de estados informacionais reflexivos, esse *zombie* somos nós.

Um fenómeno semelhante, relativamente à "falha da imaginação", ao que subjaz ao caso de Mary ocorre com a experiência mental do Quarto Chinês de Searle, que pretende, também ela, provar que há algo que falta em qualquer entendimento em terceira pessoa da mentalidade. Já no capítulo anterior foi mencionado que Dennett pensa que Searle oculta, com a sua parábola, o *insight* central de qualquer teoria materialista da mente, i.e. a possibilidade de uma interpretação mecânica do cérebro ou de outro *hardware*. De acordo com Dennett, Searle procura forçar a admissão de que para haver entendimento genuíno teria que haver alguma testemunha dos eventos que constituem o entendimento. De novo Dennett pretende que o Quarto Chinês apenas conduz à pretendida conclusão (a ausência de entendimento genuíno no sistema) mediante uma falta de detalhes da imaginação (nomeadamente uma falta de detalhes relativos ao comportamento verbal, em chinês, do Quarto Chinês). Em CE Dennett elabora um exemplo do que seria esse comportamento[648], argumentando que um programa que correspondesse às especificações de Searle, seria um sofisticado sistema cognitivo que possuiria meta-meta-....-conhecimento acerca do seu próprio comportamento e acerca do mundo e que a imaginação detalhada do sistema pura e simplesmente impediria a "intuição" pretendida. Nada poderia ser funcionalmente e informacionalmente complexo da forma comportamental evidenciada e não ser consciente. Só porque vemos que no caso de um programa curto e simples a instanciação física não gera entendimento não temos legiti-

[647] DENNETT 1991: 406.
[648] DENNETT 1991: 436-437.

midade para generalizar a um caso com a complexidade requerida pela situação de Searle a conclusão segundo a qual nenhuma organização funcional seria por si só suficiente para a consciência. Exactamente como sucede com o caso de Mary, o Quarto Chinês só persuade aqueles que não obedecem às instrucções da situação.

Quanto ao *what-it's like* nageliano, tomado como critério final e definitivo da presença de experiência sentida e ao mesmo tempo como prova de que o conhecimento da fisiologia ou de qualquer suporte nada nos diria acerca da fenomenologia do sistema, Dennett considera que o modelo apresentado em CE mostra que é simplesmente falso que o conhecimento em terceira pessoa não possa dizer-nos nada quanto a "como é ser uma determinada criatura". Aliás, o caso de Nagel baseia-se ele próprio em conhecimento em terceira pessoa acerca da ecolocalização dos morcegos. Para Dennett, um modelo como o MEM, com a sua proposta de uma Máquina Virtual implementada no cérebro como sistema de controlo, avança no sentido de preencher o abismo nageliano entre fisiologia e fenomenologia.

Os posicionamentos de Dennett face a todas estas experiências mentais pressupõem (i) uma concepção funcionalista e (ii) a crença na superveniência lógica da consciência aos funcionamentos descritos. Noutras palavras, em organizações com a complexidade descrita pelo MEM, a consciência estará necessariamente presente como papel funcional, sendo logicamente impossível que tal não aconteça.

3.3.7 *Ciência cognitiva ou teoria filosófica da consciência? O choque de intuições quanto à fundamentalidade da consciência: Dennett versus Searle, Nagel, Chalmers e Jackson. O concebível e o inconcebível.*

Em geral, os filósofos tendem a considerar CE como uma obra de ciência cognitiva, onde são feitas interessantes observações sobre processamento cognitivo, mas um quase deserto de argumentos filosóficos explícitos. Existem no entanto argumentos em CE, que correspondem a uma determinada metafísica da consciência. É certo que para os adversários de Dennett a maioria dos argumentos decorre de uma petição de princípio, já que Dennett simplesmente presume que tudo o que é necessário explicar para explicar a consciência são funções, disposições reactivas e juízos. À primeira vista, e de forma análoga àquilo que

fazem os cientistas cujas teorias empírico-especulativas foram descritas no início deste capítulo, aquilo que Dennett faz com o seu modelo de consciência é considerar os processos cerebrais correlacionados com a consciência – de acordo com o MEM mais propriamente processos do nível funcional da Máquina Virtual do que processos neurobiológicos – e tomar a correlação entre esses processos cerebrais e a consciência como um facto bruto. Por outro lado, utiliza critérios pragmáticos (nomeadamente a capacidade de reportar e controlar o comportamento) para assegurar a presença da consciência num sistema. Para a maioria dos críticos isto significa que o MEM pura e simplesmente passa ao lado da "Questão" para uma metafísica materialista, a questão da locação da consciência, com a sua subjectividade, na natureza. Dennett, no entanto, afirma que colocar que uma questão semelhante acerca da relação entre a subjectividade e os processos que a originam é comparável com a situação em que um vitalista pergunta "porque é que o ADN origina a vida?"[649]. Nenhuma das questões tem objecto.

Apesar do desentendimento entre Dennett e os filósofos que se ocupam com o problema metafísico da consciência, é possível explicitar a metafísica da consciência correspondente à teoria funcionalista de CE. No MEM não estão em causa apenas questões epistemológicas relativas a incorrigibilidade e questões fenomenológicas relativas à distinção entre veículos da representação e aquilo que é representado: o modelo corresponde também à defesa de uma posição fisicalista funcionalista e da *superveniência lógica da consciência*. De resto e apesar da semelhança da abordagem dennettiana com abordagens empíricas, o debate filosófico recente acerca da consciência transparece em vários pontos de CE. Por exemplo, num diálogo com Otto[650] afirma-se explicitamente que, ao contrário do que é usualmente admitido no debate filosófico, nada mais há a explicar numa teoria da consciência quando se explica o parecer (*seem*) em termos de disposições reactivas. Dennett exclui assim, com a sua afirmação, que algo como uma aparição fenomenal possa não ser idêntico a um juízo sobre um conteúdo de experiência. É o já referido *continuismo* entre apercebimento e consciência que sustenta tal posição. Foi aliás notório que se tratou neste capítulo muito mais de *conteúdos* conscientes

[649] DENNETT 1997.
[650] DENNETT 1991: 362-368.

do que de qualquer coisa específica que seria a "consciência", criando mesmo uma quase sobreposição temática com o Capítulo 2. Assim, aquilo que o MEM caracteriza a partir de baixo e a partir de dentro, acaba por ser o mesmo que a TSI caracteriza a partir de fora e a partir de cima: a *intencionalidade*. A noção de mundo nocional ou mundo heterofenomenológico, fundamental quer na TSI quer no MEM, acaba por ser a prova mais óbvia de uma tal confluência ou sobreposição. No entanto, para o próprio Dennett não se trata de uma confusão conceptual. O que isto significa é que Dennett, ao contrário de filósofos como Jackson ou Chalmers, considera que é desnecessário introduzir um primitivo bruto na teoria da consciência, que isso (experiências-como-juízos, i.e. conteúdo) é o que a consciência é. Um primitivo bruto apenas estaria justificado se houvesse uma divisão ontológica entre consciência e não-consciência e todo o MEM constitui um caso contra tal divisão. Começar a teoria da consciência com a introdução de um primitivo supõe admitir a impossibilidade de uma análise não circular da consciência e admitir que a consciência é algo de absolutamente distinto da não-consciência. Poder-se-ia contrapor que a consciência que Dennett explica de forma não circular não passa de apercebimento-de. Nada mais é admitido senão graus de apercebimento-de. Ora, nestas circunstâncias, de facto, de certo modo, não há como evitar a conclusão segundo a qual somos todos *zombies* e isso não faz diferença nenhuma. Note-se no entanto que nada nesta posição envolve qualquer coisa como uma abolição do sujeito: afinal, o MEM explica características do sujeito tais como a unidade, a centralidade e o auto-apercebimento. A única coisa que é de facto banida é um sentir-se ser sem mais qualificações, comum a todas as entidades conscientes, por hipótese com vidas mentais muito diferentes entre si, que corresponderia a uma descontinuidade no mundo e a algum tipo de dualismo. A ideia de uma tal forma comum inespecificada, cuja pedra de toque seria o *what-it-is-like*, é banida mas não as características de "forma" da primeira pessoa. É certo que se pode considerar que o facto de algo como a "consciência fenomenal" não poder ser introduzido de forma não circular não constitui só por si razão para a excluir ou para negar a sua existência. Nada exclui a realidade de características apenas apercebidas em primeira pessoa. Mas, precisamente, Dennett não exclui a realidade do apercebimento em primeira pessoa, tudo o que nega são os *qualia* como

qualificativo epistemológico e a descontinuidade da subjectividade consciente relativamente ao resto.

Como já se afirmou, as conclusões pretendidas por Dennett no âmbito do MEM (quer quanto a *zombies*, quer quanto a *qualia*, quer quanto ao Argumento do Conhecimento) só se seguem se já tiver sido aceite a "análise conceptual" da experiência em termos de funções, disposições reactivas e juízos. É de resto de acordo com essa análise que é possível pretender que o epifenomenismo é uma posição absurda. No entanto o epifenomenismo, a ideia segundo a qual propriedades fenomenais, embora causalmente irrelevantes, definitivamente existem, é para muitos filósofos – que não são nem reducionistas nem negam a realidade da consciência, como por exemplo precisamente F. Jackson – a posição correcta quanto à consciência. A sustentabilidade da metafísica e da análise conceptual implícitas no MEM vai portanto ter à questão: será o epifenomenismo definitivamente absurdo, como Dennett pretende? A resposta envolve convicções acerca do que é fundamental no mundo tal como ele é.

A abordagem que por exemplo D. Chalmers faz do problema da consciência é um bom exemplo do facto de ser um choque de intuições quanto ao que é fundamental que está em causa no diferendo entre filósofos que se dedicam à metafísica da consciência, mesmo que a concordância quanto à natureza da cognição seja grande. Em *The Conscious Mind*, Chalmers analisa os vários argumentos encobertos apresentados em CE e o pressuposto básico de tais argumentos. O MEM, com o seu absolutismo da terceira pessoa, pressupõe que a consciência não é fundamental, que ela deve ser caracterizada em termos de funções, disposições reactivas e efeitos. Um tal movimento não é mais do que uma intencionalização da noção de consciência[651]. Por seu lado, teóricos "fundamentalistas" da consciência (como Chalmers, Nagel, Searle e Jackson) pura e simplesmente admitem que a teoria da consciência parte da impossibilidade de uma definição não circular desta e que alguém como Dennett apenas disfarça a utilização implícita da experiência em primeira pessoa. Evidentemente, a posição de Dennett não lhe permite reconhecer que está em jogo um choque de intuições quanto à fundamentalidade da consciência e ele continua a insistir no absurdo das posições a que os qualófilos são

[651] Comum por exemplo na tradição fenomenológica.

conduzidos[652]. De facto as duas partes da discussão não se compreendem e falam de coisas diferentes. De acordo com o MEM, aquilo de que se fala quando se fala de fenomenologia não é senão apercebimento com características determinadas, juízo e texto, especificação e fixação de conteúdos representados. De acordo com os defensores da fundamentalidade da consciência aquilo de que se fala "põe-se" a si próprio, não pode ser introduzido nos termos de outra coisa e deve portanto ser um primitivo da teoria. Como Chalmers afirma, «Dennett lança-me o desafio de oferecer evidência independente para se postular a experiência (presumivelmente evidência comportamental ou funcional). Mas isto é falhar completamente o ponto: a experiência consciente não é postulada para por sua vez explicar outros fenómenos. Ela é um fenómeno de direito, para ser explicado por si próprio. E se vier a verificar-se que ela não pode ser explicada em termos de entidades mais básicas, então ela deverá ser considerada irredutível, exactamente como acontece com categorias como espaço e tempo. De novo, o desafio de Dennett pressupõe que os únicos *explananda* que contam são funções. Já agora, eu gostaria de ver a versão de Dennett da evidência independente que leva os físicos a introduzirem as categorias fundamentais de espaço e tempo. Parece-me que a evidência relevante é espacial e temporal de fio a pavio, assim como a evidência de experiência é experiencial de fio a pavio»[653].

Uma última palavra relativa a modelos cognitivos, já que é isso afinal que o MEM é antes de mais. Mesmo considerando o MEM exclusivamente como um modelo cognitivo é possível argumentar que nada nele afasta uma noção funcional de Teatro Cartesiano (que evidentemente não se identifica com uma noção neuroanatómica, "localizacionista"). De facto, a teoria apresentada inclui uma explicação da unificação ligada à consciência relacionando-a com a percepção interna. Ora, parte do que se entende por Teatro Cartesiano é exactamente essa unificação e percepção interna, que existe na vida mental humana (ou os esboços múltiplos em circulação conduziriam normalmente a vidas mentais múltiplas por cada sistema cognitvo, o que não é normalmente o caso). Alguma coisa como uma teoria do sentido interno sobrevive portanto ao colapso do materialismo cartesiano. Isto mostra, como notam B. Baars e M. Fehling[654], que o MEM vai longe demais na negação de uma função integradora da consciência, que é perfeitamente compatível com o processamento paralelo distribuído e com a sociedade de agentes.

Do que foi dito no presente capítulo conclui-se para já que enquanto teoria da consciência, o MEM não é eliminativista acerca da consciência (o eliminativismo restringe-se aos *qualia*, definidos de forma muito particular) nem pretende abolir a subjectividade (antes explica características desta como a unidade, a centralidade e o auto-apercebimento) e que apesar de tudo a explicação da consciência proposta em CE não é apenas uma investigação das correspondências entre fisiologia ou funcionamentos fisiologicamente baseados e fenomenologia mas uma proposta metafísica acerca do lugar da consciência no mundo, à qual caberia chamar *defesa da superveniência lógica*. Quanto ao pretenso eliminativismo, trata-se antes de intelectualismo, identificação do apercebimento com apercebimento-de. É o intelectualismo que provoca a rasura da especificidade daquilo a que muitos chamam "consciência fenomenal", e que aqui se tem chamado sentir-se ser. De resto, o que está sobretudo em causa na teoria dennettiana da consciência é a negação do "abismo" ontológico entre consciência e não-consciência (o que não é idêntico a negar o sujeito nem a unicidade da vida mental deste). Em termos de argumentação, o caso entre Dennett e os filósofos que se ocupam da metafísica da consciência redunda num *diferendo acerca do que é possível e impossível imaginar*. De acordo com o MEM, é logicamente impossível imaginar *zombies*, espectros invertidos, que Mary aprenda ou que exista algo de irredutível no facto de morcegos se sentirem ser. De acordo com os adversários, tal postura é injustificada.

[652] É evidente que essas posições não aparecem como absurdas aos seus defensores. Aliás Dennett por vezes extrai consequências das posições dos adversários que literalmente não são aceites por estes, como por exemplo quando assume que um qualófilo terá necessariamente que defender que uma máquina computacional não poderá nunca ter experiências (o próprio Dennett defende que pode). Ora, por exemplo D. Chalmers admite sem problemas que uma máquina computacional poderá ter experiências – nesse aspecto é tão funcionalista como Dennett – sem que tal o obrigue a abdicar da defesa da fundamentalidade da experiência no mundo tal como ele é.

[653] CHALMERS 1997: 385.

[654] BAARS & FEHLING 1992.

Capítulo 4

As Pessoas e as suas Acções: a filosofia da mente e os fundamentos da filosofia moral.

«The external view forces itself on us at the same time that we resist it. One way this occurs is through the gradual erosion of what we do by the subtraction of what happens»
Thomas Nagel[655]

4.1 *Pessoa e acção como conceitos normativos: a filosofia moral e a teoria cognitiva. Naturalismo gradualista e compatibilismo.*

As noções que sustentam a filosofia moral – como voluntariedade e racionalidade da acção, unidade e continuidade do agente responsável pela acção – estão suspensas de esclarecimentos que cabem à teoria da mente. Procurar-se-á no presente capítulo identificar e examinar as consequências na filosofia moral, e nomeadamente nas noções de pessoa e de acção, da TSI e do MEM. Embora as noções de pessoa e acção decorram sobretudo da teoria do conteúdo, elas apoiam-se em teorias cognitivas do eu, do controlo e da consciência elaboradas à luz do MEM. Muitas das ideias que permitem a Dennett abordar o problema das pessoas e das acções foram introduzidas nos três capítulos anteriores. Elas serão relembradas no momento oportuno, já que neste capítulo, como no anterior, haverá alguma violência cronológica na exposição.

Há na obra de Dennett toda uma corrente que se ocupa das questões da razão prática[656], na qual se esboça uma posição natura-

[655] NAGEL1979b: 38.

[656] Essa corrente vai desde o tratamento de pessoas e intenções em C&C, a toda a Parte IV de *Brainstorms* (intitulada *Free Will and Personhood)*, na qual aparecem dois artigos sempre apontados por

lista e gradualista acerca de pessoas e acções. A primeira consequência geral da teoria do conteúdo é a *normatividade* de tais conceitos. A análise desse aspecto normativo repartir-se-á pelas análise das condições de pessoalidade (*personhood*)[657] e da identidade pessoal. Entende-se por condições de pessoalidade determinadas características em função das quais uma entidade será ou não uma pessoa. A noção de "condições de pessoalidade" opõe-se desde logo a uma concepção absoluta de pessoa segundo a qual por exemplo um ser humano é sempre e em qualquer circunstância uma pessoa. A dissociação (ou pelo menos a distinção de conceitos) entre ser humano e pessoa tem a sua origem em J. Locke[658]. Concepções absolutas de pessoa e de acção traduzem-se em rigorosas distinções entre pessoa e não-pessoa, acção e não-acção, que Dennett não aceita. É isso que o conduz à posição gradualista de acordo com a qual é impossível separar um conceito a que se pode-

Dennett como fulcrais para a sua posição geral em teoria da mente (*Mechanism and Responsability* e *Conditions of Personhood*), a ER e DDI.

[657] O termo *personhood* é introduzido em DENNETT 1978n, *Conditions of Personhood*. Recorde-se no entanto que Dennett defende desde C&C que o problema da mente não pode ser separado do problema das pessoas, e que a eficácia de razões no controlo da acção é fundamental para a própria existência de pessoas. De BS vem uma nova contribuição para esta tese – de acordo com a qual a voluntariedade é uma condição da pessoalidade –nomeadamente a distinção entre crença e juízo (cf. DENNETT 1978q, *How to Change Your Mind*), já comentada no Capítulo 2. Recorde-se que a intenção do artigo *How to Change Your Mind*, no qual essa distinção é proposta, é marcar a diferença entre estados cognitivos atribuídos no âmbito da TSI aos SI em geral (as crenças) e as opiniões, características da vida mental humana. Com a referida distinção Dennett pretende defender que é precisamente na situação de deliberação que reside a mais relevante diferença entre os humanos (até agora as únicas entidades que são pessoas) e as outras criaturas. Utilizando a distinção entre *belief* e *assent* proposta por R. de Sousa (cf. SOUSA 1971) obtém-se que é o assentimento (*assent*) que caracteriza aquilo que é particularmente humano na posse de crenças, uma vez que as crenças são também atribuídas por exemplo aos animais. No entanto, neles não existe assentimento, e essa falta torna-os criaturas menos livres. A mudança de opinião (*change of mind*) que Dennett procura é portanto um episódio de escolha que não é um desejo mas uma espécie de juízo, e um acto que inicia um compromisso. R. de Sousa chama a esse acto *assenting* e considera que ele é uma aposta. É para o produto do acto de assentimento que Dennett reserva o termo "opinião" enquanto distinta de crença. A existência de crenças não requer criaturas linguísticas, a existência de opiniões sim e a diferença é muito importante para a possibilidade de pessoas. É por exemplo a diferença entre crenças e opiniões que abre o espaço para patologias cognitivas e não apenas morais como a *akrasia* e o auto-engano (*self deception*). A *akrasia* ou fraqueza da vontade é o caso em que o agente, apesar de saber qual curso de acção seria o melhor, opta por outro curso de acção. Também o auto-engano supõe uma divisão do agente "em dois" (aquele que sabe algo e o oculta e aquele que não sabe). A fraqueza da vontade e o auto-engano são os exemplos mais usuais na filosofia da mente de irracionalidade na acção e no pensamento devida à fragmentação da (suposta) unidade do agente. A diferença entre crença e opinião permite finalmente a Dennett isolar uma característica definitória das pessoas enquanto agentes cognitivos de algum modo especiais: o facto de poderem mudar de opinião, i.e. a existência de voluntariedade no controlo de determinadas acções (os seus próprios pensamentos).

[658] Cf. LOCKE [1690], onde J. Locke acentua o facto de "pessoa" ser uma categoria *moral*.

ria chamar *metafísico* de pessoa, de acordo com o qual uma pessoa é uma entidade auto-consciente e racional, de um conceito *moral* de pessoa de acordo com o qual uma pessoa é um agente responsável pelas suas acções. O conceito de acção é igualmente um conceito mentalista: uma acção envolve aquilo que um agente faz – e para isso acontecimentos têm que ser acontecimentos de controlo do comportamento próprio intencionalmente apercebidos pelo próprio agente – por contraste com aquilo que simplesmente acontece. O facto de "pessoa" e "acção" serem conceitos normativos torna impossível a identificação de pessoas com corpos humanos num instante de tempo e acções com os movimentos de tais corpos.

Para além do aspecto normativo ligado às condições de pessoalidade, a noção de pessoa envolve ainda a identidade pessoal. Esta não é estritamente identificável nem com as condições de pessoalidade nem com o eu. O Eu é uma questão cognitiva prévia à identidade pessoal: como se viu no capítulo anterior, o Eu é, de acordo com o MEM, um centro de controlo virtual num sistema de agentes em competição, um Agente Central "eleito" pelos agentes múltiplos de modo a que o sistema global possa representar que coisa no mundo é, distinguindo-se do resto[659]. Entende-se por outro lado

[659] Cf. *The Self as a Center of Narrative Gravity* (DENNETT 1988, inicialmente publicado sob o título *Why Everyone is a Novelist*), *The Origin of Selves*, *The Reality of Selves* (DENNETT & HUMPHREY 1998) *Speaking for Ourselves*, e sobretudo *The Mind's I* (HOFSTADTER & DENNETT 1981), uma obra "composta e arranjada" em parceria com o cientista de computadores Douglas Hofstadter cujo subtítulo é *Fantasies and Reflections on Self and Soul. The Mind's I* é uma antologia de textos literários e filosóficos e ao mesmo tempo uma investigação da noção de eu, culminando numa teoria funcionalista. Segundo um dos autores «O título do nosso livro não é apenas um possessivo mas também uma resposta a: Quem sou eu? Quem é o eu?» (HOFSTADTER 1981b: 408). A proposta apresentada é que o eu é o eu da mente. Os autores reconhecem a atracção da posição dualista quando se lida com o problema do eu e a inultrapassável cisão entre visão subjectiva e visão objectiva de si e pretendem aliás mostrar que os problemas da auto-referência e da cisão entre o ponto de vista subjectivo e o ponto de vista objectivo aparecem não apenas no âmbito do pensamento acerca de si próprio como também na física (cf. o texto 3, *Rediscovering the Mind*, de H. Morowitz), na lógica matemática, nomeadamente em torno das questões gödelianas (cf. o texto 17, *The Riddle of the Universe and its solution*, de C. Cherniak), etc. É tendo em consideração a auto-referência e a cisão entre a visão subjectiva e a visão objectiva que um sistema pode ter de si próprio que deve ser pensado o funcionamento do símbolo-do-Eu (*self-symbol*). A noção de símbolo-do-Eu é a noção central da teoria do eu proposta em *The Mind's I*, que se baseia no princípio funcionalista segundo o qual a matéria de que um efeito é feito é uma questão secundária. Tal independência da inteligência em relação ao seu substracto físico é paradigmaticamente defendida por Turing, no célebre artigo *Computing Machinery and Inteligence* (TURING 1950, Texto 4), comentado em *The Turing Test a Coffe House Conversation* (Texto 5). Naturalmente, o funcionalismo conduz os autores à defesa da IA Forte. Após a profissão de fé funcionalista e a defesa da IA Forte, Dennett e Hosftadter procuram explorar as consequências da aplicação à noção de eu da ideia funcionalista de mente como programa. Uma primeira consequência é (1) a ideia de *mind-tape*: a mente de uma pessoa, aquilo que a caracteriza como sendo exactamente

por identidade pessoal aquilo pelo qual uma pessoa se (re) identifica como sendo uma e a mesma pessoa em cada instante e através do tempo. Formulando a questão à maneira de D. Parfit, o que está em causa são as circunstâncias uma pessoa *sobreviveria*, i.e. se manteria ainda *uma pessoa*, a *mesma* pessoa e apenas *uma* pessoa. O facto de existir *uma* vida mental – uma reunião das experiências a cada instante como experiências de uma só pessoa, o facto de alguém aperceber a vida mental como sua, de forma continuada – sugere que alguma forma de continuidade (a continuidade psicológica ou a continuidade de algum sistema físico[660]) é a pedra de toque da identidade pessoal. A continuidade que aparentemente constitui o

essa pessoa, seria um padrão armazenável e replicável, independente do *hardware* que a implementa. Outra consequência é (2) a irrelevância da distinção entre natural e artificial quando se trata da matéria de que se fazem sistemas cognitivos: não é uma condição necessária para se ter consciência e para se ser um eu que se seja um entidade orgânica nascida de outra entidade orgânica. Outra consequência possível é (3) o espalhamento da implementação do programa (cf. o texto 12, *A Story of a Brain*, de A. Zuboff). Outra consequência é (4) o reconhecimento de que a possibilidade de reimplementação representa uma espécie de imortalidade: se ser uma mente específica é uma *mind-tape*, o que é importante é preservar o padrão. Preservando o padrão preserva-se a pessoa. Poder-se-á mesmo re-implementar o padrão e obter uma duplicação que de facto é a própria entidade (cf. o texto 26, A *Conversation with Einstein's Brain*, de D. Hofstadter e o texto 15, *Beyond Rejection*, de J. Leiber). Se a natureza de algo consiste em ser-padrão e em ser-efeito abre-se a possibilidade da replicação e da (5) indistinção entre cópias e originais. Esta está em causa em situações como o teletransporte de pessoas, por reprodução exacta, partícula a partícula (cf. DENNETT & HOFSTADTER 1981: 3-8), restando a questão do destino e estatuto do "original anterior", que fica para trás. Outra consequência (6) é que a escala espacial e temporal de uma organização não são importantes desde que as suas características relevantes sejam mantidas. Evidentemente, estas consequências são demasiado cruas, e os autores de *The Mind's I* referem-se indiferenciadamente ao Eu, à mente, à consciência e à pessoa, noções que aqui se procurará distinguir mas que obviamente se sobrepõem sempre. O ponto central no entanto é que quando a realidade que se pretende pensar ou simular reside em efeitos e padrões, i.e. é uma questão de nível e de *software*, como é o caso do Eu, as distinções entre simulação e realidade, natural e artificial, ficcional e real, tornam-se problemáticas. Por isso mesmo a in-distinção é o tema de muitas das selecções de Hofstadter e Dennett em *The Mind's I* (cf. nomeadamente o texto 18, *The Seventh Sally*, de S. Lem, o texto 19, *Nom Serviam, de* S. Lem, mas também o texto 12, *The Story of a Brain*, de A. Zuboff – uma "redução ao absurdo da teoria da identidade", DENNETT & HOFSTADTER 1981: 459 – e o texto 15, *Beyond Rejection*, de J. Leiber. No último caso trata-se sobretudo de re-implementação. A ficcionalidade do eu é explorada por Robert Nozick, no texto 27, *Fiction*. Os textos dos próprios organizadores, nomeadamente *Prelude... Ant Fugue* (o texto 11, de D. Hosftstadter), *Where Am I* (o texto 13, de D. Dennett) e *A Conversation with Einstein's Brain* (o texto 26, de D. Hofstadter) pretendem sustentar o materialismo da teoria defendida procurando afastar obstáculos tradicionais à compreensão materialista do sentido do eu (*sense of self*) (DENNETT 1981: 459). É nos comentários dos autores às selecções que aparece a teoria do Eu proposta, de acordo com a qual o eu é a auto-consciência (*self-consciousness*) de um sistema que é um *self*. Como a consciência de alguma coisa é a existência de uma representação interna dessa coisa, para uma entidade ser auto-consciente tem que haver um símbolo-do-Eu "mostrado" a um *self*.

[660] De acordo com o critério da continuidade física, o fundamento da persistência da identidade pessoal é o facto de uma parte suficiente do corpo do indivíduo que tem a vida mental, normalmente o seu cérebro, perdurar.

critério da sobrevivência de uma pessoa como a mesma pessoa esbarra no entanto com diversos problemas, levantados nomeadamente pelo possível variação do suporte físico[661], pelas inúmeras interrupções (normais) da continuidade[662] de si conscientemente apercebida, pela possibilidade de fissão do suporte físico da continuidade[663], pela indefinição das margens temporalmente mais distantes da suposta continuidade psicológica centrada no momento presente, a qual torna as pessoas aparentemente menos apegadas aos estádios de si temporalmente mais distantes[664], e até pelo facto de aparentemente não constituir uma característica fenomenológica específica das experiências o facto de elas serem *minhas*[665]. Não negando a importância das questões da continuidade, a teoria dennettiana conduz a um outro critério da identidade pessoal, um critério normativo.

No que diz respeito à teoria da acção, a posição naturalista e gradualista de Dennett constitui-se como alternativa quer a teorias reducionistas[666] quer a teorias da maximização racional como o utilitarismo e a deontologia[667]. A teoria cognitiva da deliberação na qual as teorias do Eu e do controlo culminam, bem como o desejo de esclarecer as implicações para a ética do facto de os seres humanos serem agentes finitos agindo sob pressão temporal, conduzem Dennett à ideia de "prudência naturalizada", que se opõe à maximização racional. Este ângulo de abordagem da questão da acção reflecte-se na ideia de um manual de "primeiros socorros morais" (*moral first-aid manual*). A proposta do manual significa que para Dennett a teoria moral não deve ser ideal ou idealizante como o são inevitavelmente

[661] Nesta situação (pelo menos até agora apenas imaginada) a mesmidade de uma vida mental experienciada deixa de corresponder à mesmidade numérica de um suporte físico (por exemplo um mesmo cérebro). É para cobrir situações como esta que D. Parfit (PARFIT 1984) fala de causa normais (como por exemplo a persistência de um mesmo cérebro) e causas *não normais* da continuidade psicológica.

[662] Estas interrupções são uma das razões para pôr dúvida a existência de uma resposta definida à questão "será que as experiências foram minhas?".

[663] Cf. por exemplo PARFIT 1984, Capítulo 12, §89.

[664] D. Parfit, que utiliza a noção de conexão psicológica (PARFIT 1984: 206) para referir a ligação entre os estádios unidos na continuidade psicológica, considera que pode existir conexão forte ou conexão fraca entre esses estádios. Apenas no primeiro caso (a conexão forte) se poderia falar de continuidade psicológica propriamente dita.

[665] Cf. DAMÁSIO 1999 para uma sugestão em contrário.

[666] Alude-se aqui às posições às quais Dennett chama *greedy reductionism* (reducionismo ganancioso), de que são exemplo a tentativa skinneriana de uma ética behaviorista "para além da liberdade e da dignidade" (SKINNER 1971) e alguns excessos da sociobiologia (cf. DENNETT 1995: 481-493). Aquilo que Dennett recusa não é, note-se, a ideia de uma análise evolucionista da normatividade ética, tal como se encontra já em autores como T. Hobbes e F. Nietzsche.

[667] Cf. DENNETT 1995, capítulo 17, ponto 1 (*Can Ethics Be Naturalized*).

as éticas da maximização racional mas sim reflectir internamente os constrangimentos práticos da situação de deliberação.

Poder-se-ia considerar, de forma aliás tradicional na filosofia, que o problema geral que unifica todas as questões da razão prática é o problema da vontade livre. No entanto, a estratégia adoptada por Dennett relativamente à vontade é idêntica à estratégia adoptada face à consciência: a vontade não é *um* problema, mas vários problemas, relativos nomeadamente ao controlo, ao Eu, à decisão, às razões na acção, ao "poder ter feito de outra maneira", às atribuições de culpa, que se interpenetram com os problemas da consciência e da racionalidade. Na corrente dedicada às questões da razão prática que atravessa a obra de Dennett o livro *Elbow Room – The Varieties of Free Will Worth Wanting* (ER)[668] ocupa uma posição central e constituirá por essa razão uma referência importante neste capítulo. Em ER Dennett propõe-se explicitar a possibilidade do agente livre a partir de uma teoria cognitiva da agência (*agency*). Na medida em que a análise cognitiva da agência estabelece a possibilidade de deliberadores deterministas, a questão da vontade é afastada do âmbito do determinismo físico para o âmbito da interpretação intencional. Se, como Dennett sugere em ER, aquilo que queremos quando queremos vontade livre é ter controlo e auto-controlo, não tem que existir qualquer incompatibilidade ou contraste entre o determinismo físico e a liberdade. O naturalismo gradualista é assim uma forma de compatibilismo e em ER Dennett acaba por caracterizar a vontade livre de deliberadores deterministas como *auto-controlo meta-reflexivo* e *avaliação forte*[669].

Um último ponto de referência para o percurso que se segue no presente capítulo é o artigo *The Moral First-Aid Manual*[670], no qual Dennett critica a falta de pertinência psicológica e cognitiva das éticas da maximização racional, propondo como alternativa uma ética centrada na ideia de prudência naturalizada.

A convicção que subjaz ao percurso do presente capítulo é que é possível retirar das teorias do conteúdo e da consciência conclu-

[668] DENNETT 1983.

[669] A ideia de avaliação forte é reportada às ideias de C. Taylor acerca da responsabilidade pela identidade própria (cf. TAYLOR 1976: 281). Um avaliador forte é um agente racional que coloca a si próprio as seguintes questões: "O que é que eu quero ser?", "Quero continuar a ser aquilo que sou?". Ele avalia assim não apenas aquilo que *faz* mas aquilo que *é*. Segundo C. Taylor, o vocabulário da avaliação é uma parte importante do vocabulário moral pós-nietzscheano, na maior parte das vezes totalmente esquecida pelas acima chamadas éticas da maximização racional.

[670] DENNETT 1988.

sões relevantes para a filosofia moral. Procurar-se-á fazê-lo a partir de três pontos focais: (i) agência e liberdade, (ii) unidade e identidade pessoal e (iii) deliberação.

4.2 *A fragmentação do problema da vontade livre e as formas que o "problema da vontade livre" não tem: vontade numénica, indeterminismo físico, capacidade mentais não mecânicas.*

> «Liberty or chance (...) is nothing but the want of determination and a *certain looseness* which we feel in passing or not passing from the idea of one to that of the other»,
>
> David Hume, *Treatise of Human Nature*, II, III, 2

Explicitar-se-á antes de mais a forma que o problema da vontade livre não tem. Dennett pensa que não é uma boa estratégia perguntar directamente o que é a vontade livre: deve-se antes perguntar o que se quer quando se quer vontade livre. Ora, aquilo que queremos quando queremos vontade livre aparece-nos fenomenologicamente na deliberação que antecede a decisão e a acção como essa *certain looseness* de que fala Hume, uma indefinição ou indeterminação na maneira como as coisas serão. De acordo com Dennett tal indeterminação não deve ser ligada à questão do determinismo[671] na física. Ao contrário do que alguns autores sugerem a questão física do determinismo não tem relação directa com o problema da vontade. Não têm também qualquer relação com liberdade da vontade as capacidades mentais humanas que estariam para além daquilo que pode ser conseguido por computações mecânicas[672], ao contrário do que sugerem os autores que abordam a vontade livre por meio de

[671] O determinismo é a ideia (metafísica, acerca da natureza do mundo) segundo a qual tudo tem uma causa. Para cada evento E existe um estado anterior e uma lei da natureza tal que dada a lei e o estado anterior, E se segue necessariamente. Se de acordo com o determinismo existe uma explicação causal para tudo aquilo que acontece, o comportamento humano supostamente livre, i.e. as acções guiadas por razões do próprio agente, está também, obviamente, aí incluído.

[672] Cf. DENNETT 1978 o, *The Abilities of Men and Machines*. Para Dennett, argumentos que supõem que Jones-o-humano, ou pelo menos o humano-matemático, transcende os limites do mecanicismo pelo facto de poder pensar aquilo que não é mecanicamente computável são maus argumentos porque (1) não consideram que os constrangimentos lógicos valem não sobre as coisas do mundo directamente mas sobre descrições das coisas, (2) supõem que a determinação das habilidades de Jones e das máquinas de computação pode fazer-se sem petição de princípio quanto à aplicabilidade do teorema de Gödel, i.e. de um modo neutro. Para opinião contrária, cf. PENROSE 1989 e PENROSE 1994. Uma apresentação clássica do argumento encontra-se em *Minds, Machines and Gödel* (J. R. LUCAS 1961).

argumentos que envolvem o teorema de Gödel. As questões da liberdade e responsabilidade de humanos são questões de macro-nível, dizem respeito a um particular entendimento ou interpretação de eventos físicos como acções[673] e não ao *micro-nível* do sistema, explicado pela teoria física. O esclarecimento do estatuto de acções de certos eventos não passa assim nem pela física nem pela lógica matemática, mas apenas pela teoria (normativa) da mente.

Em *Mechanism and Responsability*[674] Dennett pretende mostrar que o determinismo suposto na explicação científica não é nem seriamente contestável nem incompatível com a responsabilidade. Responsabilidade num mundo determinista diz respeito à relação de descrições intencionais com explicações físicas. Ora, tais descrições e explicações apenas aparentemente são conflituais. De acordo com a TSI descrições intencionais pressupõem uma atribuição holista de racionalidade. Não existem sistemas perfeitamente racionais e muitos sistemas não humanos podem ser considerados como racionais (até mesmo termostatos...) no sentido em que podem ser previstos através de atribuições de crenças e desejos. Isto poderia conduzir a concluir que "razões", sendo interpretativas e atributivas, são meramente epifenomenais e em última análise eliminâveis, não fazendo qualquer diferença naquilo que acontece (a alternativa ao epifenomenalismo parece ser a consideração de todas as entidades intencionalmente descritíveis como responsáveis, o que parece ainda mais absurdo).

É verdade que Dennett nunca abdica da ideia segundo a qual a *EI é uma pré-condição do estatuto moral de entidades no mundo*. No entanto, em lugar de essa dependência mostrar a arbitrariedade da "estratégia moral," ela mostra a necessidade de transpor para a teoria da acção ideias acerca da relação entre descrições intencionais e explicações fisicas de sistemas. Como se sabe, de acordo com a TSI, a EI e a EF não são alternativas mutuamente exclusivas. Isto impossibilita desde logo uma concepção absoluta de acção, mas não a existência de acções. O possível determinismo da explicação física apenas se traduz na irrelevância das descrições intencionais se se pressupuser o incompatibilismo, i.e. a ideia segundo a qual liberdade da vontade e determinismo são incompatíveis. Existem, evidentemente, pelo menos *dois tipos de incompatibilismo*: alguns incompatibilistas defendem que a realidade da vontade livre prova que a falsidade do determinismo, outros que o determinismo prova que não existe liberdade da vontade. Os incompatibilistas do primeiro tipo serão daqui em diante referidos como libertários (*libertarians*). Para a TSI, sendo a explicação física a única *explicação*, a única atitude possível perante a liberdade da vontade é o *compatibilismo*, i.e. a ideia segundo a qual a única concepção razoável de vontade livre a mostrará como compatível com o determinismo. No entanto, no artigo *On Giving the Libertarians What They Say They Want*[675], Dennett defende que é possível conceder aos libertários (que são incompatibilistas) o que eles afirmam querer, nomeadamente uma indeterminação anterior à decisão ou à formação da intenção na ausência da qual nunca poderia existir liberdade da vontade.

É usualmente objectado aos libertários que se se considera que as acções livres dependem assim da aleatoriedade, como ninguém pode ser responsável por algo que acontece aleatoriamente, ninguém é livre. Os libertários auto-refutar-se-iam ao quererem proteger a todo o custo a liberdade da vontade. É certo que na sua tentativa de conceber a vontade livre os libertários se inclinam para fora da natureza: de acordo com a doutrina da causação do agente (*agent-causation*), nomeadamente, agir seria provocar determinados eventos sem que o agente fosse causado de modo a causar esses eventos. Dada esta irrupção do agente, a sua insubmissão à causalidade, seria possível um agente "movido" exclusivamente por razões. Ora, Dennett considera a ideia de um agente puramente movido a razões a vários títulos indefensável: «A perfeita vontade kantiana, que seria capaz de responder com perfeita fidelidade a todas as boas razões, é uma impossibilidade física; nem o determinismo nem o indeterminismo poderiam acomodá-la»[676]. A indeterminação desejada pelos libertários não tem que supor nada disso: a indeterminação desejada pelos libertários não é nem o indeterminismo microfísico nem sintoma de uma vontade livre numénica. A indeterminação que importa para a acção é apenas a aleatoriedade da inexistência de padrões, perfeitamente possível num mundo determinista. Antes de mais, aquilo que interessa quando se trata de vontade livre não são movimentos de partículas e nem sequer movimentos esqueléticos de humanos, mas sim acções. Ora, acon-

[673] Esta é uma posição semelhante por exemplo à de P. Strawson, que defendeu que chamamos *livres* às coisas que estamos preparados para considerar como responsáveis mas que não existe nenhuma caracterização não normativa do que essas coisas livres e responsáveis possam ser.

[674] DENNETT 1978m.

[675] DENNETT 1978p.

[676] DENNETT 1983: 49.

tecimentos só podem ser considerados como acções através de descrições *mentalistas*, logo normativas. O comportamento humano poderá por isso ser, de forma independente, determinado ou indeterminado do ponto de vista físico e previsível ou imprevisível intencionalmente. Uma acção pode intencionalmente previsível mesmo que consista numa descrição de eventos resultantes de indeterminismo físico, e um agente pode ser intencionalmente imprevisível mesmo que o determinismo físico seja verdadeiro. Em *On Giving Libertarians What They Say they Want* Dennett imagina uma experiência ilustrativa exemplo destas tese. Trata-se do comportamento de uma caixa de respostas[677], um dispositivo electrónico imaginário que regista respostas de pessoas a questões simples. O dispositivo tem dois botões (sim e não) e dois pedais (sim e não) e um écran dividido ao meio, onde aparece de um lado a instrução "use os botões" e do outro lado a instrução "use os pedais". Apenas metade do écran é iluminada de cada vez na situação experimental e o sujeito é exposto a questões simples, às quais responder "sim" ou "não", seguindo as instrucções do écran. A "escolha" da metade do écran que está em causa de cada vez resulta, de cada vez, de um processo aleatório no interior do dispositivo. É prometido ao sujeito um bónus pelas respostas correctas.

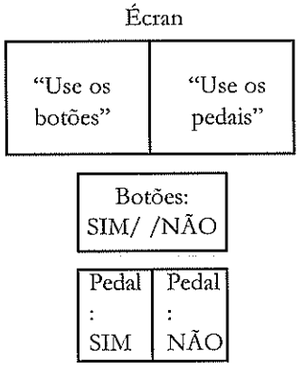


FIG 7. A caixa de respostas

O problema consiste em saber em que situação estariam um físico e um "teórico intencional" no que respeita à previsão do comportamento do sujeito (supondo que este é um sistema fisicamente determinista afectado pela aleatoridade de um processo). A ideia de Dennett é que o teórico intencional prevê melhor, não porque a sua teoria seja melhor mas porque aquilo que interessa não são movimentos físicos e sim acções. Dennett defende que processos semelhantes aos da caixa de respostas (processos indeterminados ou mesmo indeterministas que são intencionalmente previsíveis) subjazem às decisões humanas. Uma acção é implentável de diversas maneiras e factores aleatórios determinam a maneira particular de realizar uma acção intencionada. Nomeadamente, existem "pontos de escolha" possivelmente ubíquos a nível inconsciente que não aparecem fenomenologicamente. Por exemplo, quando se trata de escolher uma de entre as mil maçãs que estão na minha frente quando eu penso "quero uma maçã", não fico dias a ponderar e a calcular qual será a melhor maçã particular que satisfaz o meu querer-uma-maçã. Ora, o facto de a implementação de qualquer acção supor vários destes pontos de escolha indeterminados não torna o agente intencionalmente imprevisível.

A proposta de Dennett é que é precisamente a exploração inteligente da aleatoriedade num mundo determinista a responsável pelo espaço de manobra *(elbow room)* para a acção que queremos quando queremos vontade livre. O espaço de manobra não se identifica assim nem com o indeterminismo físico nem com uma vontade livre não física. O modelo da deliberação e decisão que Dennett propõe é esquematicamente o seguinte:

(1) *Produção aleatória.* Para que exista comportamento de decisão deve existir geração de uma variedade de alternativas.

(2) *Selecção inteligente.* É a avaliação inteligente das alternativas geradas que determina que indeterminações microscópicas serão amplificadas em determinadores macroscópicos de comportamento.

(3) *Pressão temporal* como condição e *procedimentos heurísticos* como estratégia. Agentes sob pressão temporal (como os humanos) são incapazes de considerar exaustivamente as alternativas geradas. É necessária a utilização de procedimentos heurísticos de decisão.

A exploração de modelos de decisão não é um exclusivo de filósofos morais e o modelo de decisão acima esquematizado é, segundo Dennett, um bom modelo por razões totalmente independentes do tratamento filosófico da vontade livre. De resto, o problema da decisão tem uma longa história teórica, que passa

[677]DENNETT 1978p: 290.

nomeadamente pelo cálculo de probabilidades, pela teoria dos jogos e pela lógica temporal. Trata-se de um problema geral de ciência cognitiva, mais tratado até hoje por economistas e estatísticos do que por psicólogos e filósofos. Evidentemente, para além do problema do modelo correcto para o comportamento de decisão de um agente, põe-se ainda o problema (fenomenológico) de saber *o que constitui o nosso sentimento de sermos agentes livres*. Esta é nos termos de Dennett, a questão de saber como é que um processo sub-pessoal aparece ao nível pessoal. Está fora de questão que as tomadas de decisão apareçam ao agente sob a forma do processo sub-pessoal no qual de certo modo consistem. Por exemplo, o agente considera muitas vezes retrospectivamente que existe *uma razão* para a sua decisão. Ora, tendo sido gerada por um processo como o descrito, essa razão não será a Razão *sub specie aeternitatis* ou, numa perspectiva de maximização racional, a razão única. Dennett sugere que o sentido de liberdade do agente a nível pessoal (a *certain looseness*, mais o sentimento de a ter determinado) depende menos da decisão tomada ela própria do que de decisões subsidiárias como a decisão de terminar a deliberação, ignorar outras considerações e passar ao acto.

4.3 *A liberdade num mundo determinista: aleatoriedade, controlo, espaço de manobra (elbow room) e descrição intencional.*

Mesmo que venha a ser possível considerar certos eventos como acções por meio da descrição intencional e escolhas livres devido à aleatoridade, a constatação que se impõe antes de mais a partir do ponto de vista naturalista é que os comportamentos humanos, na medida em que são eventos físicos e comportamentos de entidades biológicas determinadas, são de certo modo limitados, fixos e rígidos e portanto o exacto oposto de livres. Para nomear esta última condição Dennett fala da *sphexishness*[678] do comportamento, em referência à fixidez revelada no comportamento animal, mesmo no comportamento animal aparentemente inteligente. O exemplo dessa fixidez em ER é um comportamento de uma vespa. A revelação da limitação da inteligência da vespa escavadora *Sphex ichneumoneus* dá-se quando se interfere com os passos rotineiros do seu

[678] O termo foi proposto por Douglas Hofstadter, cf. DENNETT 1983: 11.

comportamento[679]. A vespa, quando está na altura de pôr os ovos, escava um buraco e procura um grilo que pica de modo a paralisá-lo sem o matar. A seguir coloca-o no buraco, com os ovos à volta, de modo a que as larvas venham a poder alimentar-se do grilo paralisado. O comportamento parece elaborado e inteligente. No entanto, se os experimentadores retiram o grilo da borda do buraco, interferindo na rotina da vespa, que consiste em descer primeiro, deixando-o ali e voltar para o vir buscar após ter verificado o estado do interior do buraco, todo o comportamento de verificação da vespa é reiniciado. O grilo é trazido de novo para a borda do buraco, e isto acontece tantas vezes quantas for necessário, sem que a vespa alguma vez "pense" e resolva trazer o animal imediatamente para o interior já verificado do buraco. Esta fixidez, esta repetição das rotinas definidoras do comportamento, existe em qualquer organismo, em maior ou menor grau – para não falar em máquinas e programas, sendo exactamente esta uma objecção constantemente levantada em relação à simulação de inteligência em máquinas. A fixidez e a repetição tornam-se, de resto, ainda mais óbvias nas mentes humanas em situações patológicas, por exemplo de mania e obsessão.

4.3.1 *A rigidez, o espaço de manobra e o controlo. Condições da acção: determinismo físico, determinação do design, limitações cognitivas. A impossibilidade física da vontade pura e a sua substituição pela prudência.*

O filósofo moral tem, assim, que lidar com o facto de haver uma grande rigidez naquilo que somos pelo facto de sermos seres físicos e organismos, o que se traduz numa delimitação a priori das possibilidades de movimento e comportamento. Um sistema físico determinado, por exemplo um animal terrestre A de pequenas dimensões, não pode fazer não importa o quê (não pode voar, vencer a gravidade, respirar debaixo de água, suportar pesos de 10 toneladas, sobreviver sem se alimentar). É apenas no seio de uma determinação inicial, função daquilo que o sistema é enquanto sistema físico e sistema desenhado, que ele pode ter mais ou menos espaço de manobra (*elbow room*). A noção de espaço de manobra diz respeito à existência de mais ou menos alternativas possíveis de acção. Se um humano tem muito mais *elbow room* para o seu comportamento do

[679] DENNETT 1983: 10-11.

que uma vespa *Sphex*, isso não significa no entanto que tenha descido sobre ele uma dotação especial de vontade livre ou que ele seja um sistema-para-qualquer-mundo. Não existe um sistema físico capaz de resposta apropriada a toda e qualquer condição ambiental[680]. Nem um humano é totalmente livre no sentido de não ser para um mundo determinado mas sim "para um mundo qualquer", tendo sempre, em qualquer mundo, possibilidade de acção significativa e de espaço de manobra. Ter mais *elbow room* significa apenas ter um mundo menos determinado, mais próximo do "mundo qualquer" do que do mundo extra-determinado (porque muito pouco discriminado) das entidades muito simples que são já agentes ou proto-agentes. Se o leque de possibilidades comportamentais das criaturas começa, em seres simples, por consistir em rotinas rígidas e chega até à "vontade livre", ele nunca perde no entanto o lastro de rigidez ligado à determinação das criaturas e ao determinismo[681].

Mas, «nunca ninguém disse que porque o determinismo é verdadeiro, o termostato não controla a temperatura»[682]. Esta observação de Robert Nozick introduz um outro motivo organizador da teoria da acção apresentada em *Elbow Room*, o motivo do *controlo*. É a ligação dos graus de liberdade de um sistema com a possibilidade de controlo que estabelece a aproximação do problema da vontade livre com as abordagens empíricas da cognição. Para Dennett, aquilo que queremos quando queremos vontade livre é, em larga medida, ter controlo. Não queremos não dispôr de controlo sobre a nossa situação nem sobre nós próprios nem estar sob o controlo de factores externos. É em parte o temor de se ser controlado, manipulado, joguete de forças exteriores, actor impotente numa vida já determinada, que subjaz a fantasias filosóficas como o Génio Maligno de Descartes ou o Neurocirurgião "maligno" das experiências de pensamento dos filósofos da mente. Saber em que consiste o controlo é uma tarefa básica da teoria da acção. Na medida em que apenas é possível controlar estados ou actividades que estão dentro do leque de liberdade de alguma entidade, pode ser, aliás, que não saibamos o que estamos a querer (i.e. pode ser que estejamos a querer aquilo que não faz sentido querer) quando queremos liberdade *tout court*, iniciação *ex nihilo* de não importa que acção. A abordagem da acção através das ideias de controlo e auto-controlo coloca-se evidentemente contra a ideia (kantiana e libertária) de acção livre como resultante de uma irrupção no mundo determinista. Estão em causa característica empíricas e segundo Dennett, determinações empíricas constituem limitações objectivas de dois tipos: (i) determinação física e de *design* dos agentes, (ii) limitação cognitiva. Em oposição ao kantismo e em geral às éticas da maximização racional, como as éticas utilitaristas, o teórico naturalista da acção pensa que não é possível uma "Vontade Pura", se por tal se entende uma vontade infinitamente sensível à voz da razão, que decide que acção empreender após a consideração, na deliberação, de exaustivamente todos os factores relevantes. Não é possível a entidades cognitivamente limitadas considerar exaustivamente todos os factores relevantes[683]. Na terminologia da TSI, reafirmar-se-á, agora num novo sentido, que não existem máquinas semânticas. A máquina semântica seria aqui o Agente Racional ideal, cognitivamente ilimitado. Em primeiro lugar as máquinas semânticas são fisicamente impossíveis porque todos os sistemas físicos estão submetidos à causalidade: "razões" são apenas descrições intencionais. Por outro lado, a razão na acção não é um ditame absoluto que ordena a escolha do caminho único, o caminho racional, o qual seria o caminho escolhido por qualquer ser inteligente. Capacidade de cálculo perfeita, não é plausível em sistemas finitos deliberando em tempo real. Nenhum sistema físico com *design* pode ser um agente racional puro. O agente racional de Descartes, Espinoza ou Kant não pode existir neste mundo. A teoria da acção deve partir da apreciação dos constrangimentos que o tempo real coloca sobre a decisão e da apreciação da maneira como sistemas finitos lidam com esses constrangimentos. Curiosamente, esta abordagem aproxima Dennett do reconhecimento, frequente nas chamadas éticas da virtude da importância da prudência. Simplesmente, a prudência é agora "naturalizada" i.e. considerada como uma característica do *design* de sistemas deliberadores resultantes de evolução por selecção natural.

[680] «There is no way to design a system that can be guaranteed to react appropriately under all environmental conditions (...) the notion of a perfectly rational, perfectly adaptable system, to which all input compatible with its input organs is significant and comprehensible is the notion of an unrealizable physical system» (DENNETT 1978: 244-245).

[681] «In this sense any behavior controlled by a finite mechanism must be tropistic» (DENNETT 1978: 245).

[682] NOZICK 1981: 315.

[683] Considerando a mera verificação da consistência do corpo das crenças do agente, S. Stich (STICH 1993: 152) dá um exemplo esclarecedor: «Consider, for example the *prima facie* pragmatically desirable project of testing one's beliefs from time to time for truth-functional consistency. One familiar way to do this is to use the truth-table method. But, it turns out, this is not only unfeasable for a human brain, it is unfeasable even for what we might suppose to be the ideal, physically buidable computer».

4.4 *Das teoria do controlo ao auto-controlo meta-reflexivo e à avaliação forte.*

Uma vez reportadas as ideias de acção e responsabilidade ao apercebimento das razões e ao envolvimento destas no controlo do comportamento, importa saber em que consiste o controlo e qual é a sua relação com a causação e o determinismo de forma a vir a compreender como é que sistemas cognitivos podem vir a ser caracterizados como "livres". Dennett pergunta o que pensaríamos se assistíssemos ao desenvolvimento de um sistema capaz de alterar e comandar o seu comportamento em virtude da representação de razões para si próprio. Chamar-lhe-íamos um agente racional ou considerá-lo-íamos uma entidade meramente capaz de pensamento inconsciente, de "pensamento comportamental"? É fácil prever o sentido da resposta.

4.4.1 *Deliberação, decisão, oportunidade. Previsibilidade e imprevisibilidade. A deliberação e a decisão: vantagens da insensibilidade e da arbitrariedade. Deliberação e possibilidade epistémica.*

A ideia básica de controlo com que se inicia a teoria do agente racional é a seguinte:

> Definição: A controla B sse a relação entre A e B é tal que A pode conduzir (*drive*) B a qualquer estado do leque de estados de B que A deseje. O controlo de A sobre B testa-se verificando se A consegue ou não conduzir B ao estado que A deseja para B

De acordo com a definição, para algo ser controlado precisa de ter uma variedade de estados possíveis e para algo controlar precisa de ter *desejos* acerca dos estados de qualquer outra coisa. Afirmar que A controla B não é idêntico a afirmar que A pode fazer com que B faça o que quer que seja: só é possível controlar os estados de uma entidade que pertencem ao leque dos "graus de liberdade" desta. Constituem pré-requisitos de um agente racional a existência de graus de liberdade e o controlo.

Um outro problema consiste em saber como é que pode vir a haver conhecimento do controlo. No caso do agente individual, só por experiência e variação ele descobrirá se controla e o que é que

controla. Mesmo assim, é conveniente levar em conta o facto de que, quando se controla alguma coisa (quando por exemplo um piloto controla um avião) não se controla todas as causas que influenciam esta (o piloto não controla por exemplo a gravidade ou os ventos): o controlo é conseguido através da capacidade de previsão (*foreknowledge*) e através do contacto com a coisa controlada. Além disso, a coisa só é controlada se existe *feedback* que informa o controlador da eficácia dos seus empreendimentos. O que interessa retirar daqui é que o *"controlador" pode tomar meta-decisões de modo a aumentar o controlo* (por exemplo, o piloto não conduz o avião em direcção à tempestade, onde o seu espaço de controlo seria muito menor). Este tipo de decisões são estratégias de maximização do *elbow room*.

A definição de controlo acima apresentada apresenta este como uma relação entre entidades individuais. No entanto, o controlo remonta a um ponto anterior à existência de agentes. Psicólogos behavioristas, nomeadamente B. F. Skinner, falam de controlo do comportamento de um organismo pelo ambiente. Ora, o ambiente não é um agente nem tem desejos e na definição apresentada acima a posse de desejos era criterial. Quando um behaviorista fala de comportamentos "sob o controlo de estímulos" refere-se a controlo "sem agente", nomeadamente ao facto de mudanças em A (as características do ambiente que funcionam como "estímulos") serem fiavelmente reflectidas por mudanças em B (organismos, comportamentos). Para Dennett, a utilidade da noção behaviorista de controlo do comportamento do organismo pelo ambiente é chamar a atenção para a transição conceptual (e real) entre causação física e o caso exemplar de controlo que é o controlo por um agente. O ambiente, não sendo um agente, não controla os organismos de acordo com a definição inicial de controlo. No entanto, a selecção natural é responsável pelo *design* das entidades, fazendo com que os seus comportamentos sejam apropriados. Nesse sentido, controla-as. Infelizmente, a utilização behaviorista da noção de controlo propicia uma fusão do controlo com a causação, produzindo uma ideia ameaçadora do controlo. Esta fusão estende-se nomeadamente ao entendimento de variáveis dependentes (que o experimentador *controla*) e variáveis independentes (que o experimentador *não controla*) em situações experimentais. Isolar variáveis dependentes permite que experiências semelhantes sejam interpretadas como revelando a causa do comportamento (a variável independente), aparentemente

eliminando qualquer espaço de manobra do agente. Este entendimento das causas de um comportamento, pondo em relevo o controlo exercido por uma variável independente no ambiente sobre a variável dependente, oculta o facto de o processo não controlado na situação experimental ser tão causado como os processos controlados, diferindo destes apenas por ser *imprevisível*. É fundamentalmente por isso que o processo é incontrolável, por oposição ao controlo que o experimentador exerce sobre variáveis dependentes.

As condições da previsibilidade ou imprevisibilidade de processos causais por um agente podem envolver outros agentes. Assim, estratégias de alto nível de maximização do controlo envolvem não apenas auto-conhecimento mas também conhecimento de outros agentes no ambiente. A importância da teoria dos jogos para a concepção da decisão racional reside em ter mostrado a diferença que a presença de outros agentes faz: a partir do momento em que no ambiente existem outros agentes aquilo que é previsível ou imprevisível muda. Nestas circunstâncias, o princípio geral é que quanto mais informação se possuir, mais "livre" se é. Embora seja indesejável para um agente ser controlado por outros agentes (no limite, ser controlado pelo Neurologista Maligno, que induz em nós crenças desejos e pensamentos...i.e. toda a nossa vida mental) não é possível não ser controlado por nada. Nomeadamente, a "liberdade radical" não poderia consistir numa ignorância total da informação relevante para a decisão e que portanto "controla" a acção. Pode no entanto existir outro *rationale* para a ideia de liberdade radical. Numa situação de coexistência de agentes racionais, o espaço de manobra é conseguido, segundo Dennett, através dos *usos da desordem*. É através do uso da desordem que agentes racionais iludem possíveis controladores (outros agentes que compreenderiam aquilo que o agente é, o que sabe, o que é capaz de fazer). Um agente que não quer que a sua mente seja legível não pode permitir-se revelar nas suas actividades quaisquer padrões. De novo, esta aleatoridade não é o indeterminismo físico mas aleatoridade no sentido da teoria da informação, segundo o qual uma série aleatória é informacionalmente incomprimível. A ausência ou imperscrutabilidade de padrão tem um papel importante nas técnicas de controlo e auto-controlo e as estratégias de maximização do espaço de manobra através da exploração de processos aleatórios dependem em grande parte do auto-conhecimento do sistema que a pratica. Por exemplo sistemas que precisam de decidir para agir e que são

recolhedores imperfeitos de informação, têm necessidade de incorporar despoletadores arbitrários de decisões (como a técnica da moeda ao ar). A imperscrutabilidade dos padrões estará assim em parte ligada a decisões arbitrárias que é racional substituir à ponderação completa quando existe interesse numa decisão rápida. Dennett recorda em *Elbow Room* as análise que Julian Jaynes faz do valor cognitivo das técnicas exo-psíquicas de decisão, tais como o lançamento de uma moeda ao ar ou a leitura de entranhas de pássaros[684]: a importância dessas técnicas é mostrar que alguma coisa, não importa o quê, é necessária para interromper a geração de considerações e despoletar a decisões.

Este tipo de considerações acerca dos mecanismos cognitivos sub-pessoais de deliberação e decisão está nos antípodas de noções como a Vontade Livre kantiana, a Escolha Absoluta sartriana (concebida como criação de si *ex nihilo* em cada momento de decisão) ou o Deliberador Perfeito de um utilitarismo ideal. Na imagem racionalista clássica apenas uma decisão perfeitamente racional, tomada na plena posse pelo agente de tudo aquilo que é pensável, e revelando um único curso para a acção, é considerada livre. Daí a coincidência entre liberdade, necessidade e racionalidade. Ora, uma tal imagem é completamente abalada por considerações empiricamente inspiradas acerca de informação e controlo. No contexto humano, ser racional é ser capaz de tomar decisões em tempo real (e não ponderar inteligentemente *ad infinitum*), e para isso é essencial insensibilidade à informação e um certo grau de arbitrariedade. De um ponto de vista naturalista, a cláusula "uma vez considerados todos os factores" (*all things considered*) na descrição de processos de decisão corresponde a uma descrição irrealista. Se não fosse implausível, nunca deliberações complexas (como por exemplo deliberações morais) dariam lugar a actos. A consideração de tudo o que pode ser relevante ou pertence ao quadro da situação acerca da qual se decide não acontece normalmente, pelo menos na deliberação sã comum, embora situações patológicas possam representar aproximações[685]. Entidades racionais naturais têm que estar desenhadas de modo a ignorar mais informação a partir de certo ponto e a passar ao acto. A insensibilidade informacional pode

[684] Cf. DENNETT 1983: 69 e JAYNES 1976: 245.

[685] Cf. DAMÁSIO 1994 para uma descrição do comportamento de decisão de pacientes com lesões nos lobos frontais, precisamente pacientes nos quais o papel da emoção na racionalidade se encontra perturbado, e que se dedicam a análises hiper-racionalistas das razões para a acção.

parecer rigidez, *sphexishness*. No entanto, sem ela a acção humana não poderia existir. O ingrediente de arbitrariedade revela ainda uma outra diferença relativamente ao modelo clássico da racionalidade na acção, que prevê uma exclusão entre a racionalidade e criatividade: quando se trata da motivação para a acção de seres com capacidade informacional reduzida, a *oposição entre espontaneidade e deliberação racional é totalmente ilusória*. De acordo com o modelo clássico da maximização racional, a racionalidade indicaria um único caminho de acção, tomado por qualquer agente plenamente inteligente. A racionalidade de uma inteligência finita envolve pelo contrário procedimentos heurísticos, e a terminação em certo momento, por arbitrário que seja esse momento, da análise do problema. Isto significa que mesmo o mais racional dos humanos, sendo objectivamente limitado enquanto processador da informação, não pode deixar de considerar por vezes o meta-imperativo racional que ordena "deixar-se ir espontaneamente", i.e. apostar.

4.4.2 *Razões e* self. *O eu e as suas ficções. Real, virtual. Indeterminação. Auto-exortação.*

> «People cannot see their own faces except via external aides that present images – and an image is never quite the same as the original thing. We can come close to seeing and understanding ourselves objectively, but each of us is trapped inside a powerful system with a unique point of view – and that power is also a guarantor of limitedness. And this vulnerability – this self-hook – may also be the source of the inerradicable sense of I»,
>
> Douglas Hofstadter[686]

Se aquilo que queremos quando queremos vontade livre é *auto*-controlo, não é possível compreender este sem uma teoria do *self* e do Eu. O *self*, enquanto princípio de organização pelo qual uma entidade se distingue do resto do mundo, é anterior ao Eu propriamente dito, que é uma questão cognitiva, relativa à organização funcional sub-pessoal. Pessoalidade e identidade pessoal tal como importam do ponto de vista moral e não apenas cognitivo são em grande medida possibilitadas pelo auto-controlo de sistemas nos quais e para os quais funciona um "símbolo-do-Eu" e nos quais é assim

[686] HOFSTADTER 1981b: 278.

criado um centro virtual de controlo e um "centro de gravidade narrativo". A suposição de que o Eu resulta de um funcionamento sub-pessoal afasta a ideia de um Eu pré-constituído. A racionalidade reflexiva associada à pessoalidade e à identidade pessoal apoia-se sobre um Eu que não é originário, antes resultando da instituição de um centro virtual de controlo num sistema de agentes especialistas.

A teoria da consciência apresentada no capítulo anterior envolvia já uma teoria da origem das razões. Quando existem criaturas individuadas, com interesses, passam a existir razões, *free floating rationales* que as relacionam com o ambiente. Uma criatura complexa como já o é a vespa Sphex, tem muitos interesses, age em direcção a fins hierarquizados, a partir dos fins básicos que são perseverar no seu ser e replicar-se. A história da origem das razões não é apenas a história do início da consciência, ela é também a história do início da racionalidade e a "história natural" do agente livre mostrará uma continuidade entre *free floating rationales* e a racionalidade característica de seres auto-conscientes.

Note-se antes de mais, tendo em mente a relação entre um agente e as razões para a sua acção, que por exemplo a ideia kantiana da liberdade de acordo com a qual esta consistiria em dar a si próprio as razões da acção é de certa forma indefensável do ponto de vista naturalista: se "razões" são formas intencionais de descrever processos causais uma teoria da razão na acção só pode edificar-se sobre razões que são já razões nos sistemas (*free floating rationales*) mesmo que ainda não sejam razões para os sistemas (*fully appreciated rationales*). Noutras palavras, a teoria das razões na acção feita de um ponto de vista naturalista inicia-se com a consideração de actos que são inteligentes em função de razões alheias ao actor (e basta o estudo do comportamento animal ou de sistemas artificiais para verificar que esta é uma situação comum). Como já foi dito, agentes racionais reais não são máquinas semânticas, movidas a razões. De que se fala então quanto se fala de "dar a si próprio as razões para a acção"? Antes de mais, é necessário conceber esse "si próprio", i.e. as questões de auto-referência e auto-representação essenciais à criação de uma interioridade mental em sistemas físicos. O que está em jogo é a constituição de algo como uma auto-referência estável num sistema, apesar da instabilidade última que é a não identificação do ponto de vista objectivo com o ponto de vista subjectivo. Constituem situações de auto-referência progressivamente mais conscientes o apercebimento pré-reflexivo de sentimentos próprios, o reconhecimento

de uma imagem de si próprio num espelho, a referência a si próprio usando o pronome em primeira pessoa ("Eu") e o reconhecimento reflexivo de que eu sou o sujeito deste estado mental. A ideia nuclear de uma teoria funcionalista do eu é a ideia segundo a qual o comportamento colectivo de um sistema de indivíduos ou partes pode ter propriedades globais surpreendentes, nomeadamente a propriedade de ser um indivíduo de nível mais elevado (como a Aunt Hilary de D. Hoftsadter, composta pelas formigas da colónia, em *Prelude...Ant Fugue*[687], ou como países ou partidos políticos, organizações facilmente e vulgarmente personificadas). Como qualquer *self*, também o Eu humano é, de acordo com D. Dennett, uma "propriedade" do comportamento colectivo das partes de um sistema, as quais têm finalidades próprias. Mas como é isso possível?

Quando se considera situações de auto-referência é frequente encontrar paradoxos. É também esse o caso com o eu. Como nota Hofstadter: «um objecto tem uma relação muito especial e única consigo mesmo, que limita a sua habilidade de agir sobre si próprio da mesma maneira que pode agir sobre os outros objectos. Um lápis não pode escrever em si próprio (...) uma cobra não se pode comer a si própria. As pessoas não são capazes de ver os seus rostos, a não ser mediante dispositivos externos que apresentam imagens – e uma imagem não é o mesmo que o original (...) Podemos chegar perto de nos vermos e compreendermos a nós próprios objectivamente mas cada um de nós está encerrado dentro de um poderoso sistema com um único ponto de vista – e esse poder é também uma garantia da limitação. Esta vulnerabilidade, este anel de si a si (*self-hook*) pode ser a fonte do inerradicável sentido do "eu"»[688].

Mas que poder explicativo tem a apresentação "objectiva" de si ao si "subjectivo" envolvida no eu? A "explicação pelo Eu" é uma explicação de alto nível num sistema cognitivo e Hofstadter defende que a maneira correcta de a conceber é uma maneira holista e finalista a que chama "*soulist*"[689]. Do ponto de vista cognitivo, o Eu é uma auto-representação, dependente do cruzamento de níveis num sistema, e esse resultado não pode ser capturado "a partir de baixo" e "a partir de antes". O holismo é considerado como causalidade dirigida para baixo (*downward*) e opõe-se ao reducionismo

[687] Cf. HOFSTADTER 1981b e também HOFSTADTER 1979, cujos principais temas são precisamente a auto-referência e os padrões e a relação entre o mental e estes fenómenos.

[688] HOFSTADTER, em HOFSTDTER & DENNETT 1981: 278.

[689] Que se poderia traduzir por animista...

(considerado como causalidade dirigida para cima (*upward*)). Finalismo e preditivismo são maneiras concorrentes de conceber o fluir temporal dos acontecimentos. O finalismo (*goalism*) é um nome para a previsibilidade do "passado" pelo futuro ou fim (por oposição ao preditivismo concebido como previsão do futuro a partir do passado). A opção pelo reducionismo e pelo preditivismo daria uma teoria "mecanicista" do Eu. A escolha do holismo e do finalismo gera o "*soulism*" que interessa defender na teoria do Eu. A existência do eu é uma existência de nível elevado na organização do sistema, dependente das respostas (inseparáveis) a estímulos externos e internos. Um sistema não responde apenas a estímulos externos: "responde" também a si próprio a vários níveis. De acordo com esta perspectiva, é impossível separar a auto-percepção da percepção em geral e da categorização desta por meio de modelos internos. Note-se que não existe qualquer observação interna neste funcionamento. O que mais próximo está da observação interna é o já aludido simbolo-do-Eu (*self symbol*), um modelo interno da totalidade do sistema. O símbolo-do-Eu não é, evidentemente, ele próprio capaz de percepção: *é a activação do símbolo do Eu conjuntamente com os restantes símbolos que constitui a percepção do sistema*[690].

A ideia básica desta teoria do Eu é que mostrar um símbolo-do-Eu a um *self* numa situação em que não existe reconhecimento imediato de uma representação de si pelo sistema (dada a cisão entre ponto de vista subjectivo e ponto de vista objectivo acerca de si próprio) não constitui uma redundância. Assim, algumas das representações para o sistema serão representações do próprio sistema como-se-fosse-a-partir-de-fora e não têm por que se identificar com "o sistema em si" ou o sistema tal como se sente ser. A partir do momento em que o Eu é assim explicado e também considerado, como se viu no Capítulo 3, como um centro de gravidade narrativo, torna-se menos inconcebível não apenas uma certa quebra da ligação do Eu com um cérebro específico como também a existência de vários Eus por corpo. Dennett e N. Humphrey[691] defendem, no artigo *Speaking for Ourselves*, que a primeira observação que se impõe quando se analisa perturbações da identidade é que um Eu resulta de eleições mentais. Isto significa ninguém nasce já com um – e apenas um – Eu. Todos os humanos criam automaticamente narrativas autobiográficas nas quais /das quais um Eu será

[690] HOFSTADTER 1981b: 200.

[691] DENNETT & HUMPHREY 1989.

o centro. Ora, dado o modelo de produção por competição entre agentes sugerido no MEM, não é demasiado estranho que por vezes surja mais do que um Eu. Os vários Eus resultariam de uma anomalia no processo de eleições mentais no sistema. Num sistema cognitivo que é um feixe de agências independentes, que devem comunicar de modo a gerar a representação de unidade, poderão sempre existir problemas de comunicação e de unificação. É esta sugestão que Dennet e Humphrey fazem relativamente à Desordem de Personalidade Múltipla (ou *dissociative identity disorder*). Aliás, de acordo com a ideia de produção e funcionamento de uma representação de unidade (o "símbolo-do-Eu") num sistema de partes ou de agentes em competição não é concebível que um agente cognitivo global, por exemplo um ser humano, seja à partida uma unidade representada para si. E mesmo que alguma vez venha a sê-lo continuará a ser de algum modo (ao nível sub-pessoal) uma amálgama de muitos agentes, cada um com "vontade" própria.

Uma crítica importante e generalizada ao tratamento dennettiano da consciência chama a atenção, como se sabe, para o facto de nesse tratamento a consciência ser rebatida sobre o auto-acesso, os pensamentos sobre pensamentos e o Eu. No entanto, e é apenas isso que se pretende aqui propôr, o que é em geral contestável na teoria dennettiana do Eu não é a pretensão de que o Eu, enquanto representação de unidade, não esteja presente desde o início mas sim a pretensão de que o auto-apercebimento – de forma reveladora, Dennett e Hofstadter em *The Mind's I* tanto utilizam a palavra "mente" (*mind*), como a palavra "eu" (*I*) como a palavra "consciência" (*consciousness*) – envolvido no acesso a si e nos pensamentos sobre pensamentos seja exaustivamente caracterizável de forma funcional e abstracta. De facto, a ideia de apresentação de uma representação de unidade de alguma forma exclui aquilo a que A. Damásio chama a base corpórea do sentimento de si, uma possível incorporação constitutiva do conteúdo do *self* que escapa a esta teoria funcionalista do Eu.

4.4.3 *Do eu à pessoalidade.*

Das teorias funcionalistas sub-pessoais do Eu resulta clara a existência de um fundo irredutívelmente não mental em qualquer entidade que vem a aperceber-se de si própria ao nível pessoal e

moral. Qualquer pessoa tem que ser primeiro um Eu e essa é uma questão de organização sub-pessoal. As ideias de acção e de pessoalidade devem ser extraídas gradualmente desse fundo por uma sequência de aproximações normativas e é exactamente essa a intenção de Dennett no artigo *Conditions of Personhood*[692].

Se é possível abordar as questões do controlo e da representação de unidade de um ponto de vista de terceira pessoa, o caso é diferente com a questão da pessoalidade, devido aos aspectos normativos envolvidos. De acordo com qualquer teoria normativa há dois aspectos a considerar numa entidade "postulada" pela teoria: por um lado, deve existir cumprimento de critérios para a entidade ser o tipo de entidade que é, por outro lado, não existirá adequação integral da entidade aos critérios. Assim, Dennett propõe-se saber se existem condições necessárias e suficientes para a consideração de entidades físicas como pessoas, recusando portanto, por exemplo, afirmar à partida que o auto-agenciamento ligado à moralidade é possível devido à auto-consciência. Isso seria conceder à partida a auto-consciência, sem qualquer explicação, quando o problema que se coloca consiste precisamente em saber exactamente *como é que auto-agenciamento e auto-consciência se relacionam entre si*. Já J. Locke notara que eles se sustentam mutuamente, sustentando ambos a definição de pessoa. De acordo com J. Locke «temos que ter em conta o que é que pessoa representa – e que eu penso tratar-se de um ser inteligente, pensante, que possui raciocínio e que se pode pensar a si próprio como o mesmo ser pensante em diferentes tempos e espaços; é-lhe possível fazer isto devido apenas a essa consciência que é inseparável do pensamento e, pelo que me parece, é essencial para este»[693]. A isto J. Locke ligou o facto de "pessoa" ser (não apenas mas também) «um termo forense que adequa as acções ao seu mérito, e, portanto, pertence apenas aos seres inteligentes, capazes de uma lei e da felicidade e do sofrimento»[694]. Para J. Locke a definição de pessoa conseguida era perfeitamente geral e não se atinha necessariamente aos humanos, até agora as únicas pessoas conhecidas. Os conceitos de pessoa e ser humano não são, assim, identificáveis.

É ao procurar esclarecer a relação entre os dois componentes da definição lockeana de pessoa que Dennett propõe as suas seis con-

[692] DENNETT 1978.
[693] Cf. LOCKE [1690]: 443 (Livro II, XXVII, 11).
[694] Cf. LOCKE [1690]: 459 (Livro II, XXVII, 28).

dições de pessoalidade. Todas elas são condições *necessárias* e no entanto não constituem, em conjunto, condições *suficientes*[695]. As seis condições de pessoalidade são as seguintes: (1) Racionalidade; (2) Ser objecto de atribuições intencionais; (3) Depender de uma estratégia adoptada (a EI); (4) Capacidade de reciprocidade na adopção da EI; (5) Capacidade de comunicação verbal; (6) Auto-consciência.

De acordo com Dennett a ordem de apresentação das seis condições é uma ordem de dependência. O ponto central do artigo é a análise da relação (da possível dependência) da sexta condição com as outras cinco. A pessoalidade seria alcançada através de uma edificação progressiva das condições 4, 5 e 6 sobre as três primeiras condições. As seis condições constituem aliás naturalmente dois grupos de três condições[696]. As três primeiras são interdependentes, e aliás derivam directamente da definição de Sistema Intencional. As três últimas põem em jogo a importância da reciprocidade e da comunicação linguística na auto-consciência e na pessoalidade. Talvez não seja possível considerar que elas são tão independentes entre si como Dennett propõe em *Conditions of Personhood.*

Não basta para capturar a natureza das entidades que são pessoas, afirmar que as pessoas são a sub-classe dos SI constituída pelos seres que não apenas são previsíveis como se tivessem crenças e desejos mas *realmente* os têm. Não basta também evocar a reciprocidade na adopção da EI para marcar a pessoalidade: os humanos não são os únicos SI de segunda ordem, i.e. não são as únicas entidades que para além de serem justificavelmente descritas como intencionais, são justificavelmente descritas como atribuindo intencionalidade a outras entidades, "tendo" portanto não apenas crenças mas crenças acerca de crenças. Vários animais são SI de segunda ordem, capazes, nomeadamente, de comportamento de logro. Ora, se eles são capazes de comportamento de logro, é possível e desejável atribuir-lhes *não apenas a capacidade de acreditar o falso como também o desejo de induzir crenças falsas no outro.* Retomando o exemplo de C&C, existem pássaros que fingem ter uma asa partida para enganar predadores, que assim os consideram presa fácil e os perseguem, afastando-se do ninho onde estão as crias, como era intenção do progenitor. Se o falso não é descritível sem uma disparidade entre aquilo que é acreditado e aquilo que é, na possibilidade de logro está envolvida a situação em que A acredita que q e deseja que B acredite que não-q. Esta situação envolve crenças acerca de crenças, e é uma questão empírica saber se outros sistemas físicos além de humanos têm crenças acerca de crenças. Estes pássaros, por exemplo, tê-las-iam (ou pelos menos seria justificado atribuir-lhes tais crenças acerca de crenças). Mas o que é importante para Dennett em *Conditions of Personhood* é mostrar que a linguagem natural não é uma condição necessária para a existência de um SI de segunda ordem. Criaturas não linguísticas podem ser SI de segunda ordem: a reciprocidade (condição 4) é independente quer da comunicação linguística (condição 5) quer da auto-consciência (condição 6).

Quando se descreve um comportamento fazendo apelo a crenças acerca de crenças, como acontece com o comportamento de logro do pássaro, descreve-se uma *ordem que está lá*, mas que não corresponde a intenções conscientes dos intervenientes na situação nem a eventos actuais. Trata-se não de inteligência *do* agente mas *no* agente, de um mundo proto-voluntário que não é pessoal. O facto de os comportamentos que envolvem tais razões poderem ser rigidamente fixados, instintivos, em tantos animais, mostra que não é necessário que as crenças e desejos do SI estejam representadas explicitamente para o agente para a previsão do comportamento inteligente deste ser possível. Apesar disso, a reciprocidade na atribuição de crenças acerca de crenças é um grande passo em direcção ao tipo humano de pessoalidade.

Em *Conditions of Personhood* Dennett considera que a análise griceana da significação[697] é esclarecedora relativamente à diferença que uma linguagem natural (a condição 5) faz num SI. O que é particularmente esclarecedor é a célebre ideia de significação não natural (*non natural meaning*), que Grice define como a situação em que um enunciador quer dizer alguma coisa enunciando alguma outra coisa, devendo este querer-dizer (*mean*) ser analisado em termos das intenções do enunciador (*utterer's meaning*). As análises pragmáticas de Grice são importantes na medida em que põem em evidência características que as mentes de criaturas linguísticas possuem *por serem linguísticas*, nomeadamente serem SI de terceira ordem. Para existir significação não natural, um sistema A deve pretender que

[695] Segundo Rovane (ROVANE 1994: 357), isto acontece porque em *Conditions of Personhood* Dennett faz uma *caracterização kantiano-rawlsiana de pessoalidade moral*, o que o leva a não considerar a variedade de imagens morais incompatíveis disponíveis para as pessoas. É sobre esta crítica que assenta a proposta que Rovane faz de um critério ético de pessoa, que não incorra nos defeitos do (insuficiente, de acordo com Rovane) critério moral de pessoa apresentado por Dennett.

[696] É a opinião de Rovane (ROVANE 1994).

[697] GRICE 1957.

um sistema B reconheça que A (in) tenciona que B produza uma dada resposta. Ora isto mostra que é a condição 5 (a linguagem) que depende da condição 4 (a reciprocidade) e não vice-versa. Em segundo lugar, mostra que a segunda ordem (as crenças acerca de crenças) não é suficiente como modelo da reciprocidade usual de tipo humano, nomeadamente em humanos que utilizam linguagens naturais. É preciso mais, é preciso o funcionamento de um reconhecimento tácito, um "encontro" de subjectividades, expresso aqui pela ideia de SI de terceira ordem. É este reconhecimento tácito que subjaz aos casos de Grice. Este pode ser caracterizado como uma expectativa mútua de cooperação, mais especificamente como uma expectativa de verdade, relevância, clareza e não ambiguidade quanto ao que é dito. Grice pretende mostrar que se a comunicação linguística humana envolve a possibilidade de manipulação é porque ela depende da confiança. Só a existência de expectativas determinadas permite a subsunção de intenções, a introdução de um outro discurso nas entrelinhas daquilo que é dito[698]. Deve existir uma norma em acção para este funcionamento ser possível e essa norma é a sinceridade. Dennett pensa que a análise griceana da comunicação evidencia o papel normativo (inexplícito) da sinceridade. Dizer o verdadeiro é mais normal na comunicação humana (evidentemente não no sentido voluntarista que envolveria a implausível ideia de que os humanos fazem de propósito para serem sinceros quando falam). O sucesso da manipulação linguisticamente possível não depende obviamente do facto de o entendedor se dedicar conscientemente a estes raciocínios de procura de intenções encaixadas (*nested intentions)*. Aliás, o pássaro que finge ter a asa partida também não intencionava conscientemente o seu logro. Nem quem fala nem quem compreende precisa de se dedicar conscientemente à decifração das intenções encaixadas de Grice. *O importante nas análises de Grice é o facto de elas revelarem a incontornável presença da terceira ordem na interacção de SI que comunicam linguisticamente*

[698] Cf. um exemplo de Steven Pinker (PINKER 1994: 229): «Dear Professor Pinker: I am very pleased to recommend Irving Smith to you. Mr. Smith is a model student. He dresses well and is extremely punctual. I have known Mr. Smith for three years now, and in every way I have found him to be most cooperative. His wife is charming. Sincerely, John Jones, Professor». O destinatário da carta poderia lê-la como uma recomendação «stay away from Smith, he's dumb as a tree» (PINKER 1994: 229), muito embora a carta contenha apenas asserções factuais e nada diga nesse sentido. A pertinência das análises griceanas mostra até que pontos somos «scheming, second-guessing, social animals» (PINKER 1994: 230), característica esta que constitui a base do humor, da metáfora, dos eufemismos, da etiqueta, etc.

entre si. Mas as intenções griceanas são inconscientes e Dennett está a defender, basicamente, que, mesmo se elas representam um novo aprofundamento da proto-agência, elas poderiam existir sem auto-consciência. Isto significa que todas estas condições da proto-voluntariedade são obtidas sem qualquer apelo à consciência. Dennett propõe inclusivamente que com base apenas nestas cinco condições é possível elaborar uma teoria da origem da ética (uma ética para seres inteligentes inconscientes, note-se). A ideia está aliás implícita por exemplo na teoria rawlsiana da justiça, considerada como uma parte da teoria da escolha racional. Como é sabido, de acordo com a teoria da justiça de Rawls[699] o acordo quanto aos princípios da justiça *não supõe qualquer bondade ou moralidade prévias das partes contratantes.* A solução óptima é apenas a solução racional para calculadores racionais interessados no seu próprio bem. Esse dever não é já moral, embora venha a gerar a moralidade, mas apenas racional. Nenhuma moralidade é pressuposta na escolha: trata-se apenas de um imperativo de racionalidade. Obtém-se então que assim como acreditar o verdadeiro é a norma da crença[700] e dizer o verdadeiro é a norma da enunciação[701], tratar o outro como pessoa (i.e. a justiça, ou pelo menos os princípios racionais rawlsianos da justiça) é a norma da interacção pessoal[702]. Que esta norma é uma norma, i.e. algo em relação ao qual os desvios são constantes é absolutamente claro.

Repare-se agora no dia seguinte. Nada na exploração das primeiras cinco condições explica como podemos por vezes considerar pessoas individuais como responsáveis pelas suas acções. O que foi afirmado serviria até para fornecer argumentos no sentido de defender que nunca o podemos fazer: até mesmo a justiça parece poder existir como regulação da interpessoalidade sem qualquer auto-consciência e portanto sem qualquer (ir)responsabilidade dos agentes. A pretensão de Dennett de separar a condição 6 das condições 4 e 5 é sem dúvida muito forte. C. Rovane[703], por exemplo, comenta que as três últimas condições de pessoalidade se põem conjuntamente e que as condições 4 e 5 só podem referir-se a relações que são já interpessoais, no

[699] RAWLS 1971.

[700] Cf. Capítulo 2. A ideia é defendida não apenas por Dennett como também por Davidson, Fodor, etc.

[701] GRICE 1957.

[702] Rovane nota, quanto a este princípio kantiano, que é de toda a conveniência um recuo relativamente ao kantismo que abra a porta ao desacordo ético entre as pessoas.

[703] ROVANE 1994.

sentido em que envolvem a captação dos princípios normativos que constituem a racionalidade reflexiva, ligada exclusivamente à condição 6 na proposta de Dennett. De qualquer modo, a racionalidade reflexiva constitui o topo da construção da pessoalidade descrita em *Conditions of Personhood*. Isto significa que a auto-consciência (a condição 6) é mais do que auto-apercebimento, auto-referência e auto-representação, mais do que ser considerado como racional e intencional por outras entidades que por sua vez se considera como racionais e intencionais: a *auto-consciência é pelo menos em parte auto-avaliação, algo de que um comunicador griceano pode não ser capaz*.

É através das ideias de Harry Frankfurt[704] acerca da ligação entre a liberdade na acção e o conceito de pessoa que a última condição de pessoalidade, a auto-consciência, fica esclarecida como racionalidade reflexiva e auto-avaliação. H. Frankfurt admite que ser livre aparentemente consiste em fazer aquilo que se deseja. Nota, no entanto, que é possível ser-se livre de se fazer aquilo que se deseja sem gozar de liberdade da vontade. Ora, é esta última que é crucial para o conceito de pessoa. Um viciado deseja o objecto do seu vício e no entanto, de acordo com H. Frankfurt, a sua vontade não é livre. A liberdade da vontade supõe que se seja livre para se querer o que se quer querer e portanto só pode existir em seres capazes de formar volições de segunda ordem, seres capazes de quererem que um determinado desejo seu seja (ou não seja) a sua vontade (no exemplo do viciado não existiria essa liberdade de querer ou não querer). Um comportamento de tendência para o objecto do desejo apenas constitui uma acção livre se envolver volições de segunda ordem. Dennett defende que apenas com o auto-controlo resultante da avaliação reflexiva que entra assim em jogo existe a auto-consciência que importa para a pessoalidade. De acordo com a sugestão de H. Frankfurt, é a incapacidade de deliberar racionalmente sobre o querer, e não o determinismo, que pode impedir a liberdade de acção. Ora se a liberdade de acção é a capacidade de deliberar racionalmente sobre o querer, ela envolve em termos da EI mais do que estados mentais de terceira ordem[705]. Apa-

[704] FRANKFURT 1971.

[705] Dennett abandona a terminologia das "ordens dos SI" quando chega à sexta condição. C. Rovane nota de forma perspicaz que entretanto ele começara a utilizar *dois* critérios diferentes para avaliar a sofisticação da racionalidade do agente: a *ordem dos SI* mas também o *envolvimento em actividades de avaliação*. Ora, estas actividades não requerem necessariamente ordens muito mais elevadas do que as já mencionadas. O segundo critério de Dennett não é de modo algum continuista e de certo modo contradiz a ideia segundo a qual as capacidades racionais e sociais são uma questão de grau, não existindo nenhum corte qualitativo. De facto, há um único tipo de seres envolvidos em actividades de auto-avaliação.

rentemente ela envolveria uma capacidade ilimitada e recursiva de auto-referência, uma capacidade de pensamentos sobre pensamentos sobre pensamentos sobre pensamentos, etc, e de desejos acerca de desejos acerca de desejos, etc. A estrutura não é no entanto ilimitadamente iterada, como se verá.

O controlo resultante de avaliação reflexiva supõe a adopção em relação a si próprio, por um sistema físico, do papel de exigidor de razões. Se existisse um eu naturalmente e essencialmente uno ou uma pessoa "absoluta" este auto-controlo meta-reflexivo através de razões pareceria redundante. Mas, precisamente, nenhum eu uno pode ser dado como garantia a priori, pois um Eu é ele próprio um efeito do funcionamento sub-pessoal do sistema. As "pessoas" pressupõem sistemas constituídos por um tal processo de centralização virtual. De acordo com Dennett, *é nestas circunstâncias que o auto-controlo meta-reflexivo deve ser considerado como a auto-consciência genuína*. Esta encontra-se assim indiscernivelmente ligada à agência racional: uma pessoa é uma pessoa porque é um agente racional, um auto-avaliador e um auto-controlador meta-reflexivo. Uma parte importante da pessoalidade é o pensamento acerca de crenças e desejos próprios, ligado ao poder de os interromper e controlar.

Dennett virá a aceitar uma crítica de C. Rovane[706] segundo a qual a agência racional de que se fala em *Conditions of Personhood* deve incluir um compromisso de produzir juízos caracterizados pela cláusula "uma vez considerados todos os factores" (*all things considered*). É o compromisso de entidades com esse tipo de juízos que faz delas entidades propriamente racionais. De acordo com Rovane, esses juízos são constitutivos de algo que Dennett persegue e não chega a explicitar em *Conditions of Personhood:* a relação entre a racionalidade e a unidade de um ponto de vista racional, que tornará a racionalidade racionalidade *própria*. A questão será retomada mais à frente. O artigo *Conditions of Personhood* culmina com a defesa da ideia segundo a qual apenas em situação de auto-controlo meta-reflexivo a "ordem que está lá" e que pode ser de tão variadas maneiras não consciente, não está lá a não ser que esteja em episódios de pensamento consciente. Esta é a razão da continuidade entre os conceitos moral e metafísico de pessoa, por oposição à concepção absoluta ou essencialista de pessoa.

[706] ROVANE 1994.

4.4.4 *"Quererei eu realmente ser aquilo que sou?". Os meta-problemas difíceis do controlo, a avaliação forte e a sorte moral.*

«Thinking afects our thoughts. People who program computers encounter similar problems when new programs malfunction because of unexpected interactions among their parts (any such machinery must get confused by self experiments that try to find out how it works – since such experiments must change the very records thay are trying to inspect!»,
Marvin Minsky, *The Society of Mind*, 6.3

«Nothing is entirely one's own making, unless one is God»,
Daniel Dennett, *Elbow Room*[707]

O capítulo *Self-made-selves*[708] de *Elbow Room* tem como propósito reportar a constituição da identidade pessoal no sentido moralmente relevante à possibilidade de estratégias de controlo racional reflexivo sucessivamente mais sofisticadas. Supondo que o Eu é uma auto-representação de unificação que acompanha os processos de meta-controlo, o perigo de uma explicação totalmente cognitiva e funcional é, do ponto de vista do filósofo moral, não restar nada nem ninguém acima do funcionamento, nenhum eu para ser "responsável pelo eu" nas palavras de C. Taylor[709]. Não é certamente como um eu superficial e ficcional que nos apercebemos de nós próprios: o nosso apercebimento coloca-nos mais propriamente como primeiros motores não movidos, de forma semelhante ao que pretende a doutrina libertária da causação do agente (*agent causation*). Nada garante à partida, no entanto, que esse apercebimento não seja totalmente ilusório. Se se trata de uma ilusão, é preciso saber antes de mais o que a causa e de um ponto de vista cognitivo, Dennett oferece três razões: (i) uma ilusão de escala no nosso auto-apercebimento, causada pela amplificação dos resultados do processamento de informação no sistema nervoso[710]; (ii) uma

[707] DENNETT 1983: 85.
[708] DENNETT 1983: 74.
[709] Cf. TAYLOR 1976.
[710] A introspecção mostra-nos "efeitos dramáticos" e notórios em nós próprios, tais como movimentos intencionados do corpo, e portanto grandes gastos de energia armazenada iniciados e controlados por processos que utilizam muito pouca energia como input e que são relativamente imperceptíveis. No entanto, segundo Dennett, todos os processadores de informação se apoiam em amplificadores, de modo que pequenas causas produzem grandes efeitos. Isto significa que nós nos observamos em grande escala, em escala amplificada, a fazer coisas acontecer e que o efeito é misterioso pois não conseguimos seguir introspectivamente os trajectos causais.

má avaliação natural da fenomenologia da decisão e do controlo que nos conduz a postular um eu activo; (iii) o facto de as identidades pessoais resultarem de um processo de auto-criação, i.e. de auto-avaliação[711]. Ora, é precisamente este último aspecto que possibilita o cruzamento das questões cognitiva do Eu e do controlo com a identidade pessoal. De facto, a auto-criação da pessoa consiste num processo de auto-avaliações sucessivas e Dennett convoca as ideias de C. Taylor acerca de avaliação forte e responsabilidade pela identidade própria de modo a caracterizar esse processo. Ao problema da avaliação forte chamar-se-á aqui *o problema difícil do controlo*[712].

Taylor contrasta a avaliação forte com o tipo de avaliação de que é capaz qualquer entidade desejante. No último caso, para que alguma coisa seja boa basta que seja desejada, enquanto que no primeiro caso alguém pode mesmo chegar a julgar como indesejáveis certos desejos que experimenta. A avaliação forte é assim uma proposta acerca da natureza de seres que cuidam do seu ser, que o determinam e o avaliam, e ela captura, segundo C. Taylor, um traço da moralidade esquecido com a predominância da caracterização utilitarista do agente. O avaliador forte faz mais do que uma pesagem dos seus desejos: ele liga a avaliação dos seus desejos com a sua própria identidade, tornando-se responsável pelo seu próprio ser. A auto-avaliação supõe representação e avaliação das crenças e desejos próprios, e nela estão envolvidas, na terminologia de H. Frankfurt, *volições de segunda ordem*. É com a auto-avaliação que a questão cognitiva do auto-controlo se estende ao domínio ético. São processos de auto-avaliação que possibilitam a pessoalidade e a identidade pessoal (as pessoas profundas, na terminologia de C. Taylor). Um avaliador forte, alguém que pergunta "eu quero mesmo ser aquilo que agora sou?", que escolhe e que sabe que escolhe sem necessidade por entre a pluralidade das alternativas, é, segundo C. Taylor, aquilo que se entende por "pessoa". C. Taylor liga assim a noção de "pessoa" à responsabilidade por aquilo que se é, explicitando aquilo que na identidade pessoal resulta de voluntariedade e compromisso. Segundo C. Taylor, a responsabilidade pela avaliação forte não se identifica com a escolha radical, a escolha que

[711] I.e. resultarem mais ou menos, conforme se trate de pessoas mais ou menos "profundas". A noção de "pessoas profundas" é de C. Taylor (cf. TAYLOR 1976: 299): «We consider people deep to the extent, *inter alia*, that they are capable of this kind of radical self-reflexion». De acordo com Taylor, é a auto-reflexão ou avaliação forte que torna alguém responsável pela pessoa que é.
[712] Por analogia com o problema difícil da consciência analisado no Capítulo 3.

não é fundada em qualquer razão, através da qual por exemplo J. P. Sartre caracteriza o projecto que cada pessoa é: ao contrário do avaliador forte, o sujeito da escolha radical sartriana não escolhe de acordo com razões. C. Taylor crê que é sempre possível avaliar mais as crenças e os desejos próprios, e as crenças e desejos próprios acerca de crenças e desejos próprios, etc, embora tal se vá tornando progressivamente mais árduo. No entanto, quando o problema passa a ser *avaliar as próprias estratégias de avaliação*, a prática sem limites da auto-avaliação parece tão irracional como a rejeição rígida de qualquer auto-avaliação. É bastante problemático saber se uma auto-avaliação sempre crescente melhorará algo a que se poderia chamar o "carácter" do agente. Dir-se-ia que o auto-conhecimento em excesso não pode ser bom para nenhum sistema. C. Taylor considera aliás que são precisamente as mais profundas auto-avaliações, aquelas que constituem as pessoas "profundas", as mais sujeitas a distorção e que mais se arriscam a provocar o abalo ou desmoronamento do auto-avaliador[713]. É quando se está mais próximo daquilo que se é que mais facilmente ocorre a dissolução da identidade, uma vez que uma tal auto-observação permite constatar o carácter criado e contingente daquilo que se observa. De facto, na medida em que não está disponível uma meta-linguagem que permita a avaliação das avaliações mais básicas, o agente moral não pode "saltar para fora de si próprio". É por essa razão que C. Taylor defende que *avaliações fortes não são descrições mas articulações*, tentativas de formular o que é inicialmente não formulado. No entanto, as avaliações fortes não são arbitrárias no sentido em que as escolhas radicais sartrianas o são, mas apenas no sentido em que não está disponível uma meta-linguagem para as avaliar. De acordo com Dennett, a razão para a falta de fiabilidade da auto-avaliação de nível mais alto é suficientemente clara: agentes finitos vêem-se compelidos a utilizar métodos heurísticos quando se encontram perante espaços de busca muito grandes, e esse é precisamente o caso do espaço de busca das escolhas morais, cuja contingência extrema a auto-avaliação revela.

A ideia de criação da identidade pessoal através da avaliação forte é evidentemente vulnerável às críticas que chamam a atenção

[713] «It is precisely the deepest evaluations which are least clear, least articulated, most easily subject to illusions and distorsion. It is those which are closest to what I am as a subject in the sense that short of them I would break down as a person, which are among the hardest for me to be clear about» (TAYLOR 1976: 296).

para o papel da sorte moral na constituição das pessoas. A "sorte moral" refere-se àquilo com que o agente já conta ou não conta – em termos de carácter, de circunstâncias, etc – quando se trata da constituição de si através de acções. Despossessão e involuntariedade, mais do que a voluntariedade do avaliador forte, caracterizariam nesta perspectiva, a constituição da identidade pessoal: é, em grande parte, uma questão de sorte ter-se vindo a ser aquilo que se é e ter-se agido como se agiu perante determinadas circunstâncias. Assim, a auto-criação acima tratada seria sobretudo questão de sorte e parece descabido falar de culpa ou mérito de agentes quando o que está em causa é a sorte, i.e. a intervenção de factores que escapam totalmente ao controlo.

O motivo da *sorte moral* foi tratado por B.Williams[714] e T.Nagel[715]. Fala-se de sorte moral quando um aspecto significativo daquilo que alguém faz depende de factores fora do seu controlo, continuando a pessoa a ser tratada como objecto de juízo moral. O problema é, evidentemente, que os factores fora de controlo da pessoa se alargam sem cessar, parecendo não restar qualquer espaço para o agente e para a sua vontade livre. T. Nagel considera quatro modalidades de sujeição à sorte dos objectos de avaliação moral (i.e. das acções praticadas): *o tipo de pessoa que se é* (inclinações, capacidades, temperamento), *as circunstâncias em que nos encontramos, as causas da acção* e os *efeitos da acção*. Por exemplo, um condutor embriagado que guia por uma estrada deserta e subitamente atropela e mata uma criança, poderia não ter morto criança alguma, mesmo guiando embriagado durante horas, se a criança não tivesse surgido naquele preciso momento. Um exemplo simétrico poderia ser dado com um condutor absolutamente sóbrio que também atropela uma criança. A ideia de sorte moral é assumidamente anti-kantiana (será mais propriamente kafkiana...). Uma das características do kantismo moral é afastar por princípio toda a consideração acerca de sorte (especificamente afastar da avaliação moral toda e qualquer contingência exterior à própria determinação racional da acção pelo agente). De acordo com o kantismo, o âmbito da avaliação moral restringe-se exclusivamente ao querer e às regras racionais que regem este. Como nota B. Williams[716], a atracção do kantismo é parecer justo, fazendo crer que todo o agente racional está em idênticas condições

[714] WILLIAMS 1981.
[715] NAGEL 1979b.
[716] WILLIAMS 1981: 21.

para a prática de uma vida moral bem sucedida. No entanto o kantismo parece também, à luz de exemplos de boa ou má sorte moral, estranhamente cego e irrealista. O teórico que chama a atenção para a sorte moral pensa evidentemente a "justiça" do kantismo é ilusória: as acções são moralmente avaliadas de uma forma que inclui factores dependentes da sorte, mesmo que isso pareça injusto ou irracional. Num tribunal, por exemplo[717], um réu será julgado por um crime diferente (homícidio ou tentativa de homicídio) dependendo de um factor totalmente fora do seu controlo, como o facto de a pessoa alvejada por si usar ou não um colete à prova de bala no momento em que foi alvejada. Se a pessoa não usava o colete, ele será um assassino, se a pessoa usava o colete terá apenas tentado matar alguém. Ou, recordando um exemplo chave de B. Williams[718], a atitude de Gauguin que abandonou a família para pintar será sempre avaliada pela qualidade e beleza da obra que assim produziu. Apenas o sucesso justificará Gauguin. Se ele falhar, terá agido sem justificação. Se não falhar, as coisas serão diferentes. Mas no momento em que decide, Gauguin não pode obviamente prever se o sucesso acontecerá ou não. A justificação é retrospectiva, as considerações não podem ser aplicadas no momento da deliberação. Trata-se precisamente de sorte.

4.4.5 *Sorte Moral ou responsabilidade. Deliberação. A prudência e a design do deliberador.*

Dennett pensa que um ênfase excessivo à sorte moral nos faz correr o risco de apagar o espaço de manobra entre sucesso e fracasso de acções. É certo que os agentes humanos não são totalmente responsáveis por serem responsáveis. No entanto, dispõem de oportunidades, que convergem na situação de *deliberação*. A própria ideia de deliberação seria absurda se não existisse um espaço de possibilidades. O problema real para Dennett é saber de que tipo de possibilidade se trata aqui e ele propõe que na acção se trata de *possibilidade epistémica* e não de possibilidade lógica ou possibilidade nómica.

Afirmar que a possibilidade que subjaz ao *design* do deliberador é possibilidade *epistémica* é afirmar que a situação de deliberação pressupõe que o deliberador distingue estados de coisas que conhe-

[717] O exemplo é de T. Nagel, em NAGEL 1979b: 29.
[718] WILLIAMS 1981: 22-26.

ce ou quer conhecer de outros estados de coisas. A distinção dá origem, na imagem manifesta dos deliberadores, às potencialidades das coisas. A imagem manifesta de um deliberador está ligada à sua maneira de recolher informação de modo a poder agir sobre o mundo. Na imagem manifesta do deliberador existe um fundo do qual se destacam estados de coisas possíveis, compreendendo estes estados incontroláveis e imprevisíveis, estados controláveis e estados previsiveis. Assim, do ponto de vista do deliberador, "agir sob a ideia de liberdade" significa em termos concretos que ele: (i) prevê o ambiente de forma rápida e confiável, o que é uma tarefa de tratamento de informação; (ii) considera no ambiente traços fixos, traços desprezáveis e traços mutáveis dos quais vale a pena manter a pista, sejam eles previsíveis em certas condições, sejam caóticos. O ponto da gestão de informação é chegar a tomar uma decisão que seja boa antes que seja tarde demais. Essa decisão baseia-se nos desejos do deliberador e em expectativas quanto ao que acontecerá de qualquer modo e quanto ao que acontecerá se algo não fôr feito. A auto-previsão é um componente importante nesta deliberação. De entre os eventos imprevisíveis há a considerar também os resultados das deliberações próprias. É nisto que consistem os futuros abertos e o espaço para a própria deliberação. Como já se afirmou, a deliberação e a decisão não são apenas ocasião para problemas éticos de agentes morais humanos, são também capacidades de agentes cognitivos não humanos. Deliberação e decisão são problemas para o *design* de artefactos inteligentes. Saber, nomeadamente, como investir o tempo da deliberação em algo que não seja irrelevante é um problema fulcral para o *design* do deliberador. Sem a possibilidade de saltos de imaginação o deliberador será *sphexish*, rígido, para utilizar o termo de Dennett. No entanto, com a possibilidade de saltos da imaginação demasiado grandes (pense-se por exemplo num deliberador cuja tarefa é pintar um quarto e que para tal considera, juntamente com milhares de outras alternativas, virá-lo de pernas para o ar) ele não será um deliberador eficiente. De acordo com Dennett, a chave da questão é verificar que deliberadores num mundo determinista têm *oportunidades, i.e. ocasiões para decidir alguma coisa que faz uma diferença*. Para Dennett a questão da realidade das oportunidades é absolutamente independente do determinismo físico. Oportunidade não é um termo metafórico: a ideia de oportunidade deriva da distinção entre evitável e inevitável na previsão do deliberador. Todo o deliberador

delibera sobre um fundo de antecipação do futuro, no quadro de uma determinada imagem manifesta.

Considera-se usualmente que "poder ter feito de outra maneira" (*could have done otherwise*) é um critério pacífico da responsabilidade de um agente. No entanto, Dennett defende que o princípio é de alguma forma falso e pretende prová-lo através da análise dos sentidos de poder (*can*). O ponto de Dennett é que não tem qualquer interesse procurar responder à questão "poderia (eu) ter feito de outro modo?" por exemplo após um acto lamentável. Se esta é uma questão metafísica acerca de cursos alternativos do mundo, ela é puramente especulativa, irrespondível e esgota-se em si própria. O que de facto tem interesse é a capacidade do agente de determinar o seu próprio carácter de modo a, precisamente, ser um agente tal que *não poderia fazer de outro modo* perante determinadas circunstâncias. É essa capacidade, que os agentes humanos têm ou podem ter, de se fazerem ser de um modo tal que perante circunstâncias determinadas não possam agir de outro modo que é, segundo Dennett, um componente essencial da pessoalidade e da responsabilidade. Assim, o sentido relevante do "poder ou não poder fazer de outro modo" não aponta para a liberdade contra-causal mas sim para a possibilidade de ajustamento dos hábitos de pensamento e de acção. No outro sentido de "poder fazer de outro modo" é razoável afirmar acerca de um agente artificial, também ele um deliberador determinista, que ele "poderia ter feito de outro modo". Ele pode inclusivamente ser redesenhado para fazer de outro modo. Mesmo o novo *design* não será infalível e utilizará processos heurísticos: o deliberador determinista 1 poderia já estar a usar o melhor processo tendo encontrado imprevisibilidade no ambiente, ou tendo havido falha física no sistema. Assim, ele não poderia ter feito de outro modo.

O retrato do agente apresentado resulta numa visão naturalizada e realista de racionalidade e responsabilidade. Na metáfora de Dennett em ER, a arbitrariedade e o risco (logo, a possibilidade de erros) que caracteriza qualquer sistema de controlo com sensibilidade e acesso limitados à informação é uma espécie de "pecado original naturalizado". No mundo dos deliberadores deterministas, eventos podem ser causalmente determinados não sendo inevitáveis, dadas a possibilidade epistémica, os oportunidades, e o auto-controlo. Está assim constituído espaço de manobra suficiente para a acção livre: *deliberadores deterministas capazes de auto-controlo meta-reflexivo e de avaliação forte são agentes capazes de acções livres.*

4.5 *Manual de primeiros socorros morais e ética da virtude.*

O manual de primeiros socorros morais é uma necessidade para agentes que agem em tempo real. A ideia não é uma cedência na ambição do pensamento moral face aos detalhes imperfeitos da prática e sim uma tentativa de incorporação do conhecimento empírico acerca das imperfeições da racionalidade. As ideias de prudência naturalizada e manual de primeiros socorros morais nascem da constatação de que o processo de *satisficing* é fundamental em grande parte dos processos reais de escolha e decisão. Dennett defende que o processo de *satisficing* é a estrutura básica de toda a tomada de decisões (moral, económica, evolutiva). Agentes racionais resultantes de evolução por selecção natural não são, nem poderiam ser, agentes perfeitamente racionais na medida em que o seu *design* resulta de um processo de *satisficing*. Se a continuidade evolucionista que liga organismos e processos mentais é um sustentáculo do gradualismo da teoria da acção, a atenção às situações de deliberação e de decisão conduz Dennett do darwinismo às ética da virtude. Nestas, ao contrário do que acontece nas éticas da maximização racional, considerações acerca do carácter dos agentes são centrais, de certo modo substituindo o apelo ao cálculo racional. É o carácter do agente que determina o modo como este enfrenta situações particulares de deliberação e é esta referência à importância do carácter na deliberação e decisão morais que se opõe à desconexão das acções e decisões relativamente à vida do agente, característica das éticas da maximização racional. Ora, Dennett sugere que por imperativos de *design* e pressão temporal somos constituídos-para-agir de uma forma que se aproxima da situação caracterizada pelas éticas da virtude, i.e. somos regidos em situações de ponderação prática por algo como carácter, não por maximização racional. A noção de carácter, que é obviamente uma noção de âmbito pessoal nas éticas da virtude, é aqui transposta para o âmbito sub-pessoal. É possível no entanto estabelecer uma ligação entre caracterizações sub-pessoais e teorização moral: se a visão naturalizada da racionalidade de agentes morais num mundo determinista acentua a arbitrariedade, o risco e os erros de decisão em qualquer sistema de controlo com sensibilidade e acesso limitados a informação de um ponto de vista quase apenas descritivo, uma tal caracterização descritiva tem um peso determinado na "escolha" racional de uma filosofia moral. Dennett pensa que a consequência

óbvia é o afastamento das éticas da maximização racional por falta de pertinência psicológica.

De modo a pôr em relevo as características do processo de *satisficing*, Dennett propõe o seguinte exemplo de deliberação seguida de decisão[719]. Um departamento universitário pretende conceder uma bolsa ao melhor estudante que preencha determinadas condições e publicita essas condições. Não se contava com 250 000 candidaturas, concretizadas em dossiers exaustivos, incluindo escritos dos candidatos, classificações, recomendações, etc. Perante o número das candidaturas e a quantidade de informação de cada uma, o que será razoável fazer de modo a escolher uma?

A primeira coisa a constatar é que *é impossível considerar todos os factores relevantes no tempo disponível.* Assim, será imprescindível o uso cego e impiedoso de regras grosseiras de selecção e de decisão (no caso, por exemplo, poderá ser utilizado um critério que não nem absolutamente fiável nem irrelevante, as médias das classificações). Além disso, o decisor abdicará do controlo total processo, apoiando-se num processo aleatório (por exemplo um sorteio que reduza as candidaturas a cem ou duzentas). Após o uso dessas regras de modo a cercear as alternativas em consideração, procura-se salvar alguma coisa do processo selvagem anterior, nomeadamente procedendo a uma análise exaustiva e cuidadosa das candidaturas que ficaram, das quais se vem a escolher uma. Este ultimo processo é uma tentativa, *resultante de auto-monitorização, de corrigir o resultado final.* Evidentemente, qualquer decisão tomada nestas circunstâncias é vulnerável a um olhar retrospectivo: não existe qualquer garantia de que a alternativa que acabou por ser seleccionada seja a melhor. É mesmo provável que os melhores candidatos tenham sido afastados. No entanto, o decisor pode ainda assim considerar que a sua decisão foi a melhor possível, dadas as circunstâncias.

Dennett sugere que este processo retrata de forma bastante fiel as características dos processos reais de deliberação e escolha. Ora, estas características não são usualmente admitidas pelas principais correntes da filosofia moral, i.e. quer por utilitaristas quer por kantianos. Utilitarismo e deontologia concordam num ponto importante: ambos exigem, idealmente, que sejam consideradas todas as coisas anteriormente à decisão. Ora, de acordo com Dennett, «ne-

[719] Cf. DENNETT 1988c.

nhum sistema de ética alguma vez foi tornado computacionalmente tratável, mesmo indirectamente, para problemas morais do mundo real. Por isso, mesmo se não tem havido falta de argumentos utilitaristas (e kantianos, e contratualistas, etc) a favor de particulares políticas, instituições, práticas e actos estes estiveram sempre pesadamente cercados por cláusulas *ceteris paribus* e por suposições de plausibilidade acerca das suas suposições idealizantes. Estas cercas estão feitas para ultrapassar a explosão combinatória de cálculo que ameaça se de facto se procurar –como a teoria afirma que se deve fazer – considerar todas as coisas»[720]. Considerar todas as coisas é impossível para agentes reais, e Dennett pensa que é importante que a filosofia moral leve a sério uma tal impossibilidade, já que ela caracteriza a própria natureza dos agentes morais reais e não apenas limitações contingentes. Agentes morais são por natureza objectivamente limitados: juízos "uma vez considerados todos os factores" (*all things considered)* são *realmente* impossíveis. Por esta razão, Dennett sugere que é preferível conceber a deliberação moral como um exercício de disposições ("virtudes"), não como maximização racional. Entre outras coisas entende-se por maximização racional a seguinte situação na deliberação:

> Dado o desejo D e as crenças C1, C2, Cn...de um agente a coisa mais racional a fazer, considerados todos os factores, aparece ao agente como R. Logo, o agente decide por R *como qualquer agente racional faria.*

Aparentemente, se se nega a maximização racional torna-se muito difícil continuar a defender a existência de decisões e acções racionais. Dennett defende no entanto que as decisões das pessoas são descritas de forma mais realista como resultado de processos heurísticos. As pessoas são guiadas por algo como um "cuidado não maximizante" e fazem a todo o momento escolhas não maximizantes. No entanto, essa estratégia não é irracional, e isto acontece por várias razões, nomeadamente devido à situação de insuficiência da informação e à pluralidade dos "bens" em consideração.

Dois tipos de considerações nascem da consideração realista da deliberação e decisão racionais. Um primeiro tipo diz respeito à natureza da racionalidade. A racionalidade de agentes racionais

[720] DENNETT 1995: 500.

reais é constitutivamente imperfeita, é apenas a melhor adaptação possível do comportamento ao ambiente, consideradas as crenças e desejos, sem que no entanto exista um ponto de vista exterior em função do qual se possa falar de Racionalidade *tout court,* de Racionalidade maiusculada. O erro comum às éticas da maximização racional é aliás supor que existe uma única (melhor ou mais alta) perspectiva a partir da qual é possível avaliar a racionalidade. Um segundo tipo de considerações diz respeito à questão concreta de encontrar estratégias de melhoramento das decisões, e é neste contexto que Dennett propõe o *manual de primeiros socorros morais.* O manual, que diria o que fazer imediatamente, e teria como finalidade criar estados disposicionais dos agentes morais, que têm que decidir sem tempo para calcular todas as implicações dos actos, é necessário na medida em que as decisões éticas das pessoas, feitas sob pressão temporal e debaixo de um véu de ignorância, envolvem procedimentos heurísticos e não cálculo exaustivo de alternativas. O que poderia conter o manual? Considerando o exemplo acima, apesar dos constrangimentos objectivos às decisões, o deliberador pode discernir os meta-problemas do seu próprio processo de decisão e desenvolver hábitos de pensamento que lhe permitam lidar com esses problemas. Relativamente à proliferação das considerações geradas, o deliberador poderá por exemplo desenvolver interruptores da ponderação (*conversation stoppers*), que terminem arbitrariamente a geração das considerações. Estes interruptores da ponderação são maneiras de lidar com a intratabilidade computacional e impedem de facto a consideração de todas as coisas. Um deliberador racional deve ser tal que envolva um impedimento estrutural de considerar todas as coisas. A natureza dos terminadores de considerações é obviamente uma questão problemática: eles podem ser de muitos tipos. A um nível ético, os terminadores das considerações podem ser por exemplo "direitos" ou "regras", noções que têm todas as vantagens do dogmatismo sobre a hiper-racionalidade, i.e. conduzem mais facilmente um agente à acção. Um outro conteúdo possível e bastante problemático do manual seria a ideia de que os manuais podem ser vários e alternativos.

Embora Dennett apresente as suas sugestões sob a forma de investigação do *design* de um deliberador, este pode obviamente ser um agente moral humano e de facto – embora Dennettt rejeite explicitamente a aproximação – muitas das estratégias para um certo controlo da deliberação apontadas no manual de primeiros socor-

ros morais aproximam-se de características propostas nas éticas da virtude. Como se sabe, estas opõem-se às éticas deontológicas e às éticas utilitaristas por considerarem que aquilo que deve ser procurado numa vida moral complexa é um *equilíbrio* e não uma maximização de valores e de escolhas e por considerarem que esse equilíbrio será em grande parte função das disposições do *carácter* do agente (e não da qualidade racional das regras ou das consequências da acção). O que Dennett sugere é que os problemas conceptuais com que a ideia de maximização racional se depara no domínio ético e que conduzem às propostas das éticas da virtude não são mais do que *reflexos de características estruturais do deliberador real.* Sendo esse o caso, é preferível que o pensamento moral parta de tais características estruturais em vez de pretender que elas são imperfeições desprezáveis da realidade. Nesse sentido, aquilo que começa por ser uma mera descrição das características sub-pessoais pode conduzir a propostas substantivas em filosofia moral ou pelo menos guiar a escolha e a rejeição de posições em filosofia moral. Por exemplo, pode-se imaginar uma situação em que além da sua utilidade estrutural os interruptores de considerações têm uma motivação moral. Por outro lado, a opção pelo abandono do ideal da maximização racional pode traduzir-se numa concepção não racionalista de moralidade (de facto, é razoável pensar que um agente moral totalmente racional não gozaria sequer a vida própria nem gozaria da existência de outros sujeitos[721]). Uma última ideia moral substantiva possível seria a seguinte: se existe uma pluralidade de objectos de cuidado, um cuidado não maximizante, acompanhado da admissão de que comparações de valores plurais são impossíveis, é uma estratégia racionalmente sã.

4. 6 *Problemas de fundo.*

O problema geral do presente capítulo consistiu em identificar e examinar as consequências da TSI e do MEM nas noções de pessoa e acção de modo a avaliar a consistência global da teoria da razão prática que nasce na confluência das teorias dennettianas do conteúdo e da consciência. Do ponto de vista sub-pessoal e de acordo com caracterizações do âmbito do MEM caracterizou-se os

[721] PENCE 1991. Estas duas críticas aplicam-se, note-se, quer à deontologia quer ao utilitarismo precisamente tanto quanto estas são éticas da maximização racional.

fenómenos do controlo, do Eu, da deliberação e da escolha. Do ponto de vista pessoal, ou nele culminando, e com apoio na TSI, caracterizou-se as condições da pessoalidade bem como aspectos da constituição da identidade pessoal. Em função desta dupla aproximação procurou-se explicar por que razão e até que ponto determinandos eventos podem constituir-se como acções pessoais.

A análise dennetiana tem várias imperfeições e pontos cegos. Notou-se desde logo que faltava algo à análise das condições da pessoalidade que permitisse conceber a sua aplicação a uma pessoa individual. Nesse sentido, e seguindo Rovane, propôs-se que Dennett deveria prolongar a análise da dimensão normativa da identidade pessoal examinando o exercício da racionalidade sobre si, o qual se traduz num compromisso com a vida racional própria (envolvendo resolução de conflitos entre crenças, aceitação das consequências das crenças próprias, ordenação de preferências, etc), que precisamente unifica a vida racional de um agente como *própria*. São essas características estruturais – ou uma tendência para elas, mesmo que nenhuma seja perfeitamente alcançada, já que características como a coerência total das crenças ou a ordenação totalmente coerente das preferências são de facto impossíveis – que constituem uma *vida mental pessoal*. O que está em causa é algo como um ponto de vista racional único de um agente sobre si próprio e esta ideia ultrapassa os tradicionais critérios físico e fenomenológico da identidade pessoal. Repare-se que este complemento à noção de identidade pessoal significa que a sexta condição de pessoalidade, a auto-consciência, envolve um uso de racionalidade sobre si próprio que tem a ver não apenas com auto-avaliação mas também com o apoderar-se dos processos de uma vida mental. Em ambos os casos se trata de voltar a EI sobre si, de se visar sob a perspectiva da EI de modo a controlar a acção própria e aquilo que se é. O primeiro tipo de auto-avaliação pode conduzir a uma mudança naquilo que se é, uma mudança a partir de dentro, nos casos em que aquilo que se é não se adequa àquilo que se quer ser. *Esta é uma das razões pelas quais aquilo que se entende por auto-consciência não é simplesmente auto-apercebimento*. O segundo tipo de auto-avaliação traduz-se no objectivo de uma certa completude ou unidade da racionalidade global do agente e é um compromisso com um ideal de racionalidade. Neste contexto, Rovane propõe mesmo que se deve considerar – apesar de subscrever o gradualismo e o naturalismo da teoria dennetiana que subscreve – que existe uma divisão real entre as três

primeiras e as três últimas condições de pessoalidade: entidades que satisfaçam as condições 4 e 5 já são pessoas no sentido envolvido no compromisso de racionalidade, i.e. as condições 4 e 5 não podem ser satisfeitas sem que a condição 6 esteja já a ser satisfeita. A capacidade racional que envolve pensamento de ordem mais alta não pode ser separada da capacidade de avaliação crítica e da racionalidade reflexiva em geral. Assim, de acordo com a sua proposta, a que Rovane chama "um critério ético de pessoalidade", alguma entidade é uma pessoa se e só se pode ser tratada como pessoa e pode tratar outros como pessoas[722]. Obtém-se deste modo uma concepção qualificada daquilo que se entende por reciprocidade que afasta a possibilidade de existirem intenções comunicacionais griceanas sem auto-consciência. A reciprocidade só existe de facto se existirem conexões racionais "a partir de dentro" do outro SI e nomeadamente susceptibilidade de persuasão racional, de ser movido por razões.

Para concluir, note-se que, de acordo com a teoria apresentada, pessoas e acções não são absolutamente determinadas, como nada no mundo descrito de forma mentalista o é: se a descrição intencional à qual as razões pertencem é sempre aproximativa, poderá dar-se o caso de não haver resolução completa de movimentos em acções, de causação em razões. Uma objecção possível a esta caracterização radicalmente anti-essencialista de pessoa é no entanto a seguinte: a auto-consciência é aqui caracterizada como desprovida de conteúdo sentido ou sentimento de si. Este é um dos maiores defeitos da abordagem apresentada e um defeito que possivelmente não pode ser ultrapassado a partir do próprio Dennett.

[722] Com o seu critério ético de pessoalidade Rovane pretende corrigir o preconceito universalista, kantiano-rawlsiano, implícito nas formulações de *Conditions of Personhood* e que pode ter a consequência indesejável de negar pessoalidade aos humanos que demonstrassem más performances relativamente a tais ideais (por exemplo por subscreverem teorias morais afastadas da linhagem kantiano-rawlsiana). Daí a grande insistência de Rovane em formular o critério ético de pessoalidade de modo a permitir a divergência de opiniões morais.

Capítulo 5

Questões aplicadas da TSI. Tipos de Mentes: mentes animais, artificiais e humanas.

> «In fact there is an eerie resemblance between many of the discoveries of cognitive ethologists working with lower animals and the sort of prowess mixed with stupidity one encounters in the typical products of AI»,
> Daniel Dennett[723]

5.1 *A horizontalidade da perspectiva da TSI sobre o mental. O natural e o artificial. Ciência cognitiva: engenharia invertida e/ou síntese. Cérebros, programas e robôs: da base para o topo ou do topo para a base. A IA e as experiências de pensamento reais. Tipologia (a partir de dentro e de baixo) das criaturas cognitivas.*

O funcionalismo permite a Dennett defender que não existe diferença essencial entre a inteligência "natural", humana ou animal, e a inteligência artificial. Embora a indistinção entre natural e artificial esteja em jogo na constelação de disciplinas agrupadas sob o título de ciência cognitiva – aliás, no próprio facto de a IA ser para muitos autores a disciplina matricial da ciência cognitiva[724] – a relação entre natural e artificial foi teorizada de forma marcante em certos pontos focais na literatura. Um desses pontos é o livro *The Sciences of the Artificial*, de H. Simon[725]. Neste livro, H.Simon, um dos fundadores da

[723] DENNETT 1987: 256.

[724] Cf. por exemplo HAUGELAND 1997c:1.

[725] SIMON 1969. Um outro ponto de referência, já referido anteriormente, é a experiência mental do Quarto Chinês, encarada como um teste intuitivo para discernir a diferença entre a mentalidade genuína (que seria por hipótese apenas a mentalidade natural, biologicamente realizada) e a mentalidade simulada e atribuída (que seria a artificial). Embora esta interpretação seja a fonte da adesão de muita gente à intuição convocada pela experiência, o próprio Searle não interpreta deste modo o

IA, defende que as ciências do artificial não representam de modo algum um afastamento relativamente às ciências da natureza. O artificial e o natural não são, segundo Simon, dois reinos mas dois pontos de vista, que não se opõem. Tudo o que é "*artificial*" (de acordo com H. Simon, tudo aquilo que é um artefacto adaptativo, avaliável segundo normas de funcionamento racional) é também *natural* (i.e. susceptível de uma explicação física). Aquilo que distingue o ponto de vista do artificial é outra coisa que não a existência de um reino distinto de entidades: é o facto de ele visar sistemas no seu estatuto de *interfaces* entre interior e exterior, criando assim a nova questão da racionalidade ou adaptação. Noutras palavras, segundo H. Simon, uma ciência do artificial é uma ciência do artificial porque lida com *teleologia*, com finalidades do comportamento global de sistemas, independentemente da realização material destes. Quer o interior do sistema quer o seu exterior continuam a pertencer ao departamento das ciências da natureza, é o interface que é especificamente artificial. É óbvio que também um sistema natural (entenda-se biológico) pode ser analisado de acordo com estes parâmetros e de facto, em *The Sciences of the Artificial,* H. Simon inclui a psicologia (definida como ciência dos *behaving systems*) nas ciências do artificial. Simon pretende sublinhar que a psicologia de um sistema é uma ciência do artificial por contraste, por exemplo, com a neurofisiologia do mesmo sistema, que visa exclusivamente o interior físico e não o interface interior-exterior ou as finalidades do comportamento global[726].

É evidente que o "artificial" de Simon poderia ser chamado "teleológico", e que a classificação da psicologia como ciência do artificial só é possível a partir do momento em que Simon pressupõe que considerações teleológicas são estranhas às ciências naturais. Mas o que aqui interessa sublinhar é que o que torna o funcional "artificial" não é, para Simon, o facto de ser uma característica de *artefactos* mas o facto de ser uma característica de *dispositivos adaptativos*. Ora, um dispositivo adaptativo pode ou não ser artificial no sentido vulgar de ser construído por humanos, de não "brotar" da natureza. Aliás e a propósito da irrelevância teórica da realização material dos sistemas visados nas "ciências do artificial", recorde-se que foi enquanto economista que H. Simon abordou inicialmente a questão da racionalidade, aplicando-a às organizações administrativas. De facto, para além de sistemas psicológicos e sociais, também as organizações são, para Simon, sistemas artificiais biologicamente implementados[727]. De resto, tanto organizações administrativas como cognição humana exemplificam um traço da racionalidade analisado por Simon: o *satisficing*, que se opõe à optimização e que é característico da racionalidade limitada (*bounded rationality*). Isto significa que uma outra maneira de abordar o ponto de vista do artificial é defini-lo pelos limites da racionalidade prática, que nunca será perfeita. Esta é, como se sabe, uma ideia muito importante para a TSI.

O que se pretende neste momento sublinhar é o facto de a teoria dennettiana do mental ser herdeira de uma certa indistinção entre natural e artificial que se foi estabelecendo no seio da ciência cognitiva e da qual o livro de Simon é exemplar. A indistinção entre o natural e o artificial conduz Dennett a afirmar, por exemplo, que performances cognitivas de mentes biologicamente baseadas e actuais se situam, conjuntamente com performances de outros tipos de mentes, num mesmo espaço do *design,* devendo todas ser horizontalmente analisadas numa mesma teoria do mental. É essa a justificação para o título ele próprio "horizontal" do presente capítulo, no qual se tratará de mentes humanas, animais e artificiais. De acordo com Dennett, não

Quarto Chinês. Mas muito antes do artigo de Searle teóricos da mente como A. Turing e H. Simon debatiam-se já com a questão do natural e do artificial. Turing apresentara inclusivamente, no artigo *Computing Machinery and Intelligence* (TURING 1950), no qual é proposto o Teste de Turing, uma cerrada argumentação contra possíveis objecções à ideia de uma Inteligência Artificial. Turing considerou nomeadamente a *objecção teológica* (de acordo com a qual o pensamento estaria ligado à alma, tendo Deus dado uma alma apenas aos humanos), a *objecção cabeças-na-areia* (segundo a qual seria terrível que artefactos pensassem, sendo por isso preferível acreditar que tal não acontecerá), a *objecção ligada ao Teorema de Gödel* (segundo a qual os humanos têm capacidades mentais que transcendem o que é computável), as *objecções ligadas ao sentimento e à consciência* (de acordo com as quais uma máquina não poderia ter estados como depressão, amor, emoção, etc), a *objecção ligada às incapacidades* (de acordo com a qual uma máquina nunca seria capaz de humor, aprendizagem, moralidade, paixão), a *objecção ligada à impossibilidade de originação do novo por uma máquina programada*, a *objecção ligada à continuidade do sistema nervoso*, etc. Se um dos primeiros contra-argumentos à artificialidade do artificial se bascia no facto de o artificial ser também natural, H. Simon (SIMON 1969) defenderá também que de alguma maneira certas organizações naturais são "artificiais" no sentido analisado em *The Sciences of the Artificial.* Para os propósitos do presente trabalho interessa notar que a simplicidade da diferença aristotélica entre o que é "natural" e o que é (se assim se pode dizer...) *contra-natura* é de qualquer modo posta em causa quando a teoria da cognição se alarga até à Inteligência Artificial (cf. DESCOMBES 1995:194, *Le problème de l'artificialité*). Repita-se que apesar de se considerar neste trabalho o Quarto Chinês como um foco da crítica à artificialidade das simulações da cognição, não é no entanto justo considerar que existe uma sobreposição entre a dicotomia searleana intencionalidade intrínseca / intencionalidade atribuída e a dicotomia natural /artificial. J. Searle não é um crítico simplista da IA.

[726] Vincent Descombes (cf. DESCOMBES 1995: 214) interpreta esta "des-cerebralização" da mente, do pensamento e da racionalidade feita (conscientemente) por H. Simon na sua teorização das ciências do artificial como a afirmação de uma *preponderância da natureza formal face à natureza material*, característica do cognitivismo.

[727] Para uma análise do pensamento de H. Simon, cf. MACHUCO ROSA, 2002, *Dos Sistemas da Razão às Organizações Espontâneas*, Sistemas Centrados, 3. Génese das Estruturas Hierárquicas.

são aliás sequer apenas performances cognitivas que se situam no mesmo espaço do *design:* também os sistemas às quais elas se devem (por exemplo organismos) e os artefactos desses sistemas (por exemplo instrumentos humanos enquanto exteriorizações e *off-loads* dos sistemas cognitivos) se situam num mesmo espaço do *design*. Aliás, uma segunda consideração de horizontalidade se configura, que considera a par não apenas mentes humanas, animais e artificiais mas também sistemas cognitivos, operações de cognição e veículos ou instrumentos da cognição internos ou externos. Assim se aceita, por exemplo, uma concepção não derivativa, não autoral, da relação entre o mental como funcionamento "interior" de sistemas cognitivos e os dispositivos exteriores à delimitação física dos indivíduos.

A ideia segundo a qual qualquer ciência da cognição é um campo de testes para a pertinência dos princípios da TSI, seja qual fôr a "natureza" das entidades envolvidas, vem, assim, de longe na obra de Dennett. Apesar desta consideração horizontal, existem no entanto, de acordo com Dennett, dois caminhos metodológicos distintos em ciências da cognição: engenharia invertida (*reverse engineering*[728]) e síntese. Embora o espaço do *design* seja único para todas as arquitecturas cognitivas, há diferentes formas de o percorrer. Existe, nomeadamente, uma diferença importante entre o *design* para uma mente biológica e o *design* para uma mente artificial: as soluções de *design* biológico, na medida em que nelas se impõe o aproveitamento de material pré-existente, são por natureza imperfeitas e muito distantes do melhor *design* possível. Como diria H. Simon, a natureza faz o suficientemente bom (*satisfices)*, não optimiza. Pelo contrário sistemas cognitivos sintetizados são de alguma maneira mais puros, pelo facto de serem desenhados do topo para a base, permitindo de alguma forma saltar etapas na investigação da cognição. Dennett sublinha por esta razão o contraste entre investigações da cognição "abstémias e lentas", como as investigações em neurociências, e investigações mais "directas", nomeadamente em IA, e considera as segundas exemplares. As primeiras são abstémias na medida em que a sua abordagem da base para o topo (*bottom-up*), as impede de chegar mais longe mais depressa (i.e. as preocupações propriamente biológicas e bioquímicas dos neurocientistas impedem-nos em geral de chegar rapidamente às macro-arquitecturas cognitivas ou de as visar directamente ou exclusivamente). O objecto de estudo prende as neurociências à "base" embora obviamente o trabalho se faça tendo inevitavelmente como referência mentes completas, i.e. macro-arquitecturas, bem como a perspectiva do topo para a base. De acordo com Dennett, perante um objecto de estudo como a cognição, é mais simples deduzir competências comportamentais de arquitecturas que se desenhou do que elaborar hipóteses, da base para o topo, acerca do interior de caixas negras naturais cujo comportamento se observa (esta é a "lei" da *uphill analysis* e *downhill synthesis* que Dennett evoca a partir de V. Braitenberg e dos seus "ensaios de psicologia sintética"[729]) .

Considerações como estas conduzem Dennett à apologia da IA como filosofia e como psicologia que acompanha toda a sua obra[730]. Segundo Dennett as questões da filosofia e da IA são identicamente abstractas e gerais e os cientistas que trabalham em IA assemelham-se a filósofos na medida em que procuram responder à questão epistemológica geral acerca da possibilidade da cognição. As propostas da

[728] A expressão "engenharia invertida" tem a sua origem na prática industrial de "inspecção-por-desmontagem" dos produtos de uma empresa por uma empresa rival, de modo a obter uma cópia melhorada de um determinado produto. Qualquer ciência que analisa soluções de *design* já concretizadas percorre o caminho inverso ao do "criador" das entidades e pratica assim engenharia invertida. Dada a indistinção funcionalista mencionada no início do capítulo, a própria biologia (ou melhor as suas sub-disciplinas em que se procede do topo para a base) é para Dennett um exemplo de "engenharia invertida (cf. DENNETT 1995: 187, *Biology is Engineering*, que se inicia precisamente com uma alusão às "ciências do artificial"). Em geral, Dennett atribui ao darwinismo a responsabilidade por este emparelhamento da biologia com a engenharia. Outro aspecto da ideia segundo a qual a biologia é engenharia é a ideia do mental como "híbrido deselegante". Dennett pretende combater um defeito que considera característico de muitos filósofos e cientistas cognitivos: a tendência para considerar que as únicas alternativas sérias que se colocam ao estudo científico da mente (excluído o dualismo), são a física e a matemática, i.e. as abordagens puras. Esta opção oculta a perspectiva teoricamente menos pura segundo a qual o estudo da biologia como engenharia (e por isso parente da IA) seria por excelência o âmbito das ciências da cognição. Esta perspectiva impura é para Dennett uma fonte de importantíssimos *insights* quanto à natureza impura da mente e da racionalidade (cf. DENNETT 1995: 387, onde é relatada uma discussão acerca da natureza da IA entre N. Chomsky e J. Fodor por um lado e T. Winograd e R. Schank por outro). Chomsky declara que só existem duas possibilidades interessantes quanto à natureza da psicologia: ou esta vem a revelar ser como a física (i.e. uma ciência com leis profundas e inexoráveis) ou vem a revelar não ter quaisquer leis, caso em que a melhor maneira de estudar e expor psicologia será a maneira do romancista. É M. Minsky que comenta «I think only a humanities professor at MIT could be so oblivious to the third interesting possibility: psychology could turn out to be like engineering». O episódio, passado num encontro da Society for Philosophy and Psychology em 1978, é relatado em DENNETT 1995: 387. O caso de Chomsky, dada a enorme importância da sua obra na revolução cognitiva, é muito revelador: existem intuições profundas e incompatíveis em jogo no estudo da cognição. Uma dessas "intuições" é uma esperança: a esperança de que o mental seja um tipo natural e não o efeito contingente de uma colecção de *gadgets*, e que portanto a perspectiva de engenharia sobre o mental não seja a boa perspectiva, a perspectiva central a adoptar na investigação do mental.

[729] Cf. BRAITENBERG 1984. A lei alude evidentemente ao facto de a síntese facilitar a investigação da cognição. A análise *uphill*, i.e. mais custosa, caracteriza as disciplinas da cognição que procedem, em geral, da base para o topo, tais como as neurociências.

[730] Cf. por exemplo DENNETT 1978h, *IA as Philosophy and as Psychology* e DENNETT 1998r, *When Philosophers Encounter AI*.

IA situam-se alguns degraus mais abaixo, em termos de abstracção, relativamente às respostas de filósofos do conhecimento, mas têm a enorme vantagem de poderem ser ser postas à prova da experiência e até mesmo refutadas. As investigações em IA são, para Dennett e utilizando uma expressão cara aos filósofos, "experiências mentais". Elas são aliás experiências *mentais* no duplo sentido de envolverem a imaginação do que poderia acontecer em determinadas circunstâncias e de serem explorações de arquitecturas cognitivas. São no entanto experiências mentais *reais,* por oposição às experiências mentais imaginárias dos filósofos na medida em que consistem em criação efectiva de sistemas, i.e. na engenharia de programas e de máquinas. Ao contrário do que se passa com as experiências mentais dos filósofos, nas quais o mundo vai sempre para onde o autor da experiência mental deseja, sendo a variável dependente a intuição da audiência, na medida em que as experiências mentais da IA são controladas por meio de "próteses", i.e. dos próprios dispositivos construídos, que embora ideados existem e são postos à prova no mundo, alguma coisa pode correr mal, e de forma ilustrativa.

Se é verdade que programas e robôs são para Dennett experiências puras sobre o mental possível, é importante notar que essas experiências não têm, nem têm que ter, como referência o mental natural (i.e. o mental biológico, humano ou animal). Pelo contrário, o estudo do mental num campo como a etologia cognitiva é um estudo do mental biologicamente baseado, conduzido do topo para a base, i.e. do comportamento inteligente para a implementação física. A etologia cognitiva é outra das disciplinas da cognição à qual Dennett reporta a TSI. Dennett acentuava já aliás, há mais de dez anos atrás, a pertinência teórica e metodológica de uma aproximação entre etologia cognitiva e IA. Esta aproximação está hoje estabelecida[731]. É certo que a simulação de competências cognitivas *humanas* foi cronologicamente primeira na história da IA e pareceu atraente sobretudo devido à disponibilidade de dados introspectivos. No entanto, a simulação de mentes mais simples pode vir a revelar-se um caminho mais frutífero do que a simulação de mentes humanas, já que essa simulação pode visar "mentes integrais"[732],

[731] Cf., para uma concretização de tal aproximação a revista, *Adaptive Behavior.* O palavra de ordem do movimento gerado pela aproximação é "From Animals to Animats" (cf. DENNETT 1998: 307).

[732] É o sentido da proposta de Dennett ilustrado pelo título de DENNETT 1978s, *Why Not The Whole Iguana?*

embora simples, e não sub-componentes (artificialmente) isolados de mente complexas como as mentes humanas. Dennett vê quer as criaturas sintéticas de V. Braitenberg (animais imaginários) quer os insectos de R. Brooks[733] (robôs) como tentativas nesse sentido, i.e. como tentativas de conceber e simular mentes simples mas completas. Ora, precisamente por não ser possível aqui qualquer acesso introspectivo, os trabalhos dos estudiosos do comportamento animal podem ser úteis à IA.

A verdade no entanto é que, apesar do interesse que Dennett sempre dedicou a mentes animais e artificiais, a referência central da TSI são as mentes humanas. Mas a própria horizontalidade da perspectiva abre a porta à análise de semelhanças e diferenças entre tipos de mentes. Isto significa, por exemplo, que a TSI é capaz de apresentar propostas quanto a limiares de possibilidade de mentes humanas. A ideia central é que estes se relacionam com as arquitecturas para a comunicação e para a linguagem, as quais apoiam o teor de acto de fala do pensamento humano superior. Esta característica faz com que as mentes humanas sejam de tal modo mais poderosas e sofisticadas do que, por exemplo, as mentes animais conhecidas, que frequentemente o trabalho do filósofo que se aproxima da etologia cognitiva consiste, como se verá no presente capítulo, em "deflacionar" interpretações do comportamento animal, nomeadamente atribuições de comunicação e consciência a animais. O caso característico de atribuição extra-generosa de consciência às mentes animais na etologia cognitiva é o do etologista D. Griffin[734]. Ora, ao contrário do que defende um autor como Griffin, a TSI sustenta uma posição deflacionista acerca de mentalidade animal: a distância entre as mentes humanas e mentes de outras espécies, mesmo das espécies mais inteligentes, é enorme, suficientemente grande para fazer toda a diferença em termos morais, por exemplo. Apesar disso, os princípios da TSI não permitem afirmar que se trata de uma diferença de natureza, mas apenas, de uma diferença de grau.

[733] Para uma explicação dos princípios teóricos da robótica de R. Brooks, cf. MACHUCO ROSA, 2002, 4. A Nova Robótica, 4.1. A arquitectura de subsunção de R. Brooks. Cf. também BROOKS 1997. Está aqui em causa o abandono da noção de representação, demasiado abstracta, e a sua substituição por comportamentos ou actividades que pressupõem acções reais no mundo e um nível fenomenológico de realidade.

[734] A intenção do trabalho de Griffin (cf. GRIFFIN 1976, GRIFFIN 1992) é em grande parte inspirada pelas ideias de Nagel (cf. NAGEL 1974). Cf. também ALLEN & BEKOFF 1997.

Precisamente porque a diferença entre tipos de mentes não é de natureza mas de grau, Dennett interessa-se pela gradação das mentes e pela distância que as separa no espaço do *design*. Uma das estruturas organizadoras do livro *Kinds of Minds* (KM) é uma tipologia de criaturas cognitivas feita a partir de dentro e de baixo. Este "dentro" não é o interior do "sentir-se ser" mas o interior físico dos sistemas cognitivos onde acontecem processos de geração e teste que vêm a possibilitar o "*design*-para-uma-mente". "Baixo" significa assim menos complexo do ponto de vista evolucionista e de acordo com a tipologia de Dennett as *criaturas darwinianas* iniciais, i.e. os primeiros exemplares de *design* biologicamente baseado, geradas por combinação e mutação de genes, eram testadas em campo. Nas *criaturas skinnerianas* posteriores existe já plasticidade fenotípica, i.e. o *design* dos organismos não está completo desde o nascimento, existindo espaço para um ajustamento por meio de testes em campo de uma variedade de acções que "confrontam" o ambiente até que algumas funcionem[735]. A distinção entre moldagem por selecção natural de indivíduos e moldagem do comportamento por condicionamento (a diferença entre criaturas darwinianas e skinnerianas) não é, note-se, teoricamente muito importante (a não ser para o próprio indivíduo darwiniano, dir-se-ia, uma vez que este por definição dá-a-vida-como-hipótese em vez de experimentar hipóteses de acção). Superiores às criaturas skinnerianas são as *criaturas popperianas*, que, essas sim, produzem no seu interior hipóteses que morrem em vez de si (i.e. as propostas comportamentais alternativas que as criaturas skinnerianas apresentavam ao ambiente são feitas interiormente nas criaturas popperianas). Os sucessores mais sofisticados das criaturas popperianas são criaturas em cujos cérebros está instalado *design* apoiado em porções do mundo exterior nas quais existe *design* (i.e. artefactos, inclusive a linguagem escrita). Estas são *criaturas gregoryanas*[736], capazes de pensar (fisicamente) dentro, virtualmente, com apoio em *design* "exterior". É evidentemente este o caso das mentes humanas.

[735] Dennett chama "aprendizagem ABC" a esta conjugação de associacionismo, behaviorismo e conexionismo. As criaturas capazes de aprendizagem ABC são capazes de redesenhar o seu comportamento como resultado de moldagem (*shaping*) pelo ambiente (cf. DENNETT, 1996: 81, *The Tower of Generate and Test*).

[736] A origem do termo é o nome do psicólogo inglês Richard Gregory, teórico do papel da informação corporizada em objectos na inteligência.

5.2 *Questões aplicadas da TSI: as mentes animais e o problema da interpretação, o problema do enquadramento (frame problem) na IA, o carácter incorporado das mentes.*

A primeira aplicação da TSI aqui considerada diz respeito às mentes animais. Dennett procura utilizar a TSI para fundamentar do trabalho dos etologistas cognitivos, defendendo que o problema teórico colocado pelo comportamento é idêntico no caso humano e no caso animal. Trata-se em ambos os casos de *um problema de interpretação, na continuidade do problema biológico da adaptação*. O caso específico em foco é a (suposta) comunicação entre criaturas não linguísticas. Dennett parte dos pressupostos seguintes. As mentes animais são mais simples do que as mentes humanas na medida em que, como se viu no Capítulo 3, não é possível considerar a unificação da vida cognitiva, pelo menos tanto quanto esta envolve uma auto-representação razoavelmente explícita, como uma característica uniformemente distribuída por todas as mentes biológicas. A unificação virtual é um acontecimento tardio da história evolutiva de uma única espécie, a humana, e, aliás, nem mesmo seres nos quais ela existe presentemente tiveram desde logo uma tal unificação organizacional já instalada. A unificação resulta antes de uma instalação (devida à linguagem natural) responsável pela ilusão do utilizador ou Teatro Cartesiano, que é constitutiva das mentes humanas e apenas destas. *Dennett defende assim que a ilusão do utilizador não existe noutras espécies e nem mesmo nos recém nascidos humanos, sendo todas estas mentes muito mais descentralizadas do que as mentes humanas adultas*. Como já se viu no Capítulo 3, é pelo facto de a consciência envolver, ou ser idêntica a, uma organização informacional como esta que Dennett não identifica consciência com sensiência. Ora, é a esta que usualmente se faz apelo para defender a paridade entre mentes humanas e animais. Uma tal posição opõe-se a convicções que estiveram na origem de desenvolvimentos recentes na etologia cognitiva. Nomeadamente, o etologista Donald Griffin, o responsável pela introdução da expressão "etologia cognitiva"[737], lançou um novo tipo de trabalho etológico ao interessar-se explicitamente

[737] Para uma história da disciplina, cf. ALLEN & BEKOFF 1997: 21, *A Brief Historical Account of Classical Ethology and Cognitive Ethology*, e especialmente a p. 32 e seguintes para a avaliação da importância da obra de D. Griffin. O trabalho de D. Griffin e de outros etologistas foi estimulado pelo célebre artigo de Nagel *What is It Like to Be a Bat?* (NAGEL 1974). Para todos esses etologistas tratava-se, no seu trabalho, de responder à questão de Nagel relativamente a diferentes espécies animais. Em

pela questão "como é, para criaturas não humanas, subjectivamente, existir", defendendo a existência num grande número de espécies animais de consciência, expectativas, memória e pensamento sobre objectos não existentes. A consciência seria, segundo Griffin, uma propriedade emergente que confere uma enorme vantagem permitindo aos animais seleccionarem as acções que mais provavelmente servirão para conseguirem aquilo que desejam e afastar aquilo que temem. Griffin utiliza a palavra "consciência" de forma ambígua – nomeadamente como sinónimo de apercebimento (*awareness*), mentalidade (*mentality*), cognição e consciência reflexiva – estendendo-a a muitas espécies animais. O seu trabalho foi repetidamente acusado de falta de fundamentação teórica, precisamente por utilizar de forma liberal conceitos polémicos, tais como consciência e racionalidade, sem os definir. O tratamento do comportamento animal por Griffin foi também acusado de ser antropomorfista e o seu trabalho de não passar de uma mera listagem de episódios. Assim, embora D. Griffin tenha lançado a etologia cognitiva, ele não a justificou teoricamente. É nesse ponto que a TSI entra em cena.

A segunda aplicação da TSI a ser aqui considerada diz respeito a mentes artificais, o que conduzirá a uma incursão pela filosofia da IA, nomeadamente através da abordagem do *frame problem* (problema do enquadramento), um problema relativo a representação do conhecimento, que tem sido objecto de discussão filosófica desde que foi identificado em 1969 por J. McCarthy e P. Hayes[738]. O problema consiste no seguinte: qualquer agente, para agir no mundo, deve dispor de representações internas dos traços relevantes desse mesmo mundo. Para isso é necessário um quadro de referência (*frame of reference*). Ora, acções no mundo, inclusive acções do próprio sistema, podem alterar esse quadro de referência. O problema consiste em saber o que é necessário para actualizar o quadro de referência, nomeadamente o que é necessário para que o agente saiba que coisas permanecem inalteradas num mundo em alteração e que coisas são de facto alteradas e alteráveis. Na formulação de J. McCarthy e P. Hayes, o problema do enquadramento é relativo a estratégias utiliza-

ALLEN & BEKOFF 1997 é feita uma história breve das tendências metodológicas na etologia, radicando essas tendências no debate filosófico moderno entre empiristas e racionalistas. A história inclui ainda a recepção na etologia das ideias de filósofos como Nagel, Wittgenstein e Quine. Para a influência de Griffin, cf. ainda RISTAU 1991. Para um exemplo de proposta teórica contemporânea de etologia cognitiva cf. CHENEY & SEYFARTH 1990:1, Cap. 1, *What is it Like to Be a Monkey*.

[738] Cf. McCARTHY & HAYES 1969, *Some Philosophical Problems from the Standpoint of Artificial Intelligence*.

das por agentes que fazem planeamento em tempo real para lidar com os problemas surgidos no enquadramento do planeamento. Em termos concretos, as formas de lidar com o problema do enquadramento que têm sido sugeridas vão desde a utilização de sistemas de lógica não monotónica para formular as inferências do agente, até à proposta de meios de representação analógicos que substituam vantajosamente as linguagens lógicas na representação da mudança, até novas ontologias e categorizações das propriedades do mundo[739].

Algumas das mais antigas críticas à IA, nomeadamente as críticas de Hubert Dreyfus no relatório *Alchemy and IA* que veio a dar origem à conhecida obra *What Computers Can't Do*[740], visavam já o problema da "falta de contacto" de sistemas simbólicos artificiais com a realidade, falta de contacto esta de que o problema do enquadramento pode ser considerado um sintoma. Dreyfus considerava então, inspirado pela fenomenologia europeia, que a formalização da cognição através de operações lógicas não era suficiente para dar conta da natureza desta, na medida em que aquilo a que chamamos "significação" está ligado por um lado ao mundo e por outro ao corpo próprio enquanto experienciado[741]. O problema do enquadramento seria assim apenas uma consequência particular, no âmbito do planeamento de acções, da má interpretação das relações de um sistema cognitivo com o seu ambiente implícita no uso de representações simbólicas e descontínuas. Como se verá, na interpretação que Dennett, faz do problema do enquadramento, este é um problema epistemológico novo revelado pela IA.

Considerar-se-á finalmente, no presente capítulo, algumas características da mentalidade que não foi possível tratar directamente nos capítulos 2, 3 e 4, fazendo notar que a ideia de mente incorporada (*embodied mind*) envolvida na TSI não é apenas a negação do dualismo imaterialista. O que está em causa é uma incorporação generalizada do mental, e nomeadamente a distribuição de várias "mentes" pelo corpo de um sistema cognitivo, pelo menos no caso dos sistemas cognitivos biológicos. Embore se mantenha a ideia de uma "divisão do trabalho constitutiva do mental", produto do funcionalismo psi-

[739] Fala-se de inferência não monotónica quando a adição de premissas altera aquilo que pode ser provado a partir das premissas já presentes, ao contrário do que acontece na lógica clássica, dita monotónica. Para um panorama das diferentes formas de enfrentar o problema, cf. LOUI 1996.

[740] DREYFUS 1972.

[741] Nomeadamente, de acordo com Dreyfus, à experiência fenomenológica e à *fringe consciousness* a ela ligada.

cológico, trata-se agora de admitir agora que talvez essa divisão do trabalho não seja neutra relativamente à implementação física. Se o mental se decompõe em mentes distribuídas por corpos, ele não poderá ser pensado sem se considerar a sua corporeidade ou incorporação. Este não é aliás o único sentido da mente incorporada. Um segundo sentido diz respeito ao facto de a corporeidade ser constitutiva do mental na medida em que existe um corpo no cérebro, i.e. uma representação do corpo próprio que acompanha o pensamento/cognição daquele indivíduo e um terceiro sentido diz respeito ao facto de o tipo específico de mente humana não ser concebível sem prolongamentos exteriores às fronteiras biológicas do indivíduo humano que representam *design*-para-pensar. É esse aspecto que está em causa por exemplo quando se fala das criaturas gregoryanas, cujo pensamento acontece no interior físico virtualmente, com apoio em *design* exterior. Aquilo a que se poderia chamar um *off load* ou exteriorização do mental para o ambiente é assim característico das mentes humanas[742] (senão, como saberia alguém por exemplo algo tão simples como "quanto é 234567×234678?"). Como Dennett notará em KM, não é possível comparar o pensamento possível a partir de um cérebro humano nu com o pensamento possível com apoio em todo o *design* disponível no ambiente.

5.2.1 *Mentes Animais*[743]*: etologia cognitiva.*

> «I claim that other species simply aren't beset by the illusion of the Cartesian Theater»,
> Daniel Dennett[744]

[742] Cf. DENNETT 1996: 134-135: «Our brains are modestly larger than the brains of our nearest relatives (...) but this is almost certainly not the source of our greater intelligence. The primary source, I want to suggest, is our habit of off-loading as much as possible of our cognitive tasks into the environment itself – extruding our minds (that is our mental projects and activities) into the surrounding world, where a host of peripheral devices we construct can store, process, and re-represent our meanings, streamlining, enhancing and protecting the processes of transformation that *are* our thinking. This widespread practice of off-loading releases us from the limitations of our animal brains».

[743] A questão das mentes animais tem uma presença constante na obra de Dennett. Além do artigo da revista *Behavioral and Brain Sciences* (DENNETT 1987m), que esteve no centro de um prolongado debate, toda a parte III de BC (*Ethology, Animal Mind*) é dedicada ao problema das mentes animais. Para uma abordagem anterior do problema das mentes animais, cf. HOFSTADTER & DENNETT 1981: 100, *The Soul of Martha the Beast* (Texto 7, de Terrel Miedaner). Recorde-se que em *Conditions of Personhood*, a questão era também abordada. Em C&C o problema das mentes animais é evocado a propósito do erro que o excesso de especificidade na atribuição de conteúdo a mentes animais constitui.

[744] DENNETT 1998x: 346.

Ficou estabelecido através das análises da consciência e da acção aquilo que Dennett pensa ser essencial a uma vida mental humana: a especificidade desta reside em grande medida na voluntariedade, que se encontra focalmente nos *actos de fala constitutivos do pensamento*. Esses actos de fala são (do ponto de vista teórico) juízos ou, na terminologia de Dennett, "opiniões" (por oposição ao conceito geral de crença tal como este funciona na TSI) e (do ponto de vista prático) decisões. Os outros animais são incapazes deste tipo de actos e portanto incapazes de pensamento por oposição a cognição. Uma crítica possível a esta posição salienta o facto de ela esquecer a sensiência e a presença desta em entidades biológicas não humanas. No entanto, para Dennett, a sensiência não é nem idêntica à consciência nem suficiente para esta. A possibilidade de pensamento está ligada à linguagem e à comunicação, e é por isso que pensamento não é sinónimo de cognição e inteligência[745]. Ao contrário da inteligência, o pensamento é uma especificidade humana (outros autores acrescentariam que a consciência, como a inteligência, também não é uma especificidade humana, mas não Dennett, que precisamente liga a consciência ao pensamento[746]). Deste modo, são claramente diferentes entre si as questões seguintes: (1) *será que o comportamento de criaturas não linguísticas é inteligente?* (2) *será legítimo supor que se captura a natureza do comportamento inteligente de criaturas não linguísticas através de pensamentos verbalizáveis?* A resposta à primeira questão é indubitavelmente afirmativa, a resposta à segunda é negativa. Descrevendo linguisticamente como é para uma mente animal ser não se captura a natureza desta. Isto é assim para Dennett não porque a natureza da mente animal exceda, pela sua subjectividade e sensiência, a formulação – esta seria uma posi-

[745] A comunicação por meio de linguagens naturais deverá assim ser considerada em oposição a comportamentos informativos bem sucedidos na "comunicação" de alguma coisa em função exclusivamente do seu componente ostensivo. Um exemplo seriam as danças das abelhas, referidas por R. Millikan, mas não está excluída a possibilidade de casos muito mais complicados. Assim, criaturas que ostentam comportamento informativo não necessariamente comunicam, embora o tipo de comportamento exibido possa ser considerado como um antecedente evolutivo da comunicação. Para a distinção entre comportamento informativo ostensivo e comunicação cf. SPERBER & WILSON 1986. Para um tratamento do problema da origem evolutiva da linguagem (que foi um problema tabu, a certo momento, para os linguistas) cf. PINKER 1994.

[746] Também M. Dummett, um filósofo que certamente não está tão envolvido em investigações cognitivas empíricas como Dennett mas que se tem ocupado das relações entre pensamento e linguagem, tende a defender que os pensamentos de crianças e animais são apenas proto-pensamentos. Eles podem ser expressos em palavras apenas por uma aproximação grosseira, já que tal expressão será quase certamente demasiado precisa e rica para dar correctamente o conteúdo em causa.

ção nageliana – mas, pelo contrário, porque a formulação linguística excede a mente animal[747]. Para Dennett, os animais têm crenças pela mesma razão não polémica que um termostato tem crenças: o seu comportamento é previsível a partir da EI. É certo que a conclusão aplicada a animais é bem menos polémica. A questão importante diz respeito não a saber se os animais têm ou não crenças mas ao grau de sofisticação dessas crenças.

Em 1983, Dennett publicou na revista *Behavioral and Brain Sciences* um artigo que recebeu, como é usual na prática dessa revista, inúmeros comentários. Esse artigo, *Intentional Systems in Cognitive Ethology – The Panglossian Paradigm defended*[748] representa uma primeira incursão mais elaborada nos problemas da etologia cognitiva e uma extensão explícita dos princípios da TSI às mentes animais. O artigo veio a resultar em posterior cooperação com etologistas[749] – nomeadamente Dorothy Cheney, Robert Seyfarth e Carolyn Ristau[750] – e também numa controvérsia com S. J. Gould e R. Lewontin acerca do poder do adaptacionismo na teoria evolucionista. A pro-

[747] Daí a peculiar interpretação que Dennett faz do *dictum* wittgensteiniano das *Investigações* segundo o qual "Se um leão falasse nós não seríamos capazes de o compreender" (WITTGENSTEIN 1953, II^a Parte, 220). Dennett considera que de facto o compreenderíamos. E das duas uma, ou o leão não teria nada para dizer ou seria (de alguma maneira) "humano" (ou melhor auto-consciente) não podendo por isso dar qualquer informação sobre "como é ser um leão" (cf. DENNETT 1996: 18 e DENNETT 1991: 447). Cf. também DENNETT 1998t: 306: «Wittgenstein once said if a lion could speak we could not understand him. I disagree. If a monkey could speak – really speak a language –we could understand him just fine because, if a monkey could speak his way of life his way of life would have to be very much more like ours than it is».

[748] Cf. DENNETT 1987m, *Intentional Systems in Cognitive Ethology*. O artigo foi pela primeira vez publicado em 1983, acompanhado por comentários de Jonathan Bennett, Patricia Churchland, Arthur Danto, Richard Dawkins, Michael Ghiselin, George Graham, Donald Griffin, Gilbert Harman, Nicholas Humphrey, Richard Lewontin, Dan Lloyd, J. Maynard Smith, Ruth Millikan, Carolyn Ristau e B.F. Skinner entre outros, e posteriormente incluído em DENNETT 1987.

[749] Cf. para a descrição das relações do autor com a etologia DENNETT 1998t, *Out of the Armchair and into the Field*. Segundo Dennett, a origem do seu interesse mais próximo pelos problemas da etologia teria sido o artigo *Does a Chimpanze Have a Theory of Mind?* (PREMACK & WOODRUFF 1978) que D. Premack e A. Woodruff publicaram na revista *Behavioral and Brain Sciences* em 1978. Dennett fez um comentário ao artigo (cf. DENNETT 1978t, *Beliefs about Beliefs*). Das discussões com etologistas assim iniciadas – nomeadamente discussões sobre crenças falsas e comportamentos de logro – resultou um convite para um congresso dedicado ao tema "Animal Mind – Human Mind" (Berlim, 1981). Com a sua participação no congresso Dennett aceitou o desafio de mostrar aos estudiosos do comportamento animal a utilidade da adopção da EI para o desenho de situações experimentais e a interpretação de dados. Nessa conferência entrou em contacto com os cientistas envolvidos no projecto de uma etologia cognitiva. Os participantes na conferência pediram a Dennett uma introdução à EI dirigida a não-filósofos e foi desse pedido que resultou o artigo *Intentional Systems in Cognitive Ethology*. O estudo das mentes animais foi assim definitivamente "incorporado" na TSI.

[750] Cf. CHENEY & SEYFARTH 1990 e RISTAU 1991 para referências a essa colaboração.

posta que Dennett dirige aos etologistas no artigo mencionado surgiu num momento em que o interesse pelo interior das mentes animais ressurgia[751]. Essa proposta consiste no uso da TSI como enquadramento justificativo do vocabulário teórico e do desenho de hipóteses em etologia. O método permitiria mesmo a geração de evidência episódica legítima, obviando aos problemas metodológicos sentidos pelos etologistas face ao carácter – precisamente episódico e por isso mesmo muitas vezes inutilizável – de grande parte do seu material. No entanto, a TSI está intimamente ligada ao adaptacionismo biológico e isto faz com que ela seja culpada por associação aos olhos daqueles que consideram que as conhecidas críticas de Gould e Lewontin ao adaptacionismo[752] teriam provado a falência deste. Esta possibilidade de culpa por associação é encarada logo à partida no artigo *Intentional Systems in Cognitive Ethology*, cujos propósitos são, resumidamente, os seguintes:

(1) *Propor aos etologistas a TSI como método descritivo intermédio.* Uma teoria abstracta como a TSI, que caracteriza "máquinas semânticas", pode ser útil aos etologistas cognitivos devido à grande distância existente entre o trabalho prático de observação do comportamento de animais e a validação de modelos neurofisiológicos.

(2) *Ilustrar a proposta com a análise de um particular comportamento animal, supostamente evidenciador de comunicação entre primatas não*

[751] Para dar apenas dois exemplos que fizeram história no início desse "interesse pelo interior", refira-se os artigos *The Social Function of Intellect* de N. Humphrey (HUMPHREY 1976) e *Does the Chimpanze Have a Theory of Mind* de D. Premack e G. Woodruff (PREMACK & WOODRUFF 1978). No primeiro artigo mencionado, o psicólogo N. Humphrey elabora uma hipótese acerca da função social do intelecto, de acordo com a qual teria sido a manipulação do outro indivíduo mais do que propriamente a utilização de artefactos a grande impulsionadora da evolução da inteligência biológica. No segundo artigo, os psicólogos D. Premack e G. Woodruff especulam (com base na interpretação de experiências levadas a cabo com chimpanzés) acerca a existência de uma teoria da mente nos chimpanzés. O seu objectivo é descobrir se os chimpanzés pensam, como os humanos, no que os outros indivíduos pensam, podendo esse pensamento acerca do pensamento do outro indivíduo ser utilizado para explicar o seu comportamento. As investigações iniciadas sob a égide de D. Griffin partem de um pressuposto bastante mais ambicioso: elas visam chegar a descrever como é para vários animais ser.

[752] Cf. sobretudo GOULD&LEWONTIN 1978, *The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptacionist Program*, inicialmente publicado nos Proceedings of the Royal Society, B205 em 1978 (o artigo aparece também em SOBER 1997). A principal ideia defendida é que «não devemos confundir o facto de uma estrutura ser usada de alguma maneira (...) com a razão evolutiva primária para a sua existência e conformação» (GOULD & LEWONTIN 1978:79). Os autores exemplificam o preconceito contrário relativamente ao *design* biológico com o célebre exemplo dos *spandrels* (mucarnas) da Catedral de San Marco em Veneza. Estas estruturas, apesar da sua aparentemente perfeita apropriação para a função que servem, não são senão um produto lateral (*by product*) da utilização de arcos na construcção. Para um enquadramento histórico e filosófico dos problemas do evolucionismo biológico, cf. SOBER 1993, Capítulo 5, *Adaptacionism*.

humanos[753]. Os macacos que servem de exemplo a Dennett – os macacos testiverdes (*Cercopithecus aethiops*) – utilizam vocalizações diferenciadas, nomeadamente emitem diferentes gritos de alarme (aparentemente) associados a diferentes predadores e parecem adaptar o seu comportamento comunicacional a situações concretas de forma versátil.

(3) *Explicitar a ligação entre a TSI e o adaptacionismo biológico, mostrando que os argumentos de S. J. Gould e R. Lewontin contra o adaptacionismo são análogos aos argumentos de B. F. Skinner contra o mentalismo em psicologia e não constituem um obstáculo.* Quer as críticas skinnerianas ao mentalismo quer as críticas ao de S. J. Gould e R. Lewontin ao adaptacionismo apontam, é certo, perigos reais, tais como o excesso de facilidade na explicação permitindo ignorar os detalhes concretos, a intestabilidade e consequentemente a irrefutabilidade, e a incapacidade de distinguir entre a utilidade actual e as razões da origem de alguma coisa. Elas não avançam no entanto objecções fundamentais.

Antes de passar a uma análise do artigo, convém recordar que algumas considerações aplicáveis à cognição animal apareciam já no artigo *Conditions of Personhood,* analisado no capítulo 4. Nomeadamente, era aí proposta uma gradação de tipos de mentes de acordo com ordens de intencionalidade dos SI, o que permitia entre outras coisas enquadrar o comportamento animal de logro. A diferença entre ordens de intencionalidade marcava incipientemente uma diferença crítica entre mentes humanas e animais, então reportada aos critérios griceanos para a existência de comunicação. De novo em *Intentional Systems in Cognitive Ethology* Dennett dá como exemplo de comportamento de logro o pássaro que finge ter a asa partida para afastar o predador do ninho e evoca o solilóquio com o qual R. Dawkins[754] explica as razões de tal comportamento. O solilóquio é evidentemente muito mais inteligente do que o pássaro que supostamente o "implementa", o que não deve constituir impedimento quando se trata de considerar o comportamento como comportamento de logro. O reconhecimento de *free floating rationales* é indispensável na biologia evolucionista e, precisamente, estabelece o parentesco desta com a EI. O adaptacionismo é, na definição de S.J. Gould e R. Lewontin um programa baseado em fé na selecção natural como agente *optimizador.* É esta "fé" que é caricaturada por S. J. Gould e R. Lewontin como *panglossiana*, excessivamente e injustificadamente optimista, nomeadamente por procurar razões de ser para todo e qualquer traço de *design* actual. Ora, de acordo com Dennett, quer o adaptacionismo quer a EI são estratégias para organizar dados e gerar questões e não teorias, sendo, enquanto estratégias, perfeitamente legítimos[755]. O erro dos críticos do adaptacionismo é confundirem uma estratégia heurística com uma pretensão à explicação.

Em *Intentional Systems in Cognitive Ethology* Dennett retoma a ideia das ordens de intencionalidade, considerando nomeadamente que a grande diferença entre tipos de seres mentalmente interpretáveis se relaciona com a capacidade ou incapacidade de adoptar a EI em relação a outros seres. O limiar relevante é a existência um *SI de 2ª ordem, ou psicólogo natural*[756], não apenas interpretável como tendo crenças e desejos mas também como *tendo crenças e desejos acerca das crenças e desejos de outro ser.* Um psicólogo natural pode no entanto perfeitamente ser um "psicólogo natural não pensante", i.e. uma entidade incapaz de ter crenças acerca das crenças acerca de crenças que tem. Na terminologia utilizada por Dennett, um psicólogo natural não pensante não é um SI de 3ª ordem. Estas diferenças de ordem de intencionalidade dos SI são cruciais para o trabalho dos etologistas cognitivos. Estes necessitam, de acordo com Dennett, de uma teoria e de uma linguagem descritiva neutras que lhes permitam apoiar observações e elaborar hipóteses e vir a decidir mais tarde, se e em que sentido determinados animais têm crenças e crenças acerca das crenças de outros animais. Só assim se poderá

[753] Os *vervet monkeys* ou macacos testiverdes (*Cercopithecus aethiops*) estudados especialmente por D. Cheney e R. Seyfarth (cf. SEYFARTH, CHENEY & MARLER 1980, para o trabalho ao qual Dennett se referia no artigo *Intentional Systems in Cognitive Ethology*).

[754] Cf. DAWKINS 1976.

[755] Esta forma de escapar às críticas de S. J. Gould e R. Lewontin relaciona-se obviamente com a discussão acerca do estatuto do *design* referida no Capítulo 2. Como então se afirmou, essa posição oscila entre um realismo e um interpretativismo. De qualquer modo, o estatuto do adaptacionismo é objecto de discussão entre os próprios biólogos, o que sobressai de forma bastante notória dos comentários dirigidos por biólogos ao artigo de Dennett. Os campos dividem-se, deixando de um lado nomes como J. Maynard Smith, M. Ghiselin e R. Dawkins e do outro N. Eldredge e R. Lewontin. R. Dawkins e J. Maynard Smith sublinham o quanto Gould e Lewontin são também, ainda, apesar de todas as críticas, adaptacionistas e nem poderiam deixar de sê-lo como biólogos.

[756] Como já foi referido no Capítulo 4, Dennett vai buscar a expressão ao psicólogo Nicholas Humphrey, que sugeriu a hipótese segundo a qual teria sido um efeito de corrida (cognitiva) ao armamento o responsável, de um ponto de vista evolucionista, pelo surgimento da inteligência mais sofisticada, sendo mesmo a auto-consciência um estratagema para o desenvolvimento e teste de hipóteses sobre o que se passaria nas outras mentes (cf. DENNETT 1996: 120). O efeito é o seguinte: se surge um ser que pensa sobre o pensamento de outro, surgirá a necessidade de um ser que pense sobre o que o outro ser pensa que ele pensa, de modo a estar assim em vantagem, etc.

avaliar se, por exemplo, as aparentemente versáteis trocas comunicativas e vocalizações entre indivíduos correspondem a algo de próximo da linagugem humana. Ora, Dennett crê que a TSI é capaz de cumprir esta função. De facto, como Dennett recorda no artigo *Do Animals Have Beliefs?*[757], os etologistas cognitivos partilham a problemática dos filósofos da mente (e mesmo controvérsias quanto ao carácter simbólico da implementação física do mental, quanto à importância do "sentir-se ser", etc). É evidente que existe uma agenda oculta no interesse de Dennett pela etologia: o caso das mentes animais parece apropriado para reforçar o anti-sentencialismo e a secundaridade da linguagem preconizados na TSI, bem como a indispensabilidade da linguagem para a possibilidade de (auto-) consciência.

Olhando para o comportamento animal, aquilo que se captura teoricamente acerca do comportamento animal através de linguagem mentalista não tem, de acordo com Dennett, "por essência" estrutura de linguagem nem é imediatamente revelador de consciência. No entanto, o vocabulário mentalista é perfeitamente legítimo e de resto, o etologista cognitivo já o adopta (é isso que o caracteriza como etologista *cognitivo*). Simplesmente adopta-o sem justificação e Dennett pretende apresentar a justificação que falta. As sugestões em que a sua incursão pela etologia culmina colocam Dennett em polémica com estudiosos do comportamento animal de orientações diametralmente opostas: por um lado Dennett será acusado de excesso de "romantismo" por B. F. Skinner[758], por outro lado será acusado de excesso de cautela e timidez por D. Griffin[759].

De acordo com Dennett, o etologista cognitivo está na situação do tradutor radical quineano (pior, aliás, pois não está entre os membros de uma tribo *humana*); ele não deve no entanto ter, segundo Dennett, escrúpulos behavioristas. Verá então um animal como um SI, e utilizará linguagem mentalista, para elaborar hipóteses e testá-las. Poderá fazê-lo em virtude de uma suposição de racionalidade. O problema quanto à realidade daquilo que é atribuído só se coloca posteriormente: terão os animais *realmente* crenças e desejos? Serão eles "verdadeiros crentes"? Como se sabe, a questão não pode ser posta em termos absolutos na TSI e Dennett sugere que ela seja substituída pela atenção à diferença entre tipos de SI (de 1ª, 2ª, 3ª ordem, etc).

[757] DENNETT 1998v.
[758] Cf. SKINNER 1983, *A Better Way to Deal with Selection.*
[759] Cf. GRIFFIN 1983, *Thinking about Animal Thoughts.*

A questão é pois saber, quanto a um animal específico e aos seus comportamentos, de que tipo de SI (de 1ª, 2ª, 3ª ordem) se trata.

A sugestão de análise das mentes animais em termos de ordens de intencionalidade obteve uma grande aceitação entre os etologistas, mas não é o único elemento da proposta de Dennett. Uma vez que Dennett defende, como se viu no Capítulo 4, que a existência de pelo menos uma terceira ordem da intencionalidade no falante e na audiência é uma condição necessária para a existência de comunicação, é de prever que a instanciação de espécimes aparentemente linguísticos não seja considerada suficiente para a existência de comunicação se não estiverem presentes as condições griceanas. I.e., a troca de vocalizações entre animais não deve ser imediatamente considerada como comunicação. Saber se um animal é ou não um SI de 3ª ordem é uma questão que apenas pode ser abordada empiricamente, mas é a TSI que permite a sua formulação. A proposta de classificação dos SI define uma escala de mentalidade que enquadra o significado de "comunicação".

O posterior trabalho de campo de Dennett com D. Cheney e R. Seyfarth no Quénia junto dos macacos testiverdes que supostamente exibem comportamentos de comunicação, confirmará dois pontos teóricos importantes sugeridos em *Intentional Systems in Cognitive Ethology.* Em primeiro lugar, o trabalho confirma que os macacos estudados vivem num ambiente em que o segredo é impossível, o que conduz Dennett a propor que a possibilidade de segredo é uma condição necessária para uma capacidade griceana, sofisticada, de comunicação. Ora, de acordo com Dennett, o comportamento de manter um segredo exige a seguintes condições, relativas ao ambiente e aos SI B e J quaisquer: (1) B sabe que s; (2) B crê que J não crê que s; (3) B deseja que J não venha a crer que s; (4) B acredita que pode fazer com que J não venha a crer que s. Apenas se essas condições forem satisfeitas se poderá falar de comunicação. Em segundo lugar, o trabalho confirma que mesmo em comportamentos que parecem astuciosos e inteligentes aparecem "brancos" e falhas. Isto acontece nomeadamente em casos em que a "história astuciosa" não é sustentada até ao fim pelo próprio indivíduo que a inicia, o qual é por exemplo capaz de, perante o indivíduo que desejaria enganar, se comportar de forma inconsistente[760] (por exemplo o indivíduo que emite o grito de alarme

[760] Note-se a analogia desta situação com comportamentos frequentes em crianças.

"perigo de leopardo" desce em seguida da árvore à vista do outro indivíduo em cuja presença o grito foi emitido). Dar conta destas ausências ou falhas na inteligência ou na coerência esperadas é, de acordo com Dennett, muito importante em etologia (o artigo *Why Creative Intelligence is Hard to Find*[761] explorará esta intuição). Um aspecto importante do estudo do comportamento animal deverá ser a tentativa de descobrir de que é que não são (surpreendemente) capazes certos animais inteligentes[762].

Voltando ao material concreto discutido no artigo *Intentional Systems in Cognitive Ethology*, está-se perante gritos de alarme emitidos pelos macacos testiverdes, aparentemente em correlação com a discriminação de diferentes predadores. Os predadores seriam assim classificados e a classificação seria mesmo aperfeiçoada com o tempo. Defendendo teses como esta, os estudos de D. Cheney e R. Seyfarth[763] erguem-se contra a uma certa ortodoxia estabelecida na interpretação comparativa das vocalizações de humanos e primatas não humanos, de acordo com a qual vocalizações de primatas humanos e não humanos seriam profundamente diferentes entre si, nomeadamente nos seguintes aspectos:

Vocalizações Humanas	Vocalizações de primatas não humanos
Voluntárias	Involuntárias
Referenciais	Indexicais
Discretas	Graduadas
Aprendidas	Não modificáveis

Cheney e Seyfath pretendem contestar estes estereótipos[764], e nomeadamente mostrar que os gritos de alarme denotam diferentes referentes no mundo, e que portanto são "referenciais", que os indivíduos jovens começam por cometer erros na "aplicação" dos

[761] DENNETT 1998w.

[762] Alguns exemplos destas situações encontram-se em CHENEY&SEYFARTH 1990: 256, nomeadamente o seguinte caso. Leões tinham morto um búfalo há muito pouco tempo e a carcaça deste estava fresca quando um grupo de babuínos chega. No entanto, estes só se apercebem do perigo representado pelos predadores quando os vêem. Não poderiam ter reconhecido a carcaça fresca como indicação de perigo, de possibilidade de presença dos predadores?

[763] Continuados depois do trabalho no qual Dennett se baseia em 1983. A referência é o livro CHENEY & SEYFARTH 1990.

[764] Cf CHENEY&SEYFARTH 1990, Cap. 4, *Verbal Communication*, para os dados empíricos acerca deste comportamento. Para uma apresentação de estudos recentes (e dos equívocos nesses estudos) sobre linguagem animal, cf. PINKER 1994: 335.

gritos e gradualmente melhoram, existindo portanto aprendizagem, que os indivíduos podem "falsear" gritos de alarme, e que estes são portanto voluntários, etc. Apesar de todas essas características Dennett não crê que seja legítimo considerar que existe comunicação. Noutros contextos, por exemplo em explicações sociobiológicas dos comportamentos de altruísmo[765], ordens elevadas de intencionalidade são supostas sem que se suponha que as razões em causa são de algum modo reconhecidas pelos indivíduos que manifestam os comportamentos. É necessário saber se um esquema idêntico não será suficiente para dar conta de proto-linguagens como a dos macacos testiverdes. Nesse caso, as vocalizações não seriam sintoma de um aprofundamento mental nem da existência de uma interioridade que alberga intenções para si, nem da presença de uma teoria da mente. A comunicação no sentido próprio (griceano) envolve o reconhecimento da mente do outro, e supõe portanto, por exemplo, que os indivíduos são capazes de aperceber discrepâncias entre a maneira como percebem o mundo e a maneira o outro percebe o mundo[766]. Em suma, o etologista cognitivo não deve "ler" directamente nos seus registos (relativos, no caso, às vocalizações) que eles significam o que quer que seja (que os macacos de facto "querem-dizer" (*mean*) alguma coisa com as vocalizações). Saber se as vocalizações significam alguma coisa e o que significam é um problema a ser pensado unicamente em relação com a ordem de intencionalidade dos SI presentes. Em *Intentional Systems in Cognitive Ethology*, Dennett toma como exemplo o caso de um macaco, Tom, que emite na presença de Sam, outro macaco, o grito considerado pelos etologistas como grito de alarme de leopardo. O problema consiste em saber se Tom deve ser interpretado como um SI de ordem elevada (caso em que o grito significaria por exemplo que "Tom quer que Sam acredite que Tom quer que Sam corra para as árvores") ou como não sendo sequer um SI (o grito seria então

[765] Cf. DAWKINS 1976. Cf. também SOBER 1993, Capítulo7, *Sociobiology and the Extension of Evolutionary Theory* para uma exposição das origens, problemas e modelos da sociobiologia, e ainda *The Primacy of Truth Telling and the Evolution of Lying*, in SOBER 1994: 71. E. Sober toma D. Davidson como referência e alvo de crítica. Como se sabe, Davidson estabelece um abismo entre linguagem e comunicação animais e humanas, o que não é o caso de Dennett, cujo pensamento mais facilmente se adapta à proposta de Sober e à possibilidade de uma história natural da "verdade e da mentira".

[766] O não apercebimento desta discrepância foi muito acentuado pelos psicólogos que estudaram o autismo dentro do quadro conceptual da teoria da mente. Cf. BARON-COHEN 1995, onde se afirma que o autista não teria uma teoria da mente "automática", pelo menos não como acontece no caso dos humanos que não sofrem desse défice e que portanto um autista não perceberia sem esforço os outros seres humanos como imediatamente dotados de um interior mental.

uma simples reacção à "ansiedade de leopardo", e pertenceria à mesma categoria que o comportamento involuntário de sobressalto gerado num humano pela aproximação de alguém pelas costas gritando: *Buhh*!!., sendo o efeito conseguido sem qualquer reconhecimento pelo sujeito da intenção de quem pratica o comportamento). O comportamento é analisável de ambas as formas, não sendo por isso de afastar imediatamente a possibilidade de uma intencionalidade de ordem zero, que requer uma interpretação *"killjoy"* (um termo favorito de Dennett, que se poderia traduzir por "desmancha-prazeres").

O problema metodológico consiste em saber o que permitiria decidir entre a atribuição de mentalidade sofisticada e de mera reacção, uma vez que o comportamento é o "mesmo". O behaviorismo é, por opção, uma aposta na alternativa de ordem mais baixa, a ordem zero. No entanto, também ele não é justificado por si (ser behaviorista em teoria da cognição é equivalente a optar, por exemplo, por não utilizar o conceito de "nutrição" em biologia, aceitando apenas uma análise físico-química daquilo que é ingerido). A TSI enquadra hipóteses acerca de "mais mentalidade" e permite analisar experiências no terreno de acordo com essa suposição. Por exemplo, e ainda relativamente ao comportamento dos macacos testiverdes, seria possível criar em campo a situação do rapaz que grita "Lobo!" sem qualquer lobo à vista. Os etologistas gravam vocalizações de um particular indivíduo e emitem-nas de acordo com os parâmetros dessa situação. Os outros macacos deveriam nesse caso deixar de "acreditar" no indivíduo em causa. O enquadramento da TSI permite ainda por exemplo analisar situações como a de um indivíduo que emite o grito de leopardo na ausência de leopardos de modo a assustar e afugentar indivíduos com quem se envolveu em luta. Evidentemente, experimentações geradoras de episódios comportamentais tocam no calcanhar de Aquiles da etologia cognitiva, a evidência episódica, e a má reputação desta. Aqui entra aquilo a que Dennett chama o método Sherlock Holmes, inspirado na maneira como os humanos estabelecem o seu próprio estatuto de SI de ordem elevada (os humanos estabelecem esse estatuto através de biografias episódicas). Ora o método proposto por Dennett aos etologistas consiste precisamente na tentativa de provocação do *tell-tale move*, i.e. do comportamento revelador. Trata-se portanto da geração de evidência episódica em circunstâncias controladas. A evidência episódica será de facto inútil a não ser que se mostre, construindo e controlando a situação experimental, que seria demasiado improvável que ela fosse outra coisa que não aquilo que é.

Diga-se de passagem que estes problemas metodológicos acerca de significação e comportamento animal sublinham uma importantíssima diferença na situação do etologista cognitivo relativamente ao *tradutor radical quineano*. Este último, quando procura assentimentos e dissentimentos a "estímulos" públicos que constituirão a base do manual de tradução, *pressupõe a natureza comunicacional dos nativos*. Ora, Dennett defende que sem a construção de situações a acumulação quineana, comparativa, de "respostas a estímulos" numa situação de terreno nunca será suficiente para provar intenções comunicativas de animais.

Dennett prevê desde logo uma consequência do método proposto: os excessos da atribuição de mentalidade a animais serão desmascarados, confirmando outra das suas intuições básicas relativamente ao estudo das mentes animais. Dennett defende ainda que o estudo revelará sintomas misturados e confusos de intencionalidade mais e menos sofisticada, sem que seja possível uma conclusão clara e coerente quanto à ordem de intencionalidade presente num indivíduo (a situação é aliás de certo modo paralela à humana, já que também a abordagem da vida mental humana gera a ilusão da existência de muito mais definição e precisão de conteúdo do que aquelas que legitimamente se pode supor – esta era, recorde-se, uma conclusão importante de *Brain Writing and Mind Reading*). Mesmo assim, Dennett defende que quer para humanos quer para animais, o traçado do perfil intencional de uma criatura (do seu mundo nocional) será (apesar das zonas nebulosas e falhas) valioso na medida em que é a caracterização de uma competência em termos daquilo que é representado e não em termos de dispositivos e implementação física.

Em reflexões posteriores ao trabalho de campo, nomeadamente em *Interpreting Monkeys, Theorists and Genes*[767], Dennett atenua de alguma maneira esta proposta inicial, sobretudo devido a uma impossibilidade fundamental, ligada ao facto de grande parte das experiências psicológicas de laboratório com humanos envolverem instrucções verbais. Estas são obviamente impossíveis no caso dos animais, o que de acordo com Dennett se traduz no facto de ser em última análise impossível ao experimentador controlar a informa-

[767] Notas apostas ao artigo inicial de 1983 quando este é incluído em DENNETT 1987 (cf. DENNETT 1987:269).

ção possuída pelos sujeitos. É evidente, no entanto, que um tal controlo seria essencial para os procedimentos do método de provocação de evidência episódica esclarecedora proposto. Dennett acaba por concluir que «é quase impossível estabelecer no terreno que macacos particulares foram escudados de uma particular informação»[768]. Isto torna a situação de laboratório mais atraente e portanto inclina a um recuo relativamente à etologia feita no terreno.

Os artigos *Interpreting Monkeys Theorists and Genes, Out of the Armchair and into the Field*[769] e *Cognitive Ethology Hunting for Bargains or a Wild Goose Chase*[770], todos eles escritos posteriormente ao trabalho de campo com Cheney e Seyfarth no Quénia, constituem de certo modo a continuação da abordagem inicial à etologia feita em *Intentional Systems in Cognitive Ethology*. Nestes artigos está em foco uma nova vocalização dos macacos testiverdes, o "MIO[771] *grunt*" (que se pode traduzir por "grunho de mover para espaço aberto"), uma vocalização emitida antes de o indivíduo se deslocar para um espaço aberto e repetida por outro indivíduo. Dennett elabora, em conjunto com Cheney e Seyfarth, várias interpretações das duas vocalizações de modo a verificar delas quais poderão ser eliminadas com base na evidência disponível[772].

1) Estou a ir.	1') Ouvi-te. Estás a ir.
2) Posso ir, por favor?	2') Sim, tens permissão para ir.
3) Segue-me!	3') Sim senhor.
4) Estou cheio de medo.	4') Sim, eu também.
5) Tudo pronto para eu ir?	5') Tudo pronto quando estiveres pronto.
6) Nada à vista?	6') Nada à vista. Tudo sob controlo.

Segundo Dennett, embora seja em última análise impossível decidir que uma destas interpretações é correcta, é útil explicitar as interpretações possíveis. A explicitação permite testar e excluir hipóteses, uma vez tendo avaliado se as situações se passam por exemplo entre indivíduos dominantes e indivíduos dominados, entre machos e fêmeas, etc. Assim, de entre as hipóteses acima lista-

[768] DENNETT 1987: 275.
[769] DENNETT 1998t.
[770] DENNETT 1998u.
[771] MIO é uma sigla para "*Move into the open*".
[772] Cf. DENNETT 1987:273-274, retomado em DENNETT 1998t: 301-302 e DENNETT 1998u: 317-318.

das, a sexta seria a melhor. Foram procuradas situações que a esclarecessem, e, nomeadamente, foi imaginado o caso em que a "resposta" seria falsa (i.e. o emissor aperceberia perigo à vista mas ainda assim emitiria a vocalização). A situação imaginada envolvia uma cobra píton vista apenas por um dos dois macacos. No entanto, o máximo que se pôde fazer foi concluir que a situação seria impossível de montar, devido à anteriomente mencionada impossibilidade de garantir o isolamento da informação numa situação de campo.

Para além destas conclusões deflacionárias quanto ao trabalho em etologia cognitiva, aquilo que Dennett propõe aos etologistas cognitivos como fundamentação das suas práticas é a teoria exposta por exemplo em *Three Kinds of Intentional Psychology*[773] e analisada no Capítulo 2 deste trabalho, i.e. uma visão acerca da relação entre máquinas semânticas e sistemas cognitivos físicos: uma explicação puramente causal, ao nível micro-físico, não está em competição com a explicação que dá os *rationales* e descreve os mundos nocionais. A última palavra acerca da questão da comunicação animal é a declaração da impossibilidade de traduzir qualquer vocalização animal por meio de uma qualquer linguagem humana. Uma vocalização animal não será nunca por exemplo uma questão, uma ordem, um pedido ou uma exclamação puras: ela não faz parte de um sistema que permita tais distinções sofisticadas. As vocalizações alternadamente emitidas e com grande diferenciação de natureza entre os sujeitos de Cheney e Seyfarth, não se qualificam portanto como comunicação entre estes, mesmo que possam ser esclarecedoras acerca dos inícios evolutivos da linguagem humana. Intenções explícitas no comportamento e nomeadamente nos comportamentos de comunicação acerca do ambiente e de outros comportamentos são muito mais raras do que o que poderia pensar, mesmo em primatas (e mesmo em humanos, obviamente)

Dennett afirma, e Cheney e Seyfarth acompanham-no, que quando se avalia o nível de intencionalidade das vocalizações no terreno não se encontra razões para considerar que existem intenções de ordem mais elevada. Mesmo se elas não são reflexos involuntários não é de todo claro que envolvam atribuições de estados mentais a outras criaturas. Elas são basicamente desenhadas para modificar o *comportamento* do outro indivíduo muito mais do que

[773] DENNETT 1987h.

para modificar o *pensamento* do outro indivíduo. Cheney e Seyfarth afirmam que se os macacos que estudam não são SI de ordem zero, no entanto eles não são mais do que SI de 1ª ordem. Eles «reconhecem a associação entre um som particular e um particular comportamento de fuga e fazem uso deste conhecimento para alterar o comportamento de outros indivíduos»[774] e apenas isso[775]. Assim sendo, não é de todo óbvio que apenas porque determinados animais emitem vocalizações diferenciadas e aparentemente apropriadas à situação, eles exibem comportamento linguístico e comunicacional. A conclusão de Dennett é que não se encontrará comunicação animal sem a seguinte escala: (1) complexidade do ambiente, (2) possibilidade de manter segredos, (3) linguagem, (4) pensamento. Do ponto de vista evolutivo isto significa que o pensamento esperou pela linguagem, a linguagem esperou pela possibilidade de segredo, a qual teve que esperar pela apropriada complexificação do ambiente dos indivíduos, a qual teve que esperar pela sofisticação da arquitectura física e cognitiva dos indivíduos.

5.2.2 *Mentes artificiais: o problema do enquadramento na IA*[776].

No artigo *Cognitive Wheels: The Frame Problem of AI*[777] Dennett expõe o problema do enquadramento através da história de um robô, R1, e dos seus sucessores. Os robôs têm que ir buscar uma bateria a um quarto onde existe uma bomba prestes a explodir e falham

[774] CHENEY &SEYFARTH 1990: 148. Cheney e Seyfarth interpretam a tipologia de Dennett de forma ligeiramente desviada (centram-na imediatamente no animal acerca de quem se quer verificar uma hipótese, enquanto Dennett define as ordens a partir do intérprete que assume a EI).

[775] Em DENNETT 1998x, *Animal Consciousness: What Matters and Why*, Dennett desenvolve uma posição geral acerca de consciência animal. Por acreditar que não existe uma linha de separação clara entre organismos conscientes e organismos não conscientes Dennett defende nomeadamente que a dor e o sofrimento não são o mesmo fenómeno. O sofrimento depende da sofisticação da organização cognitiva da entidade, enquanto a dor pode até existir sem sujeito, o que é por exemplo contrário ao que frequentemente os defensores dos direitos dos animais supõem sem mais problemas. Dennett defende ainda que nós não somos muito bons juízes na questão da dor e do sofrimento em outras mentes, pois somos aparentemente rigidamente determinados pelo nosso *design* biológico a empatizar com as entidades que exibam os comportamentos certos (cf. HOFSTDTER & DENNETT 1981: 109, *The Soul of Mark III Beast*, de Terry Miedaner). Definitivamente a exibição de certos tipos de comportamento pode mesmo iludir-nos acerca da presença de consciência em criaturas. No entanto é necessário admitir – e até Peter Singer o admitiria – que pode haver sensibilidade e funcionamento de organismos sem sensiência.

[776] As principais fontes que constituirão o contexto para a exposição da posição de Dennett são PYLYSHYN 1987 e FORD & PYLYSHYN 1996.

[777] DENNETT 1998j.

sucessivamente. R1, que deve supostamente cuidar de si próprio, vem a saber que a preciosa bateria suplente está fechada num quarto onde existe uma bomba relógio prestes a explodir. Localiza o quarto, a chave do quarto e formula um plano para recuperar a bateria. Existe um carrinho dentro do quarto e a bateria está nesse carrinho. R1 pensa que uma acção a que chama PULLOUT (bateria, quarto) fará com que a bateria seja retirada do quarto. Leva a cabo a acção e retira a bateria do quarto antes da bomba explodir. Só que a bomba também estava no carrinho... e, evidentemente, explode. Aliás, R1 sabia que a bomba estava no carrinho, apenas não extraiu desse conhecimento de que a bomba estava no carrinho o conhecimento de que ela "viria" com o carrinho, uma vez sendo este puxado para fora do quarto. Este primeiro fracasso conduz os engenheiros a construir um novo robô, R1D1, capaz de deduzir as implicações das descrições utilizadas na formulação de planos de acção. Uma vez no terreno, também este robô inicia a acção PULLOUT (carrinho, quarto), mas estava ainda a deduzir se puxar o carrinho para fora do quarto mudaria ou não a cor das paredes do quarto quando a bomba explodiu. Os engenheiros decidem então que o robô deve saber distinguir entre implicações relevantes e irrelevantes dos seus conhecimentos de modo a não perder tempo com a exploração das implicações irrelevantes. Constroem então um novo robô. R2D1, esse robô, estava ocupado a ignorar as implicações irrelevantes quando a bomba explodiu. A moral da história é que todos estes robôs sofrem do problema do enquadramento.

O *frame problem* ou problema do enquadramento diz respeito, como já foi referido, ao facto de as acções de um agente no mundo requererem algo como uma actualização constante das representações internas desse mundo. Algumas coisas no mundo permanecem inalteradas, algumas coisas mudam, inclusivamente devido às acções dos agents e o planeamento deve incluir previsões relativas a essa situação. Não existe no entanto acordo quanto à natureza exacta do problema do enquadramento e a posição defendida por Dennett que a seguir se expõe é apenas uma particular interpretação. Em *Cognitive Wheels: The Frame Problem of AI*, Dennett define nos seus próprios termos o problema do enquadramento, que considera um problema epistemológico novo e um problema que os humanos naturalmente resolvem[778]. Quando Dennett defende que

[778] Ou melhor, resolvem não resolvendo...de um modo que é, na expressão de Dennett "*good enough for government work*".

o problema do enquadramento é um problema epistemológico totalmente novo, por cuja descoberta a IA é responsável, ele pretende afirmar que o problema do enquadramento não se identifica com o problema humeano da indução, mesmo se existem semelhanças relevantes entre os dois. De acordo com Dennett, o problema do enquadramento é mais geral do que McCarthy e Hayes supõem quando afirmam que ele respeita a estratégias utilizadas por sistemas no enquadramento do planeamento e tem a ver com a relação da inteligência com o futuro, mais especificamente com as expectativas. O planeamento de alternativas de acção envolve cálculos acerca de futuro(s). Tal como está expresso no título do artigo *Producing Future by Telling Stories*[779], Dennett reporta o problema do enquadramento à produção de futuro em tempo real, característica dos agentes que fazem planeamento. Assim, o problema do enquadramento consistiria em encontrar uma representação útil e compacta do mundo, que possibilitasse antecipações em tempo real para o planeamento e o controlo da acção de um sistema. De acordo com Dennett, nós, os humanos, possuímos já a solução do problema. Que possuímos algum tipo de solução do problema não é muito polémico, já que obviamente – pelo menos em geral[780] – não nos comportamos como os robôs do exemplo numa situação idêntica. Mas o facto de o problema estar por natureza resolvido no tipo de ser que somos não significa evidentemente que a solução seja

[779] DENNETT 1981.

[780] Cf. DAMÁSIO 1994 para situações (ligadas a lesões nos lobos frontais) em que humanos talvez se comportem de forma semelhante. A. Damásio descreve o caso de Elliot. Devido a um menigioma, tinha sido retirado tecido do lobo frontal de Elliot. Damásio descreve do seguinte modo um dia da vida de Elliot: «para começar a manhã e preparar-se para um dia de trabalho, necessitava de incentivo. Uma vez no trabalho, era incapaz de utilizar o seu tempo adequadamente, e não era possível confiar que respeitasse os prazos prometidos. Quando o trabalho requeria a interrupção de uma actividade para passar a ocupar-se de outra, ele podia todavia persistir na primeira, perdendo aparentemente de vista o seu objectivo principal. Ou podia interromper a actividade com que estava ocupado para se dedicar a algo que o cativasse mais naquele preciso momento. Imagine, por exemplo, uma tarefa que envolva a leitura e a classificação de documentos de um determinado cliente. Elliot lê-los-ia, compreendendo inteiramente a importância do material, e saberia certamente como separar os documentos de acordo com a semelhança ou a disparidade do seu conteúdo. O problema consistia na possibilidade de abandonar subitamente a tarefa de classificação que tinha iniciado para se pôr a ler um desses papéis, de forma cuidadosa e inteligente, durante todo o resto do dia. Elliot podia passar uma tarde inteira a ponderar sobre o critério de classificação que devia ser aplicado (...). O ritmo do trabalho era quebrado. Podia dizer-se que o passo específico do trabalho em que Elliot tinha encalhado estava na realidade a ser executado com demasiada perfeição, mas à custa do objectivo global. Podia dizer-se que Elliot se tinha tornado irracional em relação ao enquadramento necessário para a sua prioridade principal, enquanto dentro de enquadramentos menores, que diziam respeito a tarefas subsidiárias, as suas acções eram desnecessariamente pormenorizadas» (DAMÁSIO 1994: 56).

explicitamente possuída. Para Dennett, essa solução, uma vez explicitada, revela ser *o hábito humeano*. Essa é de resto a semelhança entre o problema do enquadramento e o problema humeano da indução (ou melhor, a solução apresentada por Hume para o problema humeano da indução). De acordo com Dennett, o problema do enquadramento coloca-se devido à ausência do hábito humeano nos agentes artificiais.

Os dados do problema, que têm a ver com aquilo a que se chamará o comportamento de planeamento de um agente, e que se supõe agora serem comuns a qualquer criatura no mundo capaz de fazer planeamento, são os seguintes: (1) é necessário ao agente calcular as implicações dos seus actos, incluindo os efeitos laterais destes, a partir das descrições (representações de conhecimento, tenham a forma que tiverem) com as quais formula, à partida, os seus planos de acção no mundo; (2) não é possível que o agente dedique demasiado tempo a calcular as implicações dos seus actos; (3) é por conseguinte necessário distinguir de alguma maneira, no cálculo das implicações dos actos, implicações relevantes e irrelevantes; (4) não é possível "perder demasiado tempo" a ignorar as implicações relevantes dos actos.

Apesar de Dennett defender que o problema está resolvido nos humanos através do bom funcionamento do hábito humeano, ele não considera que o problema do enquadramento específico da IA diga respeito ao nível fenomenológico considerado por Hume na sua teoria do hábito. O nível fenomenológico é o nível no qual nós, humanos, "naturalmente" resolvemos o problema, enquanto o problema na IA é um problema de mecânica, i.e. relativo ao que pode fazer funcionar o raciocínio de um agente, e não relativo ao que o agente sente ou pensa. Noutras palavras, mesmo se o hábito constitui a solução para o problema do enquadramento, a própria criatura não tem qualquer ideia acerca do funcionamento do hábito em si. Essa opacidade é perfeitamente enfatizada por Hume, que fala de um "maravilhoso instinto" das nossas mentes, e se refere à razão-por-instinto em que o hábito consiste. É por esta razão que o problema do enquadramento é um problema novo, e não o problema da indução: «Mesmo que o problema da indução estivesse resolvido, e o agente acreditasse tudo o que devia acreditar, não se saberia ainda como representar, de modo a poder ser utilizada, a informação »[781].

[781] DENNETT 1998j.

O problema do enquadramento na IA é assim acerca do modo como a informação deve estar representada para estar disponível de modo a ser relevantemente utilizada na acção. Não se trata de um problema de superfície mas de bastidores. Sabemos que o problema do enquadramento não se coloca ao nível fenomenológico, porque a esse nível as coisas são individuadas pela significação. Isto não constitui problema para nós próprios que a partir de dentro nos apercebemos assim, como pensando categorias e propriedades individuadas pela significação no nosso fluxo de consciência. O problema é que os significados na IA não são gratuitos, i.e. não estão naturalmente dados.

Uma vez que para Dennett o problema do enquadramento se encontra naturalmente resolvido nos humanos devido ao hábito humeano, recapitular-se-á brevemente os pontos mais importantes da teoria humeana, de modo a esclarecer em quê, exactamente, o hábito humeano permite aos humanos a resolução do problema do enquadramento. O hábito é introduzido por Hume no âmbito do problema da indução e enquadra-se num projecto geral de solução céptica das dúvidas cépticas. O problema da indução tal como Hume o configura é o problema da legitimidade das inferências que partem de *matters of fact* observados para *matters of fact* não observados (passados, futuros, presentes para além do âmbito do observado). Segundo Hume estas inferências convocam a causalidade e, no entanto, feito um inventário daquilo que está presente numa conjugação de eventos considerados como causa (C) e efeito (E), encontra-se contiguidade, conjunção constante, prioridade no tempo do evento C em relação ao evento E mas não se encontra mais nada, quer dizer, não se encontra "causalidade". Hume, como se sabe, atribui à própria mente, com apoio na experiência, a conexão entre os fenómenos. Que para estabelecer a conexão a mente se apoia na experiência e não na razão é o ponto da conhecida experiência imaginária de Adão, vindo ao mundo com toda a inteligência mas sem experiência. A conexão não é portanto uma questão de princípios últimos residindo no mundo exterior, antes reside na mente. Mas de que forma reside "a conexão" na mente? É neste ponto que entra a teoria do hábito e da crença. A crença humeana é um sentimento, um *feeling to the mind*, descritível apenas em termos metafóricos como força, firmeza, peso, influência nas paixões e pensamentos de uma pessoa. A crença não interfere com o conteúdo cognitivo daquilo que é acreditado, não depende da vontade, nem pode ser manipulada arbitrariamente. Ela acompanha o instinto de indução e serve a Hume para dividir o concebível em duas categorias, aquilo que é acreditado e aquilo que não é acreditado. Aquilo que é acreditado não se distingue do seu contrário por este ser inconcebível, já que por princípio os contrários dos *matters of fact*, i.e. o caso em que algo determinado não é o caso, são concebíveis. Aquilo que é acreditado distingue-se do seu contrário unicamente por este não ser acreditado: afirmar que o sol não se levantará amanhã não é menos inteligível do que afirmar que ele se levantará amanhã. Ora a crença é segundo Hume gerada pelo costume ou hábito, que é uma *gentleforce* actuando sobre as mentes humanas. Hume não expõe as características da crença de modo a contrastar esta com qualquer coisa que seria o conhecimento (o ponto da teoria do conhecimento de Hume é exactamente mostrar que mesmo enquanto *reasoning creatures* os humanos são *feeling creatures*) mas sim para afirmar que nunca daremos um passo para fora das nossas mentes e do funcionamento da crença. Hume pretende aliás defender que o funcionamento da crença é suficientemente confiável embora os conteúdos assim possuídos não sejam de modo algum categóricos, e não facultem uma posse de essências ou princípios ocultos da natureza. Como se sabe, não há, de acordo com Hume, escolha a não ser entre uma falsa razão (esta razão, em que tudo é crença) e nenhuma razão. Daí a sugestão do hábito como solução céptica das dúvidas cépticas: se «a razão parece incapaz de afastar nuvens, a natureza será suficiente para tal propósito»[782], precisamente devido à instalação do hábito no tipo de criaturas que são os humanos. É esta última posição que justifica as interpretações evolucionistas da filosofia humeana do hábito.

Repare-se que o facto de não existir qualquer coisa como um fundamento lógico da indução não significa que esta tenha um fundamento meramente psicológico e como tal irrelevante. O fundamento psicológico do hábito, também chamado "instinto" e considerado como implantado em nós pela natureza, aproxima-se antes de uma explicação naturalista e adaptacionista. O naturalismo de Hume é notório aliás não apenas na abordagem da adequação ao ambiente do funcionamento do hábito como também no espírito com que Hume aborda o funcionamento dos

[782] HUME [1739]: 316.

princípios que geram e preparam a crença na mente, i.e. os princípios de associação. O funcionamento destes é tão inacessível aos seres mentais como o funcionamento de qualquer outra parte da natureza. Desconhece-se tanto a essência da mente como se desconhece a essência do mundo externo e apenas por experiência e observação se pode ir constituindo a *science of human nature*, i.e. a ciência da mente. Mas, basicamente e para o que nos interessa, o hábito humeano, apesar de ser produzido pelo funcionamento de uma mente cuja natureza é de acordo com Hume imaginação, não é uma propensão irregular. Hume declara-o aliás uma sabedoria da natureza essencial à sobrevivência, uma condição necessária da acção humana, que assegura uma harmonia pré-estabelecida entre o pensamento da criatura e a natureza exterior (aquilo a que de um ponto de vista evolucionista se chamaria precisamente adaptação).

O facto de Hume afirmar que o pensamento se apoia no hábito e não na razão não é sintoma de um cepticismo irracionalista, ou de um pessimismo quanto à possibilidade de conhecimento. No entanto é certo que o hábito ou instinto não constitui um funcionamento que possibilite a fundamentação auto-controlada das crenças possuídas de modo a assegurar por exemplo a coerência destas entre si e esse é um dos aspectos que Dennett quer sublinhar com a aproximação do hábito humeano ao problema do enquadramento. Hume baseia a sua descrição do hábito em dados fenomenológicos. Ora, o auto-acesso consciente àquilo a que Hume chama hábito é superficial ou "isolado" relativamente aos muitos níveis do processamento de informação que está a ocorrer no agente. Cada acção de um humano no mundo pressupõe (aparentemente) um conjunto imenso de factos banais, não explícitos ao auto-acesso consciente, mas sem os quais a acção do agente não seria a acção que é. O que acontece então é que, de algum modo, um humano está desde logo na posse de determinada informação que o adequa ao seu ambiente quando age. Pelo contrário na IA a tarefa de delimitação do planeamento começa do zero de pressuposições no agente. Noutras palavras, o problema do enquadramento resulta do facto de a IA lidar com a "tábua rasa" que as mentes humanas nunca foram, tornando necessário explicitar a miríade de factos banais pressupostos na mais mínimas acções. É neste quadro que se torna problemática a capacidade de trazer conhecimento possuído e relevante para a janela de pensamento que estru-

tura uma acção. Considerando que o que importa é que aquilo que o agente tem que saber esteja instalado, i.e. considerando que a distinção entre inato e adquirido não é importante, o problema tem ainda assim dois aspectos diferentes: um aspecto sintáctico (no qual se trata do formato do sistema eficiente para representar e armazenar a informação em causa) e um aspecto semântico. A controvérsia em torno do problema do enquadramento gira em torno destes seus dois aspectos.

A convicção de Dennett quanto ao aspecto sintáctico do problema é que um sistema eficaz para representar a miríade de factos banais terá as características seguintes: (1) não será um sistema de axiomas e derivações de todos os factos; (2) terá que "pôr muitos factos de uma só vez" no conhecimento do agente devido às limitações de espaço e sobretudo de *tempo*. Os constrangimentos que de acordo com Dennett necessariamente se impõem à consideração do aspecto semântico do problema do enquadramento, relativo a que informação deve ser instalada, são, por seu lado, os seguintes : (1) informação que não é acedida pelo agente quando é necessária, em tempo real, é como se não existisse; (2) a inteligência não consiste na resolução de problemas dado todo o tempo do mundo mas na resolução de problemas sob pressão temporal.

É nestas condições que é preciso desenhar um sistema capaz de fazer planeamento a partir de elementos seleccionados do seu conhecimento. A introspecção humana mostra que o planeamento consciente se faz mediante passos determinados, esquematizáveis. Evidentemente listar esses passos não significa que se saiba como é que eles conduzem àquilo a que conduzem (a uma selecção correcta de conhecimentos pelo agente, a não considerar um excesso de possiblidades antes de agir). Os passos do planeamento são: (1) Esboço da situação; (2) O agente imagina-se a levar a cabo a acção; (3) O agente imagina o resultado da acção na situação; (4) O agente avalia o resultado da acção na situação. Mesmo em tarefas triviais (o exemplo de Dennett é a tarefa "ir à cozinha buscar pão" executada por um humano) estes passos são cumpridos, ou melhor, devem estar a sê-lo, embora não o saibamos directamente pois precisamente a execução situa-se "abaixo" do nível de acesso da introspecção. É no entanto simples comprovar que a informação deve estar a ser utilizada. Basta utilizar contrafactuais: se por exemplo o agente soubesse que o pedaço de pão para o qual estende a mão estava envenenado não agiria assim, se o agente soubes-

se que o degrau das escadas estava solto não se apoiaria nele ao descer até à cozinha para ir buscar pão, etc. O mais trivial comportamento humano de planeamento, como o referido, é constantemente sensível a informação de fundo, que deve estar a ser examinada mesmo que seja de forma inconsciente e rápida. Dennett faz ainda notar o seguinte acerca dos humanos enquanto agentes capazes de planeamento e criaturas humeanas: (1) *Os agentes erram e o seu planeamento não é infalível*; (2) *Os agentes deixam-se guiar por rotinas estereotipadas espantosamente insensíveis a mudanças no ambiente*[783]. Apesar disso, produzem constantemente expectativas confiáveis (sabe-se que elas estavam lá pelo espanto sentido quando não são satisfeitas).

Quando Dennett afirma que a abordagem dedutiva não é uma boa abordagem do pleaneamento, não está a condená-la por ser psicologicamente irrealista mas por apenas funcionar em casos muito triviais. Axiomas sobre "não-mudança" em sistemas de representação do conhecimento do agente não são também a solução. Um sistema que opera sob a suposição tácita de que nada muda a não ser aquilo acerca de que se afirma explicitamente que muda tem como problema o facto de não existirem acções isoladas dos agentes. As acções envolvem efeitos laterais e se não se disser tudo explicitamente "as coisas não se comportam". Por outro lado se o sistema levar o agente a considerar todos os efeitos laterais, afundá-lo-á sem dúvida na irrelevância. A regra para um agente num ambiente complexo é ignorar muita coisa. Como no mundo real, as soluções de acção de um agente podem ser invalidadas pela adição de novos elementos ao problema. É essa a razão por que sistemas capazes de inferências não monotónicas (em que a adição de premissas altera o que pode ser provado a partir das premissas iniciais) são propostos por alguns autores como solução para o problema do enquadramento. Dennett não considera que seja esta a solução certa: de um certo ponto de vista tais sistemas (por exemplo sistemas de lógica não-monotónica e lógica temporal) constituem um melhoramento radical relativamente à abordagem dedutiva, de uma perspectiva ligeiramente diferente

[783] Imagine-se que alguém compra leite que se encontra sempre na mesma prateleira de um determinado supermercado. Essa pessoa pode dar por si a procurar afanosamente o leite na prateleira durante algum tempo só depois caindo em si para verificar que hoje a prateleira está vazia (e que ela aliás estava a ver isso desde que ali chegou).

[784] DENNETT 1998j: 200.

eles são mais do mesmo[784]. A objecção que Dennett coloca à proposta tem a ver com a sua falta de relação com o que se passa nos bastidores do sistema. Segundo Dennett, o defeito de muitas propostas actuais para lidar com o *frame problem* é o facto de elas serem "rodas cognitivas". Rodas cognitivas – a expressão que dá o título ao artigo *Cognitive Wheels* – são propostas de *design* de arquitecturas cognitivas (desde o nível semântico até ao nível dos neurónios) profundamente anti-biológicas, mesmo se elegantes do ponto de vista tecnológico. Ora um modelo cognitivo pode descer directamente de um nível fenomenologicamente reconhecível para a implementação desse nível através de "rodas cognitivas" sem lançar qualquer luz sobre o modo como os humanos conseguem ter aquela fenomenologia. Isso não é especialmente problemático por si, mas certamente não permite uma solução do problema do enquadramento. Outra hipótese de solução do problema do enquadramento consiste em repensar totalmente o nível semântico, pôr de lado a representação do conhecimento através do cálculo de predicados e a ideia segundo a qual aquilo que é acreditado são proposições. A interpretação dessas fórmulas parte (*parses*) injustificadamente o mundo em objectos e predicados. Pelo que se viu no Capítulo 2, Dennett não pode colocar-se contra esta posição e ele admite mesmo que seria preferível modelizar a capacidade do agente para manter o rastro das coisas de uma forma mais directa do que através de nomes e predicados, deixando implícita a informação sobre o que é tomado como verdadeiro. Uma outra hipótese ainda seria considerar a arquitectura serial dos agentes artificiais como origem do problema do enquadramento. No entanto a opinião final de Dennett é que o problema do enquadramento se relaciona não com qualquer das situações apontadas mas sim com *a ausência, em agentes artificiais, da forma como a mente humana permite a produção de futuro*[785].

Apenas criaturas que produzem versões-de-futuro relevantes sofrem do problema do enquadramento. Estas versões do futuro são extensões de algo que existe em criaturas mais simples: o facto de estarem instaladas para deixar que o mundo as avise. O problema original de qualquer agente, por exemplo de um animal, é "O que é que eu hei-de fazer em seguida?" e as variações nas respostas dizem respeito à finura do grão da representação

[785] Cf. DENNETT 1998l, *Producing Future By Telling Stories.*

do mundo. Daqui segue-se uma alternativa: o agente pode agir ao acaso e esperar pelo melhor ou representar o mundo e utilizar representações para guiar o seu comportamento. Embora seja tentador identificar a imagem manifesta dos humanos com aquilo que a linguagem natural representa, embora seja tentador considerar que todo o pensamento se basearia nesses termos (que por exemplo robôs vivendo entre os humanos partilhariam a imagem manifesta dos humanos), esse é um passo injustificado. O passo seguinte de Dennett é mostrar como é que o hábito humeano se relaciona com a imagem manifesta dos humanos. De um ponto de vista humeano, *a nossa imagem manifesta é constituída por um conjunto de hábitos de expectativa, derivados de experiência por condicionamento*. Essas expectativas, retomando o termo de Z. Pylyshyn, não são cognitivamente penetráveis e nada na implementação dos hábitos (que são ideias práticas, razão por instinto, feixes de como-lidar) na criatura obriga, por exemplo, a preservar a consistência do que eles instauram em termos de imagem manifesta. A questão que se coloca ao agente é saber explorar estes hábitos representacionais. Ora, de acordo com Dennett, o agente explora os seus hábitos através de meta-estratégias, nomeadamente a atribuição às coisas salientes na imagem manifesta de hábitos humeanos para ver o que acontece, mapeando assim espaços abstractos em espaços comportamentais. Estes espaços são apenas análogos aos "*frame axioms*" e a tarefa de os aperfeiçoar é indefinidamente adiada: o sistema suporta perfeitamente tê-los mal ajustados e desembaraçar-se ainda assim o melhor possível. Agentes humanos (humeanos...) fazem as coisas relevantes na situação corrente e protegem, quando isso importa, as persistências que importam. Por exemplo, em situações complexas, as funções são fechadas numa caixa negra chamada "agente", ao qual é deixada a decisão (i.e os agentes humanos e humeanos adoptam a EI relativamente a porções do mundo): considerar alguma coisa no mundo como outro agente serve para compartimentalizar a ignorância e a EI seria assim ela própria um particular hábito humeano instalado pela evolução nos humanos e noutros sistemas cognitivos. Nos agentes humanos os altos níveis de produção de versões de futuro são feitos através de histórias e narrativas, que andam sempre em busca dos seus conteúdos. *É isto que faz, de acordo com Dennett, de forma rápida e barata um planeador central imune ao problema do enquadramento.*

5.2.3 *A incorporação das mentes: perturbações do funcionalismo.*

«When you make a mind, the materials matter»,
Daniel Dennett[786]

De acordo com N. Humphrey[787], o aspecto mais interessante do livro que Dennett dedica aos vários tipos de mentes, *Kinds of Minds* (KM), é o quanto, em muitas das teses aí defendidas, Dennett se distancia do funcionalismo descerebralizado clássico exemplificado no início deste capítulo através das ideias de H. Simon quanto ao artificial. Esse afastamento é notório numa das ideias organizadoras de KM (que revela a influência de A. Damásio), a ideia segundo a qual o corpo próprio deve ser considerado parte integrante da mente de uma pessoa[788]. Como se referiu no início do Capítulo 3, a abordagem que Damásio faz do problema da consciência tem visado a questão da "propriedade" do mental por via da representação da corporeidade própria em vários níveis e de diversas formas no cérebro e está nos antípodas do tratamento da mente característica do funcionalismo puro. Nestes tratamentos[789] o Eu é apenas um símbolo-do-Eu, enquanto que a hipótese de Damásio vai no sentido de considerar que a "propriedade" é mais do que isso: os materiais que implementam a representação do Eu importam e estão de certo modo presentes eles próprios como "conteúdo".

O funcionalista clássico decompõe a mente em sub-mentes desincorporadas, cada uma correspondente a uma função e a uma caixa num fluxograma. Ora, em KM, Dennett, que sempre se declarou (e continua a declarar-se) funcionalista, vem a considerar que os homúnculos, i.e. as partes da decomposição funcional do mental, não são apenas sub-mentes mas também sub-corpos: «Não é possível separar-me do meu corpo deixando um belo e limpo bordo, como os filósofos por vezes supuseram. O meu corpo contem tanto de mim, os valores, talentos, memórias e disposições que me fazem aquilo que sou, como o meu sistemas nervoso»[790]. Em suma, a hipótese nova em

[786] DENNETT 1998b: 76.
[787] HUMPHREY 1997.
[788] Cf. DENNETT 1996, 3, *The Body and Its Minds*, DAMÁSIO 1994 e DAMÁSIO 1999. Como já se afirmou, a grande crítica de Damásio ao MEM é que este seria a teoria de uma *selfless mind* (cf. DAMÁSIO 1992), uma teoria de apenas um dos dois aspectos envolvidos no problema da consciência, aquele a que Damásio chama o filme dentro da cabeça, i.e. o fluxo fenomenológico, deixando de fora o sentido de pertença do filme, a "propriedade" da fenomenologia, que está, de acordo com Damásio, relacionada de várias maneiras com um corpo próprio particular.
[789] Neste trabalho representados por exemplo pelas ideias de Dennett e Hofstadter apresentadas nos Capítulos 3 e 4.
[790] DENNETT 1998b: 77.

KM é que afinal pode ser importante num certo sentido aquilo de que uma mente é feita. Os materiais de uma porção de matéria "mentalizada" importam. O corpo biológico, ao contrário dos dispositivos periféricos de um computador clássico, não é um mero auxiliar para a recolha de informação do exterior, informação a fornecer, nesse caso, a um sistema nervoso, o qual poderia ser descrito como sistema de controlo desse corpo em termos totalmente funcionais, voltando depois o corpo em entrar em cena para possibilitar a execução de acções em resposta. Embora sendo sustentável até certo ponto, o funcionalismo encontra problemas pelo facto de a realização física das funções mentais não ser tão indiferente quanto seria teoricamente desejável. Como nota Humphrey[791], sempre foi suficientemente claro que a ideia-chave do funcionalismo, a ideia de realizabilidade múltipla correspondente à separação entre "natureza formal" e "natureza material", fazia todo o sentido relativamente ao processamento central e à fixação de crenças complexas, mas não fazia nem nunca fez sentido relativamente à recolha de informação sensorial e à implementação motora da acção. Noutras palavras, quando se trata das periferias, dos lugares de input e output dos sistemas cognitivos, a realizabilidade múltipla não faz sentido: a realização física importa muito pois «a composição física de transdutores e efectores é ditada pelo trabalho que eles têm que fazer»[792]. O que Dennett vem afirmar em KM é que a realização física talvez importe não apenas nas periferias mas absolutamente em geral (*all the way in*), ao nível do sistema nervoso na sua totalidade. Por um lado, os sistemas nervosos evoluíram como sistemas de controlo em organismos nos quais já existiam outros sistemas de controlo (nomeadamente hormonais) distribuídos, "sobre" os quais os novos sistemas de controlo foram instalados. Por outro lado estruturas chave-fechadura químicas, são ubíquos no próprio sistema nervoso e essa é uma importante razão pela qual os materiais importam. *Nada disto impede que as mentes dos SI globais continuem, de acordo com o princípio funcionalista da divisão do trabalho, a ser consideradas como constituídas por sub-sistemas. No entanto um certo dualismo entre o mental e o corpóreo, apesar de tudo sempre latente nas teorias funcionalistas, vê-se perturbado pela indistinção entre o-que-é-implementado e os materiais-da-implementação que a ideia das mentes do corpo traz consigo.* Dennett não pretende concluir em KM que o funcionalismo está errado: «Não há mais raiva ou medo na

[791] HUMPHREY 1997. Dennett admite-o desde logo (DENNETT 1996:74).

[792] DENNETT 1996:75. Para alguma coisa, algum material, servir para detectar luz, por exemplo, tem que ser fotossensível.

adrenalina do que há tolice numa garrafa de *whisky*. Estas substâncias, *per se*, são tão irrelevantes para o mental como a gasolina ou o dióxido de carbono. É apenas quando as suas capacidades para funcionarem como componentes de sistemas funcionais maiores depende da sua composição interna que a sua assim chamada "natureza intrínseca" importa»[793]. Essa é a intuição básica do funcionalismo e ela mantém-se. No entanto, o que se sugere em KM é que: (1) o funcionalismo é uma hipótese filosófica e não uma constatação sem problemas acerca da engenharia dos sistemas cognitivos e que (2) a hipótese que diz respeito ao corpo total e não apenas ao cérebro. A insuficiência desta correcção continua a ser, de acordo com Humphrey, o facto de Dennett não a relacionar com a consciência. De facto, Dennett simplesmente não quer admitir que da sua constatação de que os materiais importam é possível passar à hipótese de que o corpo faz mais diferença no que respeita à consciência do que aquilo que o Modelo dos Esboços Múltiplos admite[794].

[793] DENNETT 1996: 76.

[794] É, como se sabe, o que o próprio Humphrey defende na sua pouco ortodoxa abordagem da questão dos *qualia* (HUMPHREY 1995). Segundo Humphrey, estar consciente (por contraste com ter intencionalidade, i.e. representar o mundo exterior) é sentir e a consciência sensorial é uma actividade que teria evoluído a partir das respostas de aceitação e de rejeição de organismos. Há portanto uma especificidade no modo de "representação" ou de subjectividade auto-cêntrico, o qual não tem que envolver nenhuma referência ao exterior e diz respeito a uma agora-idade e a uma "propriedade". A função do sentir é prover o organismo de representações do que se está a passar agora consigo. Sentir é um modo de representar o que acontece na superfície do corpo, totalmente distinto do modo envolvido na percepção ou representação do ambiente externo, que é alo-cêntrico e para a qual finalmente não importa "quem" é informado mas apenas a própria informação relativa ao exterior. É o sentir, destacável do perceber, que se caracteriza pela privacidade, por estar localizado no espaço do corpo, por ser modalmente específico e por ser no tempo presente. Ora, de acordo com Humphrey essas são características de actividades corporais. O sentir seria então uma actividade do corpo próprio e a essência do ser consciente.

Terceira Parte: Os Problemas

Capítulo 6

Fisicalismo, Conteúdo e Consciência: da filosofia da mente à ontologia.

6.1 *Primeiro e fundamental ponto crítico para a avaliação da TSI: o fisicalismo e a irredutibilidade da intencionalidade.*

Afirmou-se na Introdução que a pertinência de uma teoria filosófica se avalia pelo esclarecimento que ela produz relativamente ao conhecimento e compreensão previamente operantes, nomeadamente em áreas científicas. Ora, o objectivo da filosofia de Dennett é precisamente acompanhar a investigação empírica da cognição e propor a TSI como meta-teoria. O problema neste momento é avaliar em que medida a TSI, uma particular teoria fisicalista do conteúdo e da consciência, constitui a melhor abordagem filosófica da ciência cognitiva. Uma abordagem *fisicalista* da ciência cognitiva tal como a TSI é uma teoria do interior mental do mundo, considerado este como sendo de natureza fundamentalmente física. Sob o título de "interior mental do mundo" caem características como a intencionalidade e a consciência bem como conformações mais sofisticadas, nomeadamente humanas, a que estas possam dar origem, como a pessoalidade e a acção livre. Isto significa que a TSI tem a pretensão de analisar a subjectividade sob vários aspectos, propondo várias figuras do sujeito, desde aquela que foi neste trabalho considerada fundamental, a figura do intérprete unificador, até ao Teatro Cartesiano, aos *qualia* da experiência, ao centro de controlo virtual num sistema de agentes relativamente independentes, ao avaliador forte, responsável pela identidade pessoal e pela racionalidade da vida mental própria. Algumas dessas figuras são aceites e

incorporadas na teoria, nomeadamente e com âmbitos diferentes (respectivamente epistemológico e metafísico, cognitivo e moral) o intérprete unificador, o centro virtual de controlo e o avaliador forte. Outras aparecem para serem afastadas e desconstruídas: é o caso do eu naturalmente unificado, prévio à experiência e à identidade pessoal, de (uma certa noção de) Teatro Cartesiano e de (uma certa noção de) *qualia*. De uma tal selecção das figuras da subjectividade resulta uma nova teoria do sujeito cuja coerência constitui um problema.

Penso que se se pretende fazer a partir da TSI afirmações importantes acerca da natureza da cognição e da subjectividade se torna necessário clarificar determinados pontos insuficientemente explícitos do pensamento de Dennett. Defenderei no que se segue que se as teses acerca da EI, que apontam a subjectividade como responsável por uma unificação constitutiva da objectividade, nomeadamente da objectividade do mental são as mais fundamentais de todo o edifício (sem que possa deixar de se considerar ao mesmo tempo o sistema caracterizado por tal subjectividade como sendo "constituído", tal como os modelos mecânicos da cognição o descrevem), outras teses relativas à subjectividade, e muito especialmente o núcleo da posição deflacionária quanto à consciência não lhes fazem justiça. Neste último capítulo haverá assim um afastamento relativamente aos textos de Dennett, que orientaram o percurso até agora traçado, de modo a avançar sugestões relativas a um aprofundamento da TSI em termos de metafísica e ontologia[795]. Tais sugestões partem do problema que a irredutibilidade da intencionalidade, latente na figura do intérprete unificador (nunca "reduzido"), coloca ao professado fisicalismo da TSI[796] e conduzem não apenas a uma reformulação do dito fisicalismo mas também à recuperação de noções indevidamente afastadas por Dennett, nomeadamente as noções de intencionalidade intrínseca e de entendimento genuíno. A proposta afasta-se da letra estritamente considerada

[795] Não se insistirá aqui demasiado na possível diferença entre metafísica e ontologia. Uma particular perspectiva dessa diferença será retomada mais à frente através de B. Cantwell Smith. No entanto, em geral, utiliza-se o termo *metafísica* para nomear uma teoria da natureza da realidade, do mundo no sentido regulador e totalizante, e o termo *ontologia* para nomear respostas particulares à questão "o que é que há?".

[796] Chamar-se-á aqui *fisicalismo* à ideia segundo a qual aquilo que fundamentalmente existe corresponde àquilo com que a física, enquanto ciência básica, se compromete nas suas teorizações. Está inevitavelmente em causa na discussão acerca do fisicalismo a pretensão de que existiria uma correspondência entre o mundo-em-si, a "realidade realmente real", e uma teoria científica disciplinar.

da TSI: interessa esclarecer o que fica inexplícito quando a noção de intérprete é simplesmente assumida (por exemplo de onde vem a EI e porque é que ela surge em alguns sistemas cognitivos como postura face a outros sistemas).

Convem notar que embora Dennett declare que a TSI é uma teoria *fisicalista* do mental, a TSI não é uma exploração metafísica do fisicalismo como aquela que se encontra em autores como F. Jackson, D. Chalmers ou J. Kim. O fisicalismo da TSI é uma declaração de princípios e pouco mais. Aliás, a TSI pretende ser ao mesmo tempo uma teoria fisicalista e uma teoria quineana da interpretação e a verdade é que um certo dualismo persiste nas teorias quineanas da interpretação: tanto quanto estas propõem que o mental é *apenas uma interpretação* de algo que é *realmente físico*, elas retêm uma distinção por princípio entre o mental e o resto. É certo que não se trata de um dualismo imaterialista ou solipsista mas trata-se ainda assim de um dualismo, de resto perfeitamente notório nas discussões acerca de sintaxe e semântica, razões e causas, vocabulário intencional e vocabulário físico que povoam a filosofia da mente. A base para o questionamento do fisicalismo da TSI é o realismo moderado, baseado numa teoria evolucionista da cognição, em que a posição dennettiana acerca de intencionalidade se estabiliza. Este representa de facto uma subversão da posição básica e da última palavra atribuída no seio do fisicalismo à física e àquilo que de acordo com a física existe. A explicitação desta posição mostrará de que forma a TSI incorpora a irredutibilidade da subjectividade na teoria da mente, devendo o "absolutismo da terceira pessoa" (a expressão é de Chalmers) da teoria deflacionária da consciência ser secundarizado e corrigido. A objecção maior ao absolutismo da terceira pessoa parte do seguinte: se a filosofia da ciência cognitiva pretende oferecer uma teoria do lugar da mente na natureza ela será obrigada a encarar a questão da constituição da objectividade. Ora, a subjectividade em jogo na constituição da objectividade é uma questão mais abstracta e mais geral do que a questão do sujeito psicológico individual e dos mecanismos cognitivos sub-pessoais que garantem a unificação deste e não é abordável através de modelos cognitivos particulares, como os propostos na psicologia e na IA (ou no MEM). Aquilo que está em causa é o estatuto da representação e do entendimento no seio daquilo que existe e para abordar tal problema será necessário deixar para trás os problemas específicos da filosofia da mente, que é filosofia de uma ciência específi-

ca, a psicologia ou ciência cognitiva, de modo a enfrentar questões gerais relativas à subjectividade, à objectividade e à representação. Dennett trata estas questões em *Beyond Belief* através de uma noção muito fraca, uma noção precisamente psicologista e cientista[797], a noção de mundo nocional ou heterofenomenológico, identificado como o conjunto dos relatos do sujeito (do sujeito entendido no sentido psicológico, i.e. o indivíduo em situação experimental) sobre aquilo em que acredita, relatos que podem depois ser tomados e comparados por alguém (o intérprete, o cientista) com aquilo que se passa no cérebro do sujeito. Aquilo que se passa no cérebro seria então a realidade real, aquela que importa do ponto de vista ontológico, e que estaria (aparentemente) simplesmente lá. No entanto, de acordo com a própria TSI, o cérebro não se identifica com o pensamento, a representação ou o entendimento. Se há coisa que a investigação científica da cognição sugere é que o pensamento poderia não ser implementado ou realizado por cérebros e que "o representado na representação" não é uma realidade do mesmo nível que os mecanismos da cognição. A questão da representação não é, assim, idêntica à questão da constituição física e funcionamentos de sistemas cognitivos. Procurar-se-á por isso defender que não é o indivíduo, o sujeito do psicólogo o foco de uma teoria da representação mas sim a relação entre o registo[798] e o mundo. Noutras palavras, o que está em causa é o sujeito do filósofo, ou, nos termos da TSI, o intérprete.

6.2 *A natureza e o seu interior I. Será a TSI necessariamente fisicalista? O realismo moderado, a teoria evolucionista da cognição e o estatuto dos interfaces. A (in)justificação do fisicalismo*

Desde o início deste trabalho foi exposta uma teoria da mente fisicalista e anti-reducionista, nos seus aperfeiçoamentos e especificações ao longo das últimas décadas. A TSI é fisicalista na medida em que das três estratégias que prevê (a EI, a ED e a EF) a EF é

[797] Não se pretende afirmar que a noção de mundo heterofenomenológico é metodologicamente inútil quando se trata do estudo científico da cognição, apenas que a sua utilidade se restringe a um âmbito muito específico, a que se poderia chamar metodologia da psicologia.

[798] A noção será introduzida através de B. Cantwell Smith (CANTWELL SMITH 1996) e traduz uma apreensão de qualquer género, não apenas conceptual, entre duas partes no/do mundo, uma das quais toma-outra-como-sendo-x.

considerada a mais geral e com maior poder de previsão. Embora Dennett afirme por vezes que as noções de SI e de EI são neutras em relação ao fisicalismo e portanto apenas compatíveis com ele e que o poder de previsão da EF é na prática e na maioria dos casos inacessível aos humanos, de qualquer modo admite que apenas a existência e a natureza das entidades e leis da física não depende nunca de adscrição ou de interpretação, por oposição às entidades categorizadas como funcionais ou como intencionais (por exemplo entidades biológicas como espécies e entidades psicológicas como estados mentais). Apenas as teorias físicas são absolutamente gerais: elas aplicam-se a tudo aquilo que existe, enquanto que o âmbito de teorias funcionais ou intencionais é mais reduzido. A TSI é anti-reducionista na medida em que nega que o intencional (e aliás também o funcional) possa ser reduzido – no sentido que este termo tem em filosofia da ciência – a entidades situadas abaixo da descrição intencional numa concepção a que se pode chamar uma "concepção hierárquica de natureza" e das teorias da natureza.

Ora, um tal fisicalismo anti-reducionista pressupõe, para se opôr a ela e não a discutindo, uma determinada concepção (hierárquica) de natureza e por conseguinte da relação das teorias científicas entre si, de acordo com a qual existem níveis mais básicos e mais gerais (de entidades, de leis) na natureza que de algum modo sustentam níveis mais altos. Os níveis mais básicos e mais gerais são considerados fundamentais. Se o não-reducionismo configura uma posição em filosofia da ciência, trata-se de uma posição caracterizada pelo facto de constituir uma alternativa à ideia de "ciência unificada", de acordo com a qual (e em correspondência com a dita concepção hierárquica de natureza) a relação entre teorias é de uma eventual redução. A TSI sai fora desse quadro ao afirmar que as entidades e as teorias de nível mais alto não podem ser reduzidas às entidades e teorias de nível mais baixo. O anti-reducionismo opõe-se portanto à ideia de ciência unificada. No entanto, o compromisso de Dennett com aquilo a que chama "epistemologia e metafísica científicas *standard*" e que identifica globalmente com o materialismo ou fisicalismo, parece obrigá-lo de novo, ao contrário do que o professado anti-reducionismo parecia prometer, a uma defesa da concepção hierárquica de natureza e da correlativa ideia de ciência unificada.

A TSI, uma teoria interpretativista, não é obviamente a única posição anti-reducionista possível acerca da intencionalidade. De entre os vários autores analisados neste trabalho, também por

exemplo J. Fodor[799] defende uma posição anti-reducionista ao argumentar que as ciências especiais (no caso que interessa, a psicologia) não podem ser reduzidas à física. Fodor pensa no entanto, ao contrário de Dennett, que é possivel formular leis das ciências especiais (embora estas sejam *leis ceteris paribus* e não leis sem excepção ou leis estritas) e portanto pensar causalmente acerca do nível intencional da realidade. As categorias correspondentes a este nível seriam categorias de géneros naturais. Fodor defende assim aquilo a que se chamou um funcionalismo forte. Note-se que, enquanto o anti-reducionismo de Dennett se deve pelo menos à partida ao interpretativismo quineano, o anti-reducionismo de Fodor se explica pelo *realismo* intencional. Esta oposição foi analisada no Capítulo 2 do presente trabalho. Pelo facto de não ser uma posição realista quanto à intencionalidade, a posição de Dennett (a posição não reducionista *interpretativista*) tem mais problemas com o fisicalismo do que a posição de Fodor. O realismo intencional permite a Fodor defender determinadas posições definitivamente fechadas à TSI – nomeadamente que a explicação psicológica é nomológica (e que ela é portanto propriamente uma *explicação* e não apenas uma descrição), que as generalizações psicológicas exprimem relações *causais* entre estados mentais intencionalmente caracterizados cujas realizações físicas não são o mais importante, que as leis intencionais explicitam aquilo sobre o qual a psicologia de senso comum se apoia já e que a racionalidade de sistemas cognitivos físicos é possibilitada pela computação de *representações reais*. A estas ideias a TSI contrapropõe uma diferente relação entre a psicologia de senso comum, a psicologia cognitiva sub-pessoal e a própria TSI, de acordo com a qual a psicologia de senso comum consiste em *descrições* semânticas não explicativas e a questão da redução se coloca apenas entre a TSI (como descrição de competências) e a psicologia cognitiva sub-pessoal (como teoria da implementação). Quando considera a relação entre uma caracterização abstracta de sistemas intencionais e os mecanismos sub-pessoais Dennett não apenas pensa que não há razões para defender o reducionismo, como defende que as próprias categorias da psicologia cognitiva sub-pessoal dependem de interpretação, não correspondendo de forma alguma a géneros naturais.

[799] Cf. Cap. 2 deste trabalho e também o Cap. 2 de FODOR 1998 b, *Special Sciences: Still Autonomous After All These Years.*

A veia anti-intelectualista ryleana é óbvia na oposição de Dennett ao realismo intencional. Ao contrário de Fodor, Dennett pensa que representações reais e regras explícitas não são o bom modelo do comportamento inteligente: regras e representaçãoes não se encontram em sistemas independentemente da interpretação. Mais: a Hipótese da Linguagem do Pensamento é acerca de um *design* comum aos sistemas cognitivos e Dennett pensa que nada permite supor uma tal comunidade, que não há razão para supor correspondências entre atribuições intencionais e traços mecânicos sub-pessoais salientes em sistemas cognitivos. Sendo a disputa entre Dennett e Fodor uma disputa acerca do estatuto da explicação psicológica e dos objectos da psicologia científica (em última análise trata-se de saber se estes são géneros naturais ou não), de acordo com Dennett, a psicologia científica (a que Dennett chama psicologia cognitiva sub-pessoal) tem um estatuto menos glorioso do que aquele que Fodor – com as ideias de ciência especial e de leis intencionais – pretende atribuir-lhe. Para Dennett a psicologia cognitiva sub-pessoal é apenas uma teoria da implementação daquilo que (intencionalmente interpretado) serão representações e computações. Ela não tem como objecto géneros naturais. O veredicto final de Dennett acerca da Hipótese da Linguagem do Pensamento é que com ela Fodor comete o pecado ryleano de confundir uma resposta conceptual com uma resposta causal. O problema da mente é para Dennett um problema conceptual na medida em que é relativo à interpretação e não à descoberta de um nível intrinsecamente representacional da realidade. De acordo com Dennett, os princípios da psicologia fodoriana são impraticáveis na medida em que apenas a EI pode guiar a descoberta do *design*.

A TSI recua portanto perante o rebatimento da EI sobre a ED, i.e. perante o realismo intencional, perante representações reais independentes de interpretação, considerando antes que a própria ED é já interpretativa. É certo que isto acontece porque, como se viu no capítulo 2, a TSI alberga uma ambiguidade quanto ao estatuto do *design* ou das funções, inclinando-se ao mesmo tempo a considerar que as funções de dispositivos são relativas (a uma interpretação) e que elas são reais. É a ambiguidade quanto ao problema do *design* que explica que Dennett não defenda uma posição tão realista como os outros autores considerados, teleofuncionalistas ou defensores de uma teoria nómico-informacional. É também por estas razões interpretativistas quanto à psicologia de senso comum

e quanto à psicologia cognitiva sub-pessoal que Dennett defende, ao contrário de J. Fodor, que a racionalidade não tem verdadeira natureza, o que significará depois que a psicologia não tem verdadeira natureza, i.e. que ela não é a teoria de um nível autónomo de realidade. Esta afirmação é nuclear na oposição ao representacionismo ou sentencialismo. É importante ter sempre em mente que o diferendo entre a TRM e a TSI acerca da natureza das representações diz respeito ao estatuto de processos mecânicos, causais, sub-pessoais, descritos nas ciências da cognição e não a alguma coisa de que os sujeitos podem estar conscientes. A oposição de Dennett à TRM e ao realismo intencional, que permitiriam configurar a autonomia e a irredutibilidade do nível representacional da realidade, compromete-o com a ligação da mentalidade (da psicologia, da racionalidade, das representações em geral) a uma interpretação. O problema é saber se a perspectiva evolucionista sobre ontologia e metafísica correlativa deste interpretativismo é coerente com o fisicalismo.

A expressão, "perspectiva evolucionista sobre a ontologia e a metafísica" é utilizada por Dennett em *Kinds of Minds*[800] para se referir às consequências da sua abordagem da cognição. De acordo com a perspectiva evolucionista sobre a ontologia e a metafísica, a existência de "representações" em sistemas cognitivos físicos está ligada ao envolvimento desses sistemas no ambiente. É devido ao reconhecimento deste facto que o interpretativismo inicialmente instrumentalista vem a ser o realismo moderado de *Real Patterns*. De acordo com esta posição, um intérprete não cria arbitrariamente interpretações intencionais: as interpretações intencionais resultam do reconhecimento de representações reais "embebidas" (*embedded*) no ambiente. Mas será que o realismo (moderado) da representação, que reconhece a existência objectiva de padrões relacionais, de nível elevado, não entra em contradição com o fisicalismo?

Note-se que não conceder prioridade absoluta à EF seria perfeitamente coerente com o anti-reducionismo proclamado por Dennettt desde C&C, bem como com a pluralidade das estratégias previstas na TSI. É óbvio que existe uma semente "pluralista" na filo-

[800] «We tend to forget that our ways of thinking about the world are not the only ways, and in particular are not prerequisites for engaging the world successfully. It probably seems obvious, at first, that since they are so manifestly intelligent, dogs and dolphins and bats must have concepts more or less like ours, but on reflection it shouldn't seem obvious at all. Most of the questions we've raised from our evolutionary perspective about the ontology and epistemology of other creatures have not been answered, and the answers will no doubt be surprising» (DENNETT 1996: 146).

sofia de Dennett, mais precisamente na própria noção de estratégias. Se todas as estratégias (*stances*) são "estratégias" (i.e. posturas, abordagens, pontos de vista) é argumentável que Dennett deveria deixar de considerar a forma de existência ligada à EF como a forma de existência mais básica e tornar-se um pluralista, colocando em paralelo a existência das entidades que as várias estratégias constituem. Dir-se-ia que é apenas uma espécie de puritanismo cientista que impede Dennett de ser o pós-modernista pluralista que por exemplo R. Rorty nele vê. Se seguisse essa linha, Dennett seria levado a afirmar, como Ryle afirmava, que há várias formas de existir, sem que umas sejam mais básicas do que outras, deixando cair a questão metafísica do fisicalismo, uma questão (ligada a uma preocupação epistemológica, ou melhor à necessidade de fazer decorrer uma teoria acerca da natureza da realidade daquilo que se conhece) acerca daquilo que é mais básico na realidade. De facto, o fisicalismo enquanto posição metafísica, tal como é defendido por exemplo por F. Jackson, exclui o pluralismo, ao assumir como tarefa da metafísica a oposição às "grandes listas" de existentes e a procura aquilo que *fundamentalmente* existe.

Recorde-se que foi a adopção de uma postura quineana, centrada na epistemologia e apostando na continuidade entre filosofia e ciência, em desfavor de uma posição ryleana[801], que levou Dennett a optar por dar prioridade à linguagem das ciências naturais e que lançou todo o seu trajecto como filósofo da mente. Na medida em que entende a sua filosofia como sobretudo epistemológica Dennett pretende manter-se fiel a essa opção inicial. No entanto, a sugestão de um pluralismo mais assumido feita a Dennett não é novidade. R. Rorty é um dos leitores de Dennett que faz essa sugestão, sugerindo aliás ao mesmo tempo que Dennett deixe de se preocupar com a questão do realismo. Para Rorty, «é tolice perguntar se alguma coisa é real – por oposição a perguntar se é útil falar dessa coisa, se ela é espacialmente localizável, espacialmente divisível, tangível, visível, facilmente identificada, constituída por átomos, boa para comer, etc. A realidade é uma roda que não tem papel em nenhum mecanismo, uma vez adoptada a atitude ontológica natural. O mesmo digo da decisão acerca de ser ou não ser realista acerca de alguma coisa»[802]. O abandono do realismo conduziria naturalmente ao abandono da questão em torno do fisicalismo e da reali-

[801] Que desligaria, na prática, investigações conceptuais acerca do mental das ciências da cognição.
[802] RORTY 1993: 197-198.

dade dos padrões intencionais. Para Rorty, Dennett é demasiado realista porque é demasiado "cientista", i.e. preocupa-se demasiado com ciência e epistemologia. A sugestão de Rorty é rejeitada por Dennett[803] devido ao já referido compromisso com a ontologia e a metafísica científicas *standard*. Dennett não está disposto a deixar de ser aquilo que Rorty, enquanto filósofo, deixou de ser (e que considera um erro ser): um epistemólogo.

Ora, se Dennett não admite a sugestão de subverter a posição da EF no seu pensamento quando esta sugestão é feita a partir "de fora", nomeadamente por Rorty, procurar-se-á apresentar a mesma sugestão a partir de dentro, i.e. fazendo apelo a teóricos que estão tão preocupados com a epistemologia das ciências da cognição como o próprio Dennett. A defesa do realismo moderado será mais compreensível se for vista como uma alternativa a: (1) deixar pura e simplesmente cair a questão do realismo e do anti-realismo como Rorty sugere, (2) admitir que é afinal um eliminativista como os Churchland desde sempre lhe sugerem, (3) ser um realista intencional fodoriano.

Tomei como índicio para a proposta em seguida apresentada o louvor que Dennett faz de determinadas teorias da representação, que exemplificariam o realismo moderado, nomeadamente as sugeridas nas obras do psicólogo J. Gibson e dos filósofos A. Clark e R. Millikan[804]. O estatuto concedido à cognição por tais teorias subverte aquilo a que se poderia chamar a "neutralidade da representação" bem como a autonomia de um nível representacional da realidade. Uma ideia central comum a teóricos evolucionistas da cognição como J. Gibson, A. Clark ou R. Millikan é que as representações mais básicas que existem em sistemas cognitivos físicos são "orientadas-para-a-acção". Noutras palavras, as representações mais básicas são acerca daquilo que, no ambiente, importa ao sistema considerar. Para capturar essa característica os gibsonianos puseram a correr o expressivo termo *affordances*, R. Millikan fala de *pushmi-pullyu representations*[805] e A. Clark fala de *action-oriented representations*. Comum a todos este teóricos é o facto de pensarem que as

[803] DENNETT 1993.

[804] Cf. GIBSON 1979, *The Senses Considered as Perceptual Systems* (o trabalho de Gibson é a origem da chamada "psicologia ecológica"), CLARK 1997, MILLIKAN 1984 e MILLIKAN 1992.

[805] A expressão é um título de Millikan (MILLIKAN 1996). Já em MILLIKAN 1984 era desenvolvida a ideia de Ícones Intencionais com função "imperativa/informativa", i.e de "representações" ou proto-representações que não podem ser pensadas como puramente descritivas porque são também *directivas*.

representações orientadas para a acção, pelo facto de serem acerca daquilo que, no ambiente, é relevante para o sistema, não são de modo algum neutras. Se as representações orientadas para a acção são o tipo básico de representação, o mundo representado por sistemas cognitivos físicos, por exemplo sistemas biológicos, muito diferentes dos humanos, é muito diferente do mundo representado pelos humanos. Ao longo da evolução biológica dos animais, por exemplo, estes desenvolveram sistemas perceptivos prontos a responder às oportunidades que o ambiente "providencia" (a palavra *affordances* poderia ser traduzida por providenciamentos, oportunidades para o comportamento apresentadas pelo ambiente) para a satisfação das suas necessidades específicas. Estas necessidades diferem entre si, e portanto os sistemas perceptivos, e logo os animais, constituirão os seus mundos de formas muito diferentes (num mundo serão especialmente salientes flores-que-providenciam-néctar-que-deve-ser-colhido, noutro mundo a nutrição-possível-que-passa-rapidamente-pelo-campo-visual-e-deve-ser-capturada, noutro mundo a toxicidade-da-água-à-superfície-que-deve-ser-evitada, etc). Pelo que foi dito nos capítulos anteriores acerca do *design* natural e do processo de *satisficing* que preside à criação deste, é claro que nenhum sistema perceptivo biológico constitui uma solução ideal para os "problemas" colocados pelo ambiente ao organismo. O problema resolvido de forma suficientemente satisfatória pelos sistemas cognitivos não é "capturar o mundo com exactidão" mas "capturá-lo suficientemente bem de modo a que o sistema global sobreviva". Igualmente, nada existe de absoluto, do ponto de vista cognitivo, na forma como os humanos vão resolvendo o problema de capturar o mundo suficientemente bem. Alguém que aceite os pressupostos de uma teoria evolucionista da cognição não pode admitir que o "mundo em si" ou o "mundo descrito de uma forma neutra" corresponde de alguma maneira às categorias humanas comuns (i.e. à imagem manifesta de uma particular espécie animal, a espécie humana). O conceito de "mundos nocionais" traduz uma tal ideia e é esse aspecto que Dennett pretende acentuar com a perspectiva evolucionista sobre ontologia e metafísica explorada em KM. Mas se é assim, é muito contestável que se possa considerar que o mundo em si corresponde ao que a física – a imagem científica de uma particular espécie – fôr afirmando mais do que ao que outras "estratégias" forem afirmando. Consequentemente, o fisicalismo, com a sua referência constitutiva à física, à qual é conce-

dido o "lugar de honra" na hierarquia da natureza, terá que ser posto em causa. Esta constatação não tem por que conduzir ao abandono da metodologia e da convicção naturalistas e nem sequer necessariamente ao abandono do realismo científico. O que está em causa é apenas a interpretação filosófica do estatuto da cognição e o enquadramento do pluralismo dos "mundos". A perspectiva evolucionista "acerca de ontologia e de metafísica" sugere que de algum modo existem mundos representados diferentes e mesmo incompatíveis entre si não havendo razão suficiente para afirmar desde logo que uns são simplesmente mais reais do que os outros.

Começar-se-á por deixar nas mãos de John Haugeland[806] uma interpretação do realismo moderado da TSI que torna este mais coerente com as perspectivas evolucionistas acerca da cognição. A interpretação de Haugeland coloca a EI numa posição de privilégio relativamente às outras estratégias, uma posição que a EI não tem no pensamento do próprio Dennett. Uma vez que a interpretação de Haugeland vai no sentido de incorporar a TSI na sua própria metafísica da mente, convém notar o que Dennett pensa acerca desta. Ora, Dennett pensa que o facto de Haugeland acompanhar a sua defesa de que a cognição é constituição com a constante demonstração da dificuldade de constituir um mundo é suficiente para afastar qualquer relativismo (por contraste, por exemplo, com a sugestão de Rorty)[807]. Isto é assim mesmo se a explicação que Haugeland dá daquilo em que consiste a constituição de um mundo não permite afirmar que o relativismo deve ser eliminado porque existe uma maneira privilegiada de constituir o mundo[808]. De facto, Haugeland salienta a ausência de necessidade na constituição humana do mundo através da representação, acentuando no entanto a ideia de *constituição.*

6.2.1 *J.Haugeland e a teoria do entendimento (ou como reconciliar a Estratégia Intencional com a Intencionalidade Intrínseca de Searle).*

John Haugeland, discípulo de H. Dreyfus, célebre crítico das primeiras tentativas da IA, será o primeiro guia na tentativa de en-

[806] A referência será HAUGELAND 1997, *Having Thought*, obra na qual estão reunidos vários ensaios sobre Mente, Matéria, Significação e Verdade.

[807] Dennett refere-se concretamente ao trabalho prático de Haugeland com modelos de cognição, ao que se poderia acrescentar a proximidade relativamente à IA. Como afirma Dennett (DENNETT1999a: 434) «he has plenty of hands-on acquaintance with the relation between mechanisms and codes».

[808] É o comentário de Dennett em DENNETT 1999: 433.

contrar um aprofundamento metafísico e ontológico coerente para a TSI. Haugeland cita Dreyfus para explicitar a posição acerca de significado e representação que defende em *Having Thought*: «Quando estamos em casa no mundo (*at home in the world*), os objectos significativos embebidos[809] (*embedded*) nos seu contexto de referências (...) não são um modelo do mundo armazenado na nossa mente: eles são o mundo ele próprio»[810]. Esta é uma afirmação acerca do "lugar do significativo" (o significativo não é qualquer coisa de "interior"), acerca do carácter do significativo (o significativo não é primeiramente "representacional", não é um modelo do mundo na mente), e acerca da nossa posição em relação ao significativo (nós lidamos com isso). Ao desenvolver a sua tese acerca daquilo que é significativo sob a forma de uma teoria do entendimento (*understanding*) Haugeland reconcilia, *en passant*, e é esse o facto que me interessa aqui, duas das mais aparentemente irreconciliáveis posições quanto à intencionalidade da filosofia da mente contemporânea, as posições de D. Dennett e de J. Searle[811].

Como se viu no Capítulo 2, a ideia de intencionalidade intrínseca defendida por Searle, é um dos principais alvos da TSI. Ora, através da interpretação que faz do confronto entre Dennett e Searle quanto à natureza da intencionalidade, nomeadamente tomando de Dennett a noção de EI e de Searle a noção de intencionalidade intrínseca, Haugeland propõe-se ultrapassar dificuldades que quer Dennett quer Searle encontram no desenvolvimento das respectivas propostas quanto à intencionalidade. Para Haugeland, a fundamentação das teorias cognitivas requer uma teoria da *constituição* da objectividade. Haugeland enfrenta directamente o problema da normatividade na cognição e no pensamento., não apenas do ponto de vista da filosofia da mente enfrentando o problema semântico mas de um ponto de vista epistemológico e ontológico mais amplo, visando o estatuto ontológico da representação. De acordo com Haugeland, constituir é trazer a ser e simultaneamente deixar ser: as normas da razão e da verdade objectiva são em última análise as mesmas. Para explicar esse facto é necessário dispôr de uma teoria da *unidade* em causa na normatividade. Ora, essa unidade é não apenas a unidade do sujeito mas também a unidade da "objectividade" por este constituída.

[809] Adopta-se a tradução portuguesa corrente na IA.

[810] Hubert DREYFUS 1972, *What Computers Can't Do*, Capítulo *The role of the body in intelligent behavior*. Cf. comentário em HAUGELAND 1997 b: 231.

[811] Cf. HAUGELAND 1997 a, *Understanding: Dennett and Searle.*

Dennett trata a "unidade da subjectividade" de modo muito diferente na teoria do conteúdo e na teoria da consciência. Na teoria do conteúdo, a unidade é a unidade do intérprete e daquilo que este visa. Na teoria da consciência a unidade é sobretudo tratada como a realidade virtual de um Eu, uma auto-representação de unidade num sistema de agentes múltiplos. Com esta maneira dupla de tratar a unidade da subjectividade, e embora declare sempre que a teoria da consciência pressupõe a teoria do conteúdo, Dennett parece esquecer-se totalmente, na abordagem do problema da consciência por meio do MEM, da ligação entre sujeito, normatividade e objectividade que era clara na teoria do conteúdo. De facto, o MEM trata a unidade do "sujeito" como unidade *constituída*, centro virtual, esquecendo a responsabilidade pela constituição do conteúdo. É como se se tratasse de duas unidades diferentes, a unidade envolvida na intencionalidade, ligada à interpretação supondo a racionalidade a a unidade virtual de controlo num sistema de agentes múltiplos que possibilita a auto-consciência. A situação torna aliás estranha à TSI qualquer questão acerca da (auto)-consciência do intérprete, o que é no mínimo bizarro. Esta duplicidade é sintomática: é a teoria da consciência e não a teoria do conteúdo o lugar do absolutismo da terceira pessoa na teoria dennettiana do mental. No entanto de acordo com as relações de precedência na ordem da teoria que Dennett sugere existirem entre teoria do conteúdo e teoria da consciência, o absolutismo da terceira pessoa depende de uma teoria na qual o intérprete é irredutível e na qual, portanto, a terceira pessoa não é de modo algum absoluta. Noutras palavras, se é possível uma teoria do Eu feita pelo menos em parte a partir de fora não é possível fazer uma teoria do intérprete a partir de fora. O grande obstáculo na teoria da consciência parece ser aos olhos de Dennett a identificação da consciência com um ponto ou um lugar (ou com algum "um" centrado por pessoa, a não ser um "um" virtual, resultante do funcionamento descrito pelo próprio modelo). No entanto, a unidade normativa em causa no conteúdo ou intencionalidade, não tem de modo algum que ser um ponto ou um lugar ou um Teatro Cartesiano. A unidade em causa na normatividade é a unidade de um intérprete, ligada à constituição (de objectos) através da EI. Recorde-se[812] que já R. Rorty apelidava (pejorativamente) de cartesiana a forma como Dennett colocava a questão da consciência em

[812] Cf. Capítulo 1, Ponto 1.3.1.

C&C, precisamente devido ao esquecimento do aspecto normativo da intencionalidade (que Rorty por seu lado ligava à linguagem, e como tal à exterioridade, o que de resto impedia a "interiorização" funcionalista da incorrigibilidade pretendida por Dennett em C&C[813]) e à obsessão com o centro, considerado como a unidade da auto-referência. Em suma, o que me interessa é que quer Haugeland quer Rorty chamam a atenção para o facto de uma teoria cognitiva da consciência não constituir só por si uma teoria epistemológica e ontológica da subjectividade enquanto foco de normatividade e unidade. É essa a razão pela qual uma explicação científica da consciência, um modelo da consciência, não é desde logo ou só por si uma teoria do pensamento ou da subjectividade. Ora, enquanto teórico da consciência Dennett parece por vezes esquecer este facto.

Uma teoria da consciência não é por si uma teoria da subjectividade na medida em que parte do que é preciso explicar acerca da subjectividade é a responsabilidade desta na constituição da "objectividade", algo que uma teoria por exemplo neurobiológica da consciência não faz nem tem que fazer. Dennett tem a intuição correcta quando considera a teoria do conteúdo mais fundamental do que a teoria da consciência na medida em que essa ordem de precedência sublinha a unidade envolvida na noção de intencionalidade, ou conteúdo ou representação, uma unidade que é alguma forma de visar e não um ponto ou lugar num sistema físico. O interesse da interpretação que Haugeland faz da TSI reside precisamente no facto de ele tomar a EI como um primeiro passo para uma teoria da normatividade desse visar que institui unidade.

É no âmbito de tal problema que Haugeland considera que o prolongado confronto entre Searle e Dennett em torno da natureza da intencionalidade é exemplar[814]. É preciso notar que apesar de tudo Dennett e Searle estão de acordo quanto à abordagem metodológica apropriada para a teoria da mente: ambos se declaram naturalistas e anti-reducionistas. A discordância existente entre os dois é recondutível ao ponto em que se trata de decidir exactamen-

[813] E na medida em que a incorrigibilidade explicava a consciência, uma tal proposta pura e simplesmente impediria a consciência-segundo-Dennett...A incorrigibilidade deveria antes ser explicada segundo Rorty (cf. RORTY 1972a e RORTY 1972 b) por uma teoria social da significação.

[814] Se relativamente ao problema da intencionalidade Fodor e Dennett exemplificam o choque entre o realismo intencional (no sentido de autonomia de um nível autónomo de realidade intrinsecamente representacional) e a recusa desse realismo intencional, quando se trata de discutir a pertinência da noção de intencionalidade intrínseca por oposição a intencionalidade atribuída Searle e Dennett exemplificam o choque de intuições.

te que sistemas são intencionais. Para Searle, a intencionalidade é uma característica *intrínseca* de sistemas, que envolve a consciência. Nenhum sistema incapaz de consciência poderia alguma vez ser (genuinamente) intencional. Para Dennett, a intencionalidade é ao mesmo tempo *relativa a um intérprete e objectiva* (i.e. pode ser imputada, atribuída, reconhecida a partir de fora: os padrões da mentalidade sobressaem ao olhar de um intérprete sem que por isso sejam arbitrários ou ficcionais – ao contrário por exemplo da narrativa do Eu, uma questão muito mais "superficial"). Cada uma destas posições deixa alguma coisa por explicar. No caso de Searle, falta saber em que consistiria a intencionalidade intrínseca por oposição à "intencionalidade" que não é intrínseca mas apenas atribuída. No caso de Dennett fica pelo menos um intérprete como resto.

Haugeland pensa que a descrição que Searle faz da natureza da intencionalidade[815] é mais acertada, na medida em que relaciona a intencionalidade intrínseca com a subjectividade (e para Haugeland a subjectividade que importa, nomeadamente numa teoria do pensamento, é a intencionalidade intrínseca e não a intencionalidade atribuída). Existe uma diferença importante entre cada um de nós, seres conscientes capazes de ler estas linhas, e determinados SI previstos com sucesso através da EI, como um computador pessoal ou o termostato de *True Believers*. Haugeland crê no entanto que a noção searleana de intencionalidade ganha ao ser esclarecida através da noção dennetiana de *stance*. Quer Dennett quer Searle admitem que a intencionalidade é normativa, i.e. que a *aboutness* da mentalidade é um ser-suposto-ser. Para Dennett isso significa, como se sabe, que a mentalidade está ligada a um intérprete e que a racionalidade é um critério da interpretação do observador, que aborda um SI submetendo-o a um standard normativo. Para Searle, a normatividade significa que existem condições para que um estado seja intencional. Esta normatividade é aliás capturada por Searle com o conceito de *forma aspectual*: o mental tem forma aspectual, i.e. toma aquilo que visa sob determinados aspectos e não sob outros (é por isso, de resto, que é possível distinguir como mental algo de actualmente inconsciente, por oposição a algo que também é inconsciente mas que não é mental, mesmo que seja por exemplo cerebral, como a mielinização dos axónios). Searle não tem no entanto como explicar a forma aspectual da intencionalidade. Ora a TSI tem indu-

[815] Cf. SEARLE 1992.

bitavelmente recursos para isso, ela pode explicar porque é que uma "história causal" acerca de estados físicos de sistemas, admitida que seja toda a sua relevância, não justifica a intencionalidade, não justifica a razão por que um particular estádio da história causal numa região de espaço-tempo, a região que engloba os particulares relacionados como representando e representado, é mais importante do que outros estádios e do que tudo aquilo que fisicamente medeia essa "relação", ou, para colocar a questão como um psicólogo gibsoniano, a história causal não explica porque é que um sistema cognitivo, por exemplo um animal, vê uma árvore e não tudo o que fisicamente medeia o ver de uma árvore e o ser-visto de um particular no mundo por outro particular no mundo.

Para explicar como é possível que configurações de um cérebro físico resultante da evolução sejam "intrínsecamente normativas" Haugeland retoma o exemplo do jogo de xadrez. Como é que um jogador de xadrez vê, por exemplo, num jogo de xadrez, que "um cavalo" "ameaça" "um peão"? De acordo com Haugeland, não é uma história causal que explica essa "percepção" mas sim um compromisso constitutivo, do qual depende a própria realidade em causa. Fenómenos-de-xadrez não são identificáveis com formas ou cores de peças (as peças de xadrez nem sequer têm que ser de madeira ou de marfim, podem ser por exemplo peças num écran de computador ou letras dispostas numa folha branca). Fenómenos-de-xadrez são constituídos por standards e o "jogador" (o sujeito, o intérprete) não é um factor indiferente no que respeita ao facto de os standards estarem ou não a ser cumpridos. É neste ponto que reside, segundo Haugeland, a importância da EI: o papel constitutivo-de-realidade das regras do xadrez relativamente aos fenómenos-de-xadrez é análogo ao papel constitutivo (de mentalidade) que Dennett atribui à interpretação-supondo-a-racionalidade. O facto de existir um compromisso constitutivo significa que: (1) se alguma coisa não estiver de acordo com os standards, dever-se-á verificar se existe um erro, (2) se a anomalia persistir e se não forem encontrados erros, ou se modifica os standards ou se desiste deles (desce-se por exemplo à ED ou à EF, na terminologia de Dennett). Mas modificar standards ou desistir deles não é uma decisão indiferente: trata-se de modificar ou fazer desaparecer mundos e entidades nesses mundos. A constituição de um domínio de fenómenos de acordo com standards não é um resultado trivial. Afirmar que a objectividade é constituída, é afirmar que ela é resultado do visar a

partir de uma (nos termos de Dennett) estratégia (*stance*), algo bem patente no caso do jogo de xadrez: as questões de correcção põem-se quanto a saber que peças são "cavalos", e não quanto ao embate de fotões na retina do observador visual do sistema.

A ideia de constituição por meio da qual Haugeland interpreta a TSI é evidentemente kantiana e Haugeland considera que ela esclarece a noção dennetiana de estratégia. O ponto desta é a importância da constituição na cognição, o facto de objectos de percepção, acção ou pensamento apenas serem inteligíveis como os objectos mentais que são nos termos de algum compromisso prévio (um compromisso de alguma subjectividade-*qua*-unidade, i.e. do intérprete). São precisamente objectos constituídos pelo compromisso da EI que têm a "forma aspectual"[816] searliana, que os distingue como intencionais.

Por sua vez, o problema da posição de Searle quanto à intencionalidade intrínseca (defendida emblematicamente com o Quarto Chinês) é o facto de ela não explicar como é possível que algo (o intencional) que por outro lado se distingue, se caracteriza, pela forma *aspectual* seja *intrínseco*. Searle propõe uma caracterização de intencionalidade genuína na qual enumera tudo aquilo que a intencionalidade não é: ela não é derivada, atribuída, etc. No entanto não explica como é que a intencionalidade intrínseca – aquela que Searle manipulando símbolos incompreensíveis dentro do quarto teria, ao contrário do próprio Quarto Chinês globalmente considerado – é possível. Ora, Haugeland defende a intencionalidade intrínseca de alguns sistemas, aquela que o Quarto Chinês precisamente põe em relevo, consiste no acima descrito compromisso com normas. Usando da conveniente justiça interpretativa para com Searle, é preciso notar

[816] De acordo com Searle (SEARLE 1992, Capítulo VII, *The Unconscious and Its Relation to Consciousness*), a *forma aspectual* é a pedra de toque que distingue o mental do não mental, mesmo se o mental é actualmente inconsciente. A forma aspectual dos estados intencionais, aquilo pelo qual uma crença minha na qual não penso no instante t e os axónios dos neurónios do meu cérebro diferem, consiste no facto de os estados intencionais apenas representarem as suas "condições de satisfação" sob determinados aspectos (e não sob outros), aspectos estes que devem ser importantes para o agente (logo, a forma aspectual não poderá ser caracterizada em terceira pessoa – o que é exactamente o que a componente interpretativista da ideia de mundo nocional determina na TSI). Para Searle, as representações mentais sintácticas de cognitivistas como Fodor não têm forma aspectual e por isso não devem ser consideradas como mentais. É esta convicção que justifica a ideia de Searle quanto ao que está em causa na filosofia da mente: está em causa a consciência, o cérebro e nada mais, i.e. não existe um nível cognitivo intermédio, sintacticamente estruturado e real entre consciência e cérebro. Tudo aquilo que é mental está ligado actualmente ou possivelmente com a consciência (este é o princípio a que Searle chama *princípio da conexão*).

que ele não afirma nunca que a intencionalidade de uma particular crença é intrínseca: Searle não fala jamais de *intencionalidade intrínseca* relativamente a estados mentais individuais, mas apenas relativamente a *sistemas cognitivos globais*. Afirmar que um estado mental isolado teria intencionalidade intrínseca seria incompatível com outro tema caro a Searle, central na sua teoria da mente, o holismo da intencionalidade e a dependência desta em relação ao fundo (*background*). Se Searle apenas fala de intencionalidade intrínseca relativamente ao sistema, o que intencionalidade *intrínseca* significa é intencionalidade-intrínseca-de-um-particular-sistema, i.e. independência da intencionalidade desse sistema relativamente à intencionalidade de qualquer outro sistema. Ora é precisamente aqui que Searle tem mais razão do que Dennett. A noção de estratégia (*stance*) é boa tanto quanto evidencia a importância da unidade-que-unifica, e portanto de um compromisso, na cognição, ela não é boa na medida em que implica a dependência da intencionalidade face a uma estratégia que é primeiramente uma estratégia de um intérprete face a outro sistema. De acordo com Haugeland, a intencionalidade (genuína) só pode ser entendida a partir da estratégia adoptada pelo próprio sistema. A estratégia poderá ser uma qualquer (seja, nos termos da TSI, a EI, a ED ou a EF). O que importa é que é o compromisso com normas constitutivas de objectividade (a *stance*) que justifica a existência de intencionalidade intrínseca num sistema. É precisamente a unidade desse compromisso, a unidade que está em causa na objectividade, que é na TSI "o intérprete"[817]. Uma tal unidade não é dada desde sempre ou desde logo, i.e. fora do envolvimento na representação: a unidade de um sistema-sujeito é concebível apenas em virtude de normatividade constitutiva de objectividade. Nas palavras de Haugeland, «A unidade do sistema é a unidade de um único compromisso consistente, nos termos do qual uma pluralidade de estados intencionais pode ser normativamente considerada em relação com as suas condições de satisfação constituídas, e que é base da necessária subjectividade dos estados intencionais»[818]. Haugeland de-

[817] Saber exactamente de que unidade se está a falar quando se fala de *um* sistema cognitivo é, note-se, desde logo problemático. Searle, por exemplo, quando fala de intencionalidade intrínseca de um sistema está frequentemente a referir-se ao cérebro. Mas neste trabalho já foram analisados vários e diferentes sentidos da unidade, nomeadamente quando se tratou as questões da intencionalidade, do eu, da identidade pessoal e da auto-consciência. Há no entanto um sentido em que a própria individuação e distinção entre sistemas cognitivos vários se relaciona com a unidade da/na representação de que Haugeland fala.

[818] HAUGELAND 1997 a: 299.

fende portanto que é o compromisso de cada sistema de tomar os seus próprios estados intencionais correctamente que faz da intencionalidade desses estados a intencionalidade do sistema, i.e. a *intencionalidade intrínseca* do sistema. A mentalidade de percepções, pensamentos e acções requer aquilo a que Haugeland chama um "eu sempre em guarda", mesmo que não explicitamente, i.e. não necessariamente de forma reflexiva. Uma *stance*, e portanto uma normatividade, é o fundamento da objectividade e da subjectividade. Sem a consideração deste aspecto, é impossível compreender a intencionalidade. Esta interpretação da TSI retoma afinal a ideia segundo a qual o mental consiste em padrões percebidos por uma mente, aplicando-a à mentalidade do próprio sistema.

É esta, em suma, a sugestão aparentemente bizarra que Haugeland faz: juntar à intencionalidade intrínseca de Searle a noção dennetiana de *stance* de modo a explicar aquilo que Searle só por si não pode explicar e apenas afirma, i.e. que existe intencionalidade intrínseca em certos sistemas e não noutros e que ela é diferente de uma forma importante da intencionalidade atribuída. Searle tem razão contra o liberalismo excessivo da TSI, que se traduz na total indistinção entre intencionalidade genuína e intencionalidade atribuída, ao defender que podem existir sistemas aparentemente intencionais, previsíveis como intencionais e racionais, mas que não têm entendimento nem subjectividade. Tem também razão na simplicidade com que afirma em relação a certos produtos da IA que é importante poder defender que esses sistemas não possuem entendimento genuíno. Falta-lhe no entanto notar que isso é assim apenas porque sem o compromisso da estratégia pelo (ou no) próprio sistema, o qual estabelece uma unidade, não existe entendimento num sistema

Não esquecendo a variedade de mentes consideradas no Capítulo 5, é preciso notar que Haugeland não defende, com esta proposta, que o entendimento (ou intencionalidade intrínseca) não humano é impossível, apenas que pelo menos por enquanto subsiste uma diferença importante entre o uso de "representações" em sistemas artificiais, até mesmo em sistemas sofisticados como por exemplo robôs (cuja relação com essas "representações" é aliás, Haugeland admite-o, muito diferente da relação de uma folha de papel com a escrita nela inscrita), e o pensamento humano. No caso do robô, ao contrário do que se passa com os humanos, os standards são impostos a partir de fora, não existindo compromissos

constitutivos no/do próprio sistema (não existe *do-it-yourself understanding,* nos termos de Dretske). Nada exclui no entanto essa possibilidade para sistemas "artificiais". É no entanto óbvio que também os animais de algum modo não têm compromissos constitutivos semelhantes aos dos humanos. É por essa razão que segundo Haugeland (e de uma forma que indignaria muitos dos etologistas cognitivos referidos no Capítulo 5) os animais não têm senão intencionalidade-*ersatz*. Haugeland pensa aliás que toda a mentalidade animal (memórias, percepções, desejos, etc, que os animais possam ter) embora exista é *ersatz*, um sucedâneo. Para Haugeland a mentalidade animal é análoga à teleologia biológica: só aparece em função de standards humanos. Esta mentalidade-*ersatz* dos animais representa uma classificação intermédia entre as categorias searleanas de intencionalidade genuína e intencionalidade atribuída, e, ao aplicar-se a SI naturais, mostra que aquilo que Haugeland não está simplesmente a defender uma distinção apriorista entre natural e artificial. O que há aqui é uma defesa fundamentada da intencionalidade genuína de certos sistemas e a partir dos pressupostos da própria TSI. Se a distinção entre intencionalidade intrínseca e intencionalidade atribuída fosse totalmente abolida, tal como Dennett pretende, tudo no mundo seria susceptível de ter intencionalidade. Aquilo que Haugeland faz é mostrar que Dennett não deveria ser tão taxativo na eliminação da distinção entre intencionalidade intrínseca e atribuída. Mesmo se à partida a intencionalidade intrínseca não tem que coincidir com a consciência, ela é pelo menos uma explicitação possível da natureza de um intérprete e esta é, sem qualquer dúvida, uma noção sem a qual a TSI não se sustenta.

A interpretação que Haugeland faz da TSI e que nivela as estratégias, subvertendo a posição básica da EF, não é, como se disse atrás, a posição do próprio Dennett. Na TSI, de acordo com as formulações de Dennett, a EI é apenas uma particular estratégia entre outras. Mas, precisamente, isso só acontece devido a uma certa cegueira da TSI perante a necessidade de explicar o intérprete que assegura as estratégias. É para oferecer uma tal explicação que Haugeland estende a noção de estratégia a qualquer estratégia de pensamento, obtendo assim os princípios de uma teoria do entendimento e da objectividade. A sugestão de Haugeland, de tomar todas as estratégias como estratégias é, como já se fez notar, semelhante à sugestão de Rorty. As duas sugestões enfrentam um mesmo problema que deve ser colocado à TSI. O problema é o se-

guinte: que razões tem Dennett para considerar a EF como fundamental, quase como uma não-estratégia, fazendo assim como se a física, enquanto ciência, não fosse intencional e como se os existentes com os quais a teorização física se compromete constituíssem a única realidade real, independente de ser-pensada? A ideia comum a Rorty e a Haugeland é que Dennett não tem nada a não ser um preconceito. Haugeland vai mais longe e sugere que primeira no sentido próprio (no sentido de *constitutiva*, relativamente a qualquer nível e forma de realidade, e relativamente à mentalidade do próprio sistema) é a EI.

6. 2. 2 *Pós-antropologismo: B. Cantwell Smith e a origem dos objectos.*

Uma crítica possível à teoria do entendimento apresentada por Haugeland é o facto de ela ser "antropologista", i.e. o facto de considerar o entendimento ligado à intencionalidade genuína como um apanágio (se não exclusivamente e definitivamente pelo menos centralmente e por enquanto) humano, distinguindo a mentalidade humana por entre a mentalidade possível ao considerar explicitamente que o compromisso constituinte é marca do humano e que a mentalidade de animais e de sistemas de IA (pelo menos até ao momento presente) não é intrínseca. Contraria-se assim a posição ecuménica característica da TSI. Mas esse ecumenismo, ou pelos menos a consideração horizontal das mentes de que se tratou no Capítulo 5, não é um exclusivo da TSI. Mesmo filósofos que são apologistas da intencionalidade intrínseca como J. Fodor e F. Dretske não restringem esta aos humanos.

A teoria apresentada pelo cientista de computadores e metafísico B. Cantwell Smith na sua obra *The Origin of Objects* permitir-me-á investigar a possibilidade de evitar o antropologismo de Haugeland continuando a explorar dúvidas acerca da pertinência do fisicalismo da TSI, dada a "perspectiva evolucionista acerca de ontologia e metafísica" e o pluralismo a que esta naturalmente conduz. A teoria ontológica da cognição que B. Cantwell Smith apresenta em *The Origin of Objects* é anti-fisicalista, no sentido em que se opõe à ideia segundo a qual o mundo-em-si corresponde ao modo como a física, a ciência básica, o teoriza. Nos termos de Dennett, Cantwell Smith não concede uma posição privilegiada à EF numa "teoria dos objectos". Ele pretende fazer justiça ao pluralismo "a partir de

dentro" construindo uma teoria não-reducionista mas fundacional da origem dos objectos, pela qual pretende recuperar noções como objecto, objectivo, verdadeiro, formal, lógico, físico, etc, num quadro que não o da concepção hierárquica de natureza. Em *The Origin of Objects* Cantwell Smith procura precisamente estabelecer os princípios daquilo a que chama uma metafísica sucessora da tradição moderna da hierarquia da natureza. De acordo com esta ideia, tudo aquilo que existe seria constituído a partir dos ingredientes postos pela física, ao que se acrescenta, para completar a ontologia, a lógica matemática e a teoria dos conjuntos. Ora, segundo Cantwell Smith, de acordo com essa perspectiva nem sequer se poderia falar de *objectos,* com a individuação e a discretização implicadas no termo.

A importância da questão da origem dos objectos na teoria da cognição é a seguinte. Nenhuma teoria naturalista da intencionalidade (incluindo, de acordo com Cantwell Smith teorias da computação, da semântica, ontologias, etc) pode presumir ou pressupor a identidade ou existência de um único objecto intencional que seja. Objectos simplesmente não existem a não ser em determinadas circunstâncias e para determinadas entidades. Não há indivíduos para a física, apenas particularidade e localidade. Ser um objecto supõe sustentar-se de algum modo como unidade, o que supõe reunião e transporte ao longo do tempo, e portanto *história*, dimensão esta que não é introduzida pela física. Este facto permite prever desde logo que a noção de objecto não será causal. Esta é, nos termos de Dennett, a questão da EI, a questão do intérprete, afinal, que origina ou constitui, através de uma estratégia, a realidade intencional, aqui chamada a realidade dos objectos, bem como o problema da *explicação* psicológica. Cantwell Smith chama-lhe a questão do registo (que poderá ser representação), noção cuja metafísica procura elaborar em *The Origin of Objects*. Afirmar que objectos não estão aí por si, que a sua realidade é intermédia e não "fundamental" é, antes de mais, constatar que objectos não são de modo algum constituintes últimos numa hierarquia da natureza. A realidade dos objectos é uma realidade intermédia entre conexão e desconexão: existe, na origem dos objectos, uma forma de separação intransponível, uma separação que subjaz à abstracção.

A estratégia de Cantwell Smith na sua investigação acerca da origem dos objectos é aceitar provisoriamente a física como caracterização básica do mundo natural e perguntar depois que mais é

necessário para que existam intencionalidade, semântica e ontologia, e portanto objectos. A questão da origem dos objectos também poderia ser formulada perguntando como podem surgir indivíduos num substracto de particularidade não delimitada. De qualquer modo, a necessidade de formular a questão pressupõe que não será possível, de acordo com Cantwell Smith, falar de representações e computações sem explicitar uma teoria ontológica, uma teoria dos objectos. I.e., não é possível, apesar da natural tendência dos teóricos da cognição, tomar simplesmente as noções de representação e computação como se elas não fossem problemáticas do ponto de vista naturalista.

Especificamente, e de forma central para a tese anti-fisicalista acerca da ontologia da cognição que defenderá, Cantwell Smith nota, como se disse, que a física não afirma o que quer que seja acerca da individualidade de objectos, embora sustente sem dúvida noções de particularidade e localidade. As questões da individualidade e da individuação não lhe pertencem[819], elas são do âmbito da teoria da representação. Este facto não torna as questões da individualidade e da individuação menos importantes na teoria da cognição, pelo contrário mostra que elas são questões especificamente cognitivas. O que acontece é que o facto de as questões da individualidade e da individuação não pertencerem à física abala o fisicalismo da teoria da cognição. Cantwell Smith considera também, por esta razão (contra Fodor, por exemplo) que uma teoria de objectos intencionais, de representações, nunca poderá ser inteiramente causal[820]. Como se disse, o mero facto de "objecto" ser uma noção histórica faz com que ela não possa ser uma noção físicamente efectiva: ser-um-objecto não é uma propriedade local de uma região espaço-temporal.

Cantwell Smith chama à sua teoria metafísica "realismo simétrico". De acordo com o realismo simétrico o mundo é Um, e esta suposição é propriamente *metafísica*. Embora a *Oneness* de particulares seja uma característica de indivíduos e não do mundo, afirmar que o mundo é Um é afirmar que este inclui sujeito e objecto e que é ao nível dos interfaces sujeito/objecto no seio de um mundo que é metafisicamente Um que se definirão *ontologias*, mundos de objectos. Ao nível mais básico, aquilo que existe é particularidade inefá-

[819] Cf. CANTWELL SMITH 1996, onde Cantwell Smith lança o seguinte repto: retire-se os animais, os artefactos, e talvez as árvores e tente-se individuar aquilo que fica.

[820] Cantwell Smith defende mais especificamente que nenhuma *teoria* poderá ser inteiramente *causal*.

vel, ou, nos termos de Cantwell Smith, um "fluxo deíctico que tudo envolve inclusive o registo". O registo é aliás definível como uma forma de interacção subsumindo *representação* e *ontologia*, «uma forma de interacção (...) pela qual "regiões-s" ou sujeitos estabilizam amostras (*patches*) do fluxo, em parte através de processos de intervenção (...) e também através de padrões de desconexão e de coordenação à distância, necessários para tomar a amostra como um objecto, ou, mais em geral, como alguma coisa no e do mundo»[821]. Embora os padrões de conexão parcial e de desconexão parcial sejam em detalhe tão variados que desafiam a descrição, isso não é importante porque eles são precisamente regularidades. Os padrões do realismo moderado da TSI, noção em que desagua a teoria da intencionalidade, são nos termos de Cantwell Smith regularidades no registo. Do mesmo modo que Dennett, Cantwell Smith considera essas estabilidades ou padrões como (1) não sendo em primeira instância clara e distintamente definidas, (2) disrompendo-se em estrutura fina. Este é exactamente o estatuto dos padrões reais da TSI, por oposição às representações sentenciais da TRM.

Na definição de registo afirmou-se que este subsumia "representação" e "ontologia". Os termos são tomados do seguinte modo por Cantwell Smith: «a ontologia [*i.e. aquilo que há considerado em termos de objectos*] é a projecção do registo no mundo. A representação é a projecção do registo no sujeito»[822]. Sendo projecções, nenhuma delas pode em última análise manter-se, o que significa que a representação não poderá ser uma categoria explicativa última (não poderá por exemplo ser um título geral para a intencionalidade, como a TRM de Fodor pretende), desligada de compromissos constitutivos de ontologia de algum sistema no mundo. É impossível separar representação e ontologia, elas são as duas faces do registo, que é mais primitivo. Alguma forma de gradualismo deverá portanto ser aperfeiçoada, sob a forma de uma teoria da independência parcial conseguida entre partes do mundo que representam e os seus representados.

A noção de registo mantém uma importante característica que os filósofos da psicologia – Cantwell Smith pensa precisamente em Dennett e Fodor – atribuíram à percepção: o registo vai ao mundo e é "por defeito" (*by default*) verídico. A noção de registo tem portanto o sentido de alinhamento com a situação exterior, ou mesmo

[821] CANTWELL SMITH 1996: 347.

[822] CANTWELL SMITH 1996: 349.

de calibração activa (Cantwell Smith virá a afirmar que o registo é assimétrico e normativo). Registar é registar-como (noutras palavras, reencontra-se aqui a forma aspectual de J. Searle, embora talvez excedendo o âmbito que o próprio Searle lhe atribuiria). No entanto, o registo não extrai desde logo "objectos". A noção de registo atenua o excesso envolvido na menção a objectos, no que esta supõe já de individuação, desde o início numa teoria da representação. Nomeadamente, processos naturais comuns como aquele a que Cantwell Smith chama seguir-efectivo (*effective tracking*)[823] não são ainda representação: nem toda a coordenação entre particulares no mundo supõe individuação abstracta. Seguir alguma coisa é o comportamento percursor da semântica mas não é desde logo representação num sentido mais específico, sentido esse que envolverá aquilo a que Cantwell Smith chama "responsabilidade da região-s". Cantwell Smith retoma o exemplo já aqui várias vezes discutido da rã-que-segue-a-mosca defendendo que aí não existe ainda representação mas pura conexão, descritível em termos físicos. Trata-se de um acoplamento, sem qualquer distinção clara, ainda, entre sujeito e objecto. Este acoplamento é comparado por Cantwell Smith com a relação existente entre uma pessoa e uma peça de roupa que esta tenha vestida: neste caso não se pensaria em dizer que a peça de roupa detecta e segue a pessoa, apesar da coordenação com e do acompanhamento do movimento da pessoa. Retomando o muito discutido caso da rã e da mosca, Cantwell Smith considera que ontologicamente o proto-sujeito e o proto-objecto são aí ainda insuficientemente distintos para que se possa falar de representação, e que fisicamente são ainda insuficientemente separados. Na situação referida existem alterações físicas numa região de espaço-tempo que envolve aquilo que está a ser nomeado como constituindo dois intervenientes, um proto-sujeito, um proto-objecto. Como diz Cantwell Smith «tendemos a ver o ar que medeia como conceptualmente além de opticamente transparente (...) mas a transparência do ar não é evidentemente ainda um facto para a emergente cena participatória. De facto, os padrões de radiação electromagnética no ar são tão (...) reais como os sinais eléctricos saindo da retina – e ainda não obviamente distintos deles (...). De facto (...) saber como é possível que nós e talvez as rãs veja-

[823] Como por exemplo (um exemplo já várias vezes aqui utilizado) o seguir-da-mosca-que-voa pela rã. Para Cantwell Smith não se trata ainda propriamente de registo na medida em que não está presente a desconexão criterial.

mos moscas é o núcleo do problema do registo. Por isso mesmo é vital não colocar aquilo que é conseguido pelo registo na nossa concepção da situação inicial»[824].

O registo supõe o envolvimento de um sistema no mundo e é, para Cantwell Smith, o acto intencional original, o processo pelo qual o mundo se apresenta imediatamente, um processo que depende da presença localizada de região-s e região-o (é aliás a partir desta presença indissolúvel que esta metafísica é baptizada como "filosofia da presença"). É muito importante para Cantwell Smith, como se viu pelo exemplo anterior, estabelecer que uma tal apresentação não é necessariamente conceptual e que ela acontece em vários tipos de sofisticação, não apenas o humano. A representação propriamente dita será apresentação de novo, em contexto de uma muito maior *desconexão* do que aquela que existe nos dois exemplos acima referidos e no contexto daquilo a que Cantwell Smith chama "a responsabilidade da região-s pela estabilização de uma região-o". Apenas nessa situação poderão vir a existir indivíduos. Na natureza de indivíduos está assim envolvido algo que falta aos fenómenos físicos: abstracção. Na origem da sua possibilidade está a assimetria entre região-s e região-o. Esse é o caso mesmo ao nível da percepção visual de objectos: de acordo com Cantwell Smith, as propriedades normalmente associadas à individuação visual de objectos, tais como limites ou bordos, são propriedades da interacção sujeito/objecto. Por esta razão que Cantwell Smith defende que não há indivíduos puramente físicos, i.e. que a individualidade não é feita apenas de "material" físico (*physical stuff*), já que supõe a responsabilidade pela estabilização e a capacidade de manter uma relação com a região estabilizada mesmo se o acoplamento for quebrado. O facto de não ser usualmente considerada, no estudo da cognição, a natureza primitiva de *registo,* realidade de dupla face constituída por representação e ontologia conduz ao erro que, de acordo com Cantwell Smith, mais frequentemente impede a compreensão da natureza da cognição: *o erro de inscrição.* O erro de inscrição é a tendência que o teórico da cognição tem para projectar suposições ontológicas num sistema, "lendo-as" em seguida de volta como se isso constituísse uma descoberta empírica independente ou um resultado. O realismo intencional seria um exemplo, e a TSI assume-se desde sempre como crítica de tal erro.

[824] CANTWELL SMITH 1996: 217.

Em suma, o caso de Cantwell Smith contra o fisicalismo prende-se com o facto de a física não ter nada a dizer directamente sobre as questões da objectividade, universalidade ou verdade. Cantwell Smith considera este problema através de uma esquematização das relações entre aqueles a que chama os 3 reinos (uma separação provisória e em última análise insustentável, é claro, dado o monismo metafísico subjacente), sendo o primeiro reino o âmbito do *particular* (o âmbito de forças, campos, posição espaço-temporal – nele não existem indivíduos), o segundo reino o âmbito do *material* (o âmbito dos físicos, dos documentos, das experiências) e o terceiro reino o âmbito do *Universal*, (o âmbito dos tipos, dos conjuntos e das leis da matemática, onde existe individuação abstracta). Uma maneira de formular o problema da origem dos objectos é notar que as relações entre os sujeitos (do segundo reino) e o *plenum* físico (do primeiro reino) simplesmente não são assunto da física.

Como se vê, a distinção entre *particularidade* e *indivíduo* é essencial a uma teoria da cognição entendida como teoria da origem dos objectos. É importante porque é essencial separar o sentido de *local* associado à ideia de particularidade do sentido de *discreto*, ou partilhável em unidades ou todos, associado à ideia de individualidade. É importante que a teoria da cognição comece por constatar que particularidade e individualidade são noções com fontes muito diferentes. A particularidade será o pedaço localizado do fluxo, ocorrente, local, que existe aí, em especificidade inexorável. A individualidade de um indivíduo é o que quer que seja acerca do ocorrente que sustenta critérios de individuação, fazendo os objectos discretos, unidades coerentes separadas do fundo. É o que nos permite dizer de um objecto que ele é um, ou de dois objectos que são dois. Apenas indivíduos possuem a propriedade da *oneness* vulgar que "falta" ao mundo como um todo. Ora, isso mostra que um essencial acto de abstracção subjaz a todo o indivíduo, a todo o objecto. Essa é de resto a razão pela qual existe realmente, como Dennett pretende, um elemento de interpretação na caracterização de qualquer mundo nocional, que não é como se viu no Capítulo 2, nem interior nem exterior, tem precisamente uma realidade intermédia, de interface.

Retomando a ideia de Millikan acerca de identidade, segundo a qual o que importa na identidade biopsiocologicamente importante é ser capaz de identificar *more mama, more milk, more mouse...*, admite-se que é a capacidade de reidentificar o mesmo como mesmo, e não

uma suposta identidade prévia, pura, o que importa na identidade, de acordo com Cantwell Smith. Esta identidade é *identificabilidade*, e identificabilidade é identificabilidade-*por* (um sujeito de cognição). As formulações de acordo com as quais a identidade é uma propriedade de um indivíduo (a propriedade possuída pelo indivíduo de ser esse indivíduo, a única propriedade que se afirma desse objecto e não de outro, uma relação binária de uma coisa consigo própria) são, para Cantwell Smith, pós-ontológicas, *after the fact*, pressupõem uma distinção entre indivíduos e propriedades, que não se sustenta sem referência a indivíduos, que não é simplesmente dada. Indivíduos ou objectos têm que ser *originados* e a história desse originar é uma história acerca de *abstracção*, num contexto de *conexão/desconexão* entre *regiões locais* no/do mundo. Se há uma história a contar é uma história acerca da origem de objectos. A identidade como identicalidade, a identidade puramente extensional, não é de modo algum primitiva.

Cantwell Smith considera que o registo é assimétrico e normativo e que a origem da assimetria é precisamente a responsabilidade de manter a coordenação, a qual cabe à região-s. Neste contexto de assimetria, nenhum objecto para um sujeito dado estará totalmente lá, no sentido de ser perfeitamente acessível de forma efectiva: a presença de um objecto envolve de forma constitutiva a sua ausência. Mais do que a não localidade, é o facto de a noção de objecto ser, repita-se, uma noção histórica que mostra que ela não pode ser uma noção de algo fisicamente efectivo. Estes são para Cantwell Smith princípios ontológicos gerais, princípios da constituição de mundos. Se a origem dos objectos acontece nas condições descritas, os mundos com os quais as mentes estão relacionadas não podem ser pura e simplesmente identificados ou directamente relacionados com o mundo-tal-como-vai-sendo-descrito pela física: noções como "objecto" e "propriedade" são inseparáveis do registo, uma noção como "verdade" é inseparável da assimetria e da normatividade do registo e só pode ser pensada ser pensada em função da natureza participatória e de interface deste. Todas as ontologias serão assim reconduzidas a práticas de registo das quais objectos, propriedades e verdade surgirão. Ser um objecto, ser uma propriedade, ser verdadeiro são questões concebíveis apenas no contexto do registo de alguma parte do mundo por alguma outra parte do mundo envolvida em cognição.

Afirmei que a teoria da origem dos objectos de Cantwell Smith permitiria evitar o antropologismo da teoria do entendimento de

Haugeland. Penso que isso é conseguido distinguindo o *monismo metafísico*, a ideia segundo a qual o mundo (num sentido regulador, totalizante) é Um, de um *pluralismo ontológico*, a ideia segundo a qual os mundos constituídos por objectos (que nunca são inteiramente físicos nem inteiramente locais nem inteiramente presentes) são vários. Ao contrário da imagem associada à concepção hierárquica de natureza, na metafísica do registo apresentada por Cantwell Smith «começa-se com o mundo como um todo (...) que é (...) dividido em objectos (...). As partes são (parcialmente) (...) extraídas do todo; não é o todo que é constituído a partir das partes»[825]; «O registo é inexoravelmente plural (...) embora não discreto, (...) o fluxo do qual ele surge é inexoravelmente singular»[826].

Procurei mostrar, apoiando-me em Haugeland e Cantwell Smith, que aquilo que falta à TSI para que nela não permaneça um inexplicado intérprete como resto é um aprofundamento ontológico e metafísico, uma análise da responsabilidade da subjectividade (ou região-s, de forma menos antropologista) na constituição da objectividade, nomeadamente na constituição da objectividade do mental e que a partir do momento em que essa análise aparece, a fundamentalidade da EF na teoria é inevitavelmente posta em causa. Apesar disso, e é isso que me importa, o que até aqui se afirmou vai no sentido de confirmar as intuições de Dennett quanto ao estatuto das representações, i.e. vai no sentido de confirmar a dependência dos padrões da mentalidade relativamente a um intérprete, e portanto a oposição ao postulado de um nível de realidade intrínsecamente representacional. Nada vai no entanto, note-se, no sentido de qualquer eliminação da fenomenologia. Pelo contrário, recuperou-se a noção de intencionalidade intrínseca – que poderia ser captada através da noção "como é para o sistema intencional ser um sistema intencional" – que poderia vir a aproximar-se do que se entende por fenomenologia. O que é certo é que a constituição de sistemas como intencionais a partir de uma estratégia, tese que é o teor central da TSI, não pode excluir uma certa unificação pela subjectividade, ligada ou não a uma consciência fenomenal, mas pelo menos geradora em determinadas condições daquilo a que se chamou intencionalidade genuína do próprio intérprete. A teoria da subjectividade inscrita na EI distingue-se assim do absolutismo da terceira pessoa que rege a teoria dennetiana da consciência e constitui base suficiente para discernir e corrigir os defeitos cientistas do tratamento da consciência no MEM.

[825] CANTWELL SMITH 1996: 270.
[825] CANTWELL SMITH 1996: 374-375.

6.3 *A natureza e o seu interior II. Racionalidade: a impossibilidade de irracionalidade e a racionalidade mínima.*

Não foi até agora explicitamente retomada a questão da racionalidade, embora ela fosse suposta de cada vez que se evocou um "intérprete": sem suposição de racionalidade não existe intérprete, sem intérprete não existem Sistemas Intencionais. A questão da racionalidade é nuclear para a TSI, por várias razões. Por um lado, num sentido imediato, ela diz respeito às transições cognitivas num SI particular, admitindo que estas não consistem em computações de representações. Esta última proposta seria evidentemente a proposta da TRM quanto à natureza dos processos cognitivos "que não resultam da irrupção bruta do nível fisiológico"[827]. Abdicando desta ideia, a TSI arrisca-se a não dispôr de qualquer teoria dos mecanismos da racionalidade. De facto, entre a TSI e a TRM joga-se uma decisão quanto ao que é mais fundamental quando se trata de cognição: intencionalidade ou racionalidade. Para a TSI a racionalidade é mais fundamental: é em função dela que podem ser *atribuídos* estados intencionais a sistemas físicos. Para a TRM apenas se pode falar de racionalidade enquanto transformação de representações se existem representações, i.e. estados intencionais *reais* e não apenas atribuídos. Evidentemente se não existem *realmente* representações, a racionalidade terá que ser qualquer outra coisa que não computação de representações.

Mas o problema da natureza das transições cognitivas ou inferências num particular sistema cognitivo é apenas um dos aspectos da questão da racionalidade. Aquilo que torna a questão da racionalidade decisiva na TSI é o facto de esta se apoiar num argumento a favor da *impossibilidade* de irracionalidade dos SI (tem aliás sido acusada disso[828]). De facto, o argumento é duplo: por um lado trata-se de um argumento a favor da *impossibilidade conceptual de irracionalidade*, cuja

[827] É nestes termos que Fodor coloca a questão em FODOR 1975 para referir por exemplo estados e processos dos quais não existe descrição útil em termos de representações e computações (por exemplo sensações).

[828] STICH 1993, Capítulo 2, *Good Reasoning and Intentional Content: How Irrational Can We Be?* Se no núcleo do MEM está uma "relação conceptual" entre experiência e memória, a relação conceptual aqui em jogo é uma "relação conceptual" entre intencionalidade e racionalidade.

raíz é quineana[829], por outro lado de um argumento a favor da *impossibilidade empírica de irracionalidade*, dada a submissão dos sistemas cognitivos à selecção natural[830]. Que o argumento seja duplo revela, de resto, uma ambiguidade: não fica totalmente claro a partir da TSI tal como Dennett a formula se é ou não possível uma *teoria naturalista da racionalidade*. A racionalidade é simultaneamente considerada como uma noção normativa, com apoio na qual sistemas físicos são interpretados como mentais e como algo que comportamentos adequados de sistemas físicos resultantes de selecção natural realizam no seu ambiente (por contraste com, repita-se, um mecanismo de representações e computações *à la* Fodor, o qual seria uma caracterização puramente interna aos sistemas cognitivos daquilo que se entende por racionalidade). Tudo na TSI conduz à conclusão, não perfeitamente explicitada por Dennett mas natural, segundo a qual a racionalidade não é nem uma noção muito profunda (caso em que "racionalidade" seria o nome para qualquer coisa como "valor cognitivo intrínseco") nem uma noção unificada. Isto é assim não apenas no sentido em que a racionalidade é uma noção do nível do agente, i.e. do nível de sistemas cognitivos físicos que se comportam adequadamente, do ponto de vista instrumental, nos seus mundos (nocionais), e não uma noção apoiada na existência de géneros naturais representacionais, mas também porque a racionalidade é por princípio insusceptível de uma caracterização *precisa* (e de *uma só* caracterização). É neste sentido, e na medida em que embora imprescindível à teoria da mente ela só pode ser identificada e avaliada instrumental-

[829] Recapitulando, o princípio que rege a TSI é o princípio segundo o qual *nada pode ser interpretado como intencional se não for suficientemente racional*. Recorde-se o caso de Fido, que aparece em C&C e que foi analisado no Capítulo 1 (ponto 1.2.3), e o caso do vendedor de limonada, que aparece no artigo *Making Sense of Ourselves* e que foi analisado no Capítulo 2 (cf. ponto 2.1.1.7.2). Este princípio afirma que existe uma conexão incontornável entre intencionalidade e racionalidade: a atribuição de crenças a um sistema pressupõe a racionalidade desse sistema, a qual é comportamentalmente avaliada. Já em C&C Dennett afirmava que o carácter adaptativo do comportamento é uma condição necessária da inteligência. De resto, o aludido princípio provem de Quine e acompanha a ideia de teoria da mente como teoria da interpretação (cf. QUINE 1960: 58-59). A proposta de Quine – que pode ser considerada como o princípio acerca de racionalidade que rege todas as teorias quineanas da interpretação, nomeadamente as de Dennett e de Davidson– é a seguinte: para além de um certo ponto, a (suposta) irracionalidade do indivíduo sob interpretação deve fazer com que o intérprete duvide das suas próprias hipóteses de tradução. É mais provável que estas hipóteses estejam mal formuladas do que que o indivíduo a ser interpretado seja desse modo irracional. Esta é obviamente a raíz do princípio da caridade interpretativa, comum a Quine, Davidson e Dennett. O exemplo de irracionalidade improvável do sistema sob interpretação que Quine dá em *Word and Object* é o exemplo de uma crença contraditória, a crença de um indivíduo em p e ~p.

[830] Como se viu no Capítulo 2, a TSI considera que organismos resultantes de selecção natural têm crenças que são na sua maioria verdadeiras e fazem inferências que são na maior parte racionais.

mente, que a racionalidade não tem "valor intrínseco". Este último ponto é muito importante quando se colocam questões como a caracterização "psicologicamente realista" da racionalidade de sujeitos ou a comprovação empírica da irracionalidade desses sujeitos. Estas questões são por exemplo colocadas no âmbito de estudos psicológicos que teriam supostamente implicações pessimistas no que respeita à racionalidade de agentes reais (i.e. esses estudos provariam que agentes reais são bastante irracionais na maior parte das circunstâncias de inferência e decisão[831]).

Do ponto de vista biológico e considerando a evolução por selecção natural de sistemas inteligentes, aquilo de que se fala quando se fala de racionalidade é uma adaptação evolutiva com propósitos determinados. A acepção mínima de racionalidade em causa é a racionalidade instrumental, característica do comportamento de agentes dotados de *estruturas de finalidades*, guiados por *representações*, que fazem uma correcta gestão de meios em ordem aos fins, nomeadamente a uma adaptação suficientemente boa ao ambiente. Uma tal caracterização de racionalidade vale para qualquer sistema cognitivo, consciente ou não, pois diz respeito à adequação do *design*-para-agir[832]. Não é apenas a TSI que reporta a racionalidade à evolução por selecção natural de sistemas cognitivos: essa é uma suposição comum em teorias psicológicas e filosóficas da racionalidade. No entanto, desde logo, considerar a racionalidade a partir da perspectiva da selecção natural como uma adaptação com um propósito determinado conduz à constatação de que o *design* desta não é óptimo mas

[831] Cf. Capítulo 2, ponto 2.1.1.3. Nos aludidos estudos chama-se *irracionalidade* por exemplo à existência de tendências ou inclinações prévias (*bias*) que determinam o enquadramento (*framing*) das escolhas (supostamente racionais) de agentes, às más estimativas de probabilidades, à intransitividade (ou mesmo inversão) das preferências dos agentes, ao facto de os agentes incorrerem sistematicamente em falácias (cf. por exemplo a falácia dos *sunk costs*), etc. Em suma, chama-se irracionais a (1) processos pelos quais agentes chegam a conclusões que não podem ser justificadas a partir do seu próprio conhecimento e a (2) processos que conduzem a uma conclusão ou decisão que não é a melhor que poderia ter sido alcançada à luz da evidência disponível, com os recursos temporais disponíveis. São assim possíveis situações em que agentes acreditam em contradições, não acreditam nas consequências daquilo em que acreditam, incorrem em fraqueza da vontade e em auto-engano, etc. Em todos estes exemplos trata-se de irracionalidade na gestão dos meios, e portanto de algo relativo ao núcleo da noção de racionalidade que é a racionalidade instrumental (este aspecto foi também sublinhado, recorde-se, no Capítulo 2). Não se trata da racionalidade dos fins.

[832] Noutras palavras, agir de acordo com "razões" é um funcionamento que pode ter bases diversas nos sistemas cognitivos, desde a pura e simples instalação (*wiring*) que propicia a adaptação das funções do sistema ao seu ambiente, até ao condicionamento, até ao reconhecimento consciente de conexões entre conteúdos. Este último, embora possa parecer destacar-se relativamente às "bases" anteriores pode igualmente ter sido seleccionado (cf. NOZICK 1993: 108).

apenas satisfatório, suficientemente bom, uma vez que esse é o estatuto geral dos produtos da selecção natural. Ora, isso significa no limite, como R. Nozick sublinha[833], que se o facto de a produção de inferências (em alguns agentes, nomeadamente humanos) ser confiada a regras cumpre uma função biológica interessante para os seres em quem tal funcionamento está instalado, não se deve esquecer que as regras são seleccionadas pelo êxito obtido na acção pelos agentes que as seguem e que esse êxito não depende de modo algum da perfeição das regras mas sim do facto de elas conseguirem o máximo possível na negociação (*trade-off*) entre as capacidades cognitivas (limitadas), o tempo de resposta e a quantidade de informação obtida. A reportação da racionalidade à selecção natural é uma reportação ao êxito no comportamento do agente e apenas isso e é portanto também, inevitavelmente, um passo no sentido de duvidar da qualidade intrínseca dos produtos e resultados da racionalidade.

A abordagem a partir da selecção natural oferece assim um argumento a favor da impossibilidade de irracionalidade mas coloca em relevo a ausência de qualquer constrangimento de "qualidade" sobre os processos e produtos da racionalidade: para serem seleccionados os mecanismos da racionalidade devem ser satisfatoriamente bons, não necessariamente perfeitos. A abordagem a partir da teoria da interpretação coloca a questão dos constrangimentos de forma bem diferente. O princípio da caridade interpretativa comum aos quineanos conduz a uma dada perspectiva da impossibilidade de irracionalidade em Sistemas Intencionais[834]. Os quineanos

[833] NOZICK 1993, Capítulo 4, *Evolutionary Reasons*.

[834] O lugar originário do princípio da caridade é o pensamento de Quine (cf. o parágrafo 13 do Capítulo 2 de *Word and Object*, capítulo no qual se trata de tradução e significado, sobre a tradução dos conectivos lógicos). O princípio da caridade acompanha assim desde logo aquele que no Capítulo 1 do presente trabalho (cf. 1.1) foi considerado como sendo o ponto matricial da filosofia da mente contemporânea, a observação de Quine acerca da tese de Brentano (§45 de *Word and Object*) que gera as teorias do mental que são teorias da interpretação, nomeadamente a tradução radical do próprio Quine, a TSI de Dennett e a interpretação radical de Davidson. Quine afirma aí o seguinte: «To accept the extreme case, let us suppose that certain natives are said to accept as true certain sentences translatable in the form "p and not p". Now this claim is absurd under our criteria [*para traduzir conectivos lógicos*]. And, not to be dogmatic about them, what criteria might one prefer? Wanton translation can make natives sound as queer as one pleases. Better translation imposes our logic upon them, and would beg the question of prelogicality, if there were a question to beg (...) The maxim of translation underlying all of this is that assertions startingly false on the face of them are likely to turn on hidden differences in language (...) The common sense behind the maxim is that one's interlocutor silliness, beyond a certain point, is less likely than bad translation-or, in the domestic case, linguistic divergence» (QUINE 1960: 58-59). O núcleo de qualquer princípio de caridade, incluindo o da TSI é portanto, repita-se, a suposição de que os agentes racionais têm crenças que são na sua maioria verdadeiras e fazem inferências que são na sua maioria racionais.

podem, é certo, ter posições diferentes quanto ao estatuto da racionalidade suposta na interpetação, como é claro nos casos de Dennett e Davidson. Davidson sugere que a verdadeira natureza da racionalidade suposta na interpretação é claramente exposta pela teoria da decisão[835], a TSI conduz a pensar que a racionalidade não tem verdadeira natureza, sendo um conceito sistematicamente pré-teórico e sendo o estatuto das caracterizações de uma teoria explícita da racionalidade sempre condicional e provisório. Mas o que é certo é que quer Davidson quer Dennett consideram que a interpretação supondo a racionalidade (a caridade, na terminologia difundida por Davidson) não é uma opção mas sim uma condição para uma teoria do mental. No entanto, essa condição deixa em aberto saber que racionalidade é essa que é assim inapelavelmente suposta, bem como quanta racionalidade tem que ser suposta. Estes problemas foram nomeados, recorde-se, nos Capítulos 1 e 2 do presente trabalho (nomeadamente na sequência dos já aludidos exemplos de Fido e do vendedor de limonada). O problema que se coloca é portanto saber se se poderá ou deverá supor, por razões de caridade interpretativa, uma racionalidade perfeita, apesar do que as abordagens a partir da selecção natural e os estudos psicológicos têm a dizer.

O que as teorias quineanas da interpretação normalmente fazem é apelar a intuições acerca de irracionalidade (pense-se no exemplo de irracionalidade improvável do próprio Quine, o indivíduo que crê p e ~p). Mas onde se vai buscar os standards de racionalidade quando se apela a intuições de irracionalidade? E sobretudo, que standards de racionalidade mantém ainda alguém que afirma, como se viu Dennett fazer no Capítulo 2 deste trabalho[836], que a racionalidade não consiste nem em consistência das crenças, nem em fechamento dedutivo nem na capacidade inferencial perfeita de um agente? No que toca à racionalidade, ao longo das exposições da teoria dennetiana do mental, passou-se por (i) situações em que se afirmou, verificado o choque entre a racionalidade e a incorrigibilidade dos SI, que a atribuição de racionalidade era sempre global ou holista e assim de certo modo "permissiva" no sentido em que admitia alguma inconsistência[837], (ii) situações em que a insuficiência computacional, a incapacidade de inferências perfeitas considerando todos os factores, foi evocada como razão suficiente para a

[835] Cf. DAVIDSON 1980 c, *Hempel on Explaining Action*.

[836] Ponto 2.1.1.7.2.

[837] Cf. Capítulo 2 (ponto 2.1.1.6).

oposição às éticas da maximização racional[838], (iii) situações em que se afirmou que a solução para (aquilo a que na IA se chama) o problema do enquadramento passava, nos humanos, por uma não obrigação à coerência das crenças do sistema[839]. A racionalidade em causa foi, assim, sempre, de algum modo e desde logo, limitada.

Não se chega, seguramente, aos standards de racionalidade necessários às teorias da interpretação a partir de evidência indutiva (se há evidência indutiva de alguma coisa é de irracionalidade e insuficiência cognitiva dos SI). Logo, o caminho para a racionalidade só pode passar por qualquer coisa como um constrangimento conceptual: se um agente é insuficientemente racional ele não terá estados cognitivos intencionalmente descritíveis. Este é um ponto do próprio Quine e um princípio adoptado por Dennet. Mas, de novo, mesmo admitindo que existe uma conexão "conceptual" entre racionalidade e intencionalidade, quanta racionalidade é necessário supor quando se interpreta? Um ponto de acordo não muito promissor, aparentemente, é o facto de, se nos usamos a nós próprios como modelo, ser consensual que a nossa racionalidade não é muita ou muito perfeita[840].

Ao contrário do que alguns críticos de Dennett afirmam[841], a TSI não supõe uma racionalidade perfeita. Torna-se então necessário avançar alguma proposta quanto à "quantidade" de racionalidade que lhe é necessária. A melhor sugestão é a de uma racionalidade mínima. Na proposta de C. Cherniak[842] a expressão "racionalidade mínima" significa que não há um conjunto específico de inferências que um agente tenha que manifestar para ser candidato à descrição intencional, o agente deve manifestar apenas um conjunto razoavelmente subtancial das inferências que seriam requeridas de um agente racional perfeito. O importante na noção é o facto de ela se opor

[838] Cf. Capítulo 4 (ponto 4.5).

[839] Cf. Capítulo 5 (5.2.2). Recorde-se que se afirmou então que «Nada na implementação dos hábitos (que são ideias práticas, razão por instinto, feixes de como-lidar) na criatura obriga a preservar a consistência do que eles instauram em termos de imagem manifesta».

[840] Partindo do constrangimento ou obrigação de nos supormos a nós próprios como modelo, S. Stich (STICH 1993: 50) fala de um "princípio de chauvinismo intencional". Ser intencionalmente descritível requer a presença uma dinâmica cognitiva semelhante à nossa e criaturas muito diferentes de nós não serão, assim, intencionalmente descritíveis. No entanto, um tal limite imposto à (ir)racionalidade tem razões (relativas ao observador) que são perfeitamente desinteressantes. A in-descritibilidade intencional não assinala, nestas condições, nenhuma fronteira natural. Em STICH 1993 Stich procura extrair as consequências (curiosas) deste chauvinismo intencional.

[841] Cf. por exemplo STICH 1993.

[842] CHERNIAK 1986.

a um apriorismo definitivo: não é simplesmente possível enumerar a priori um núcleo de crenças verdadeiras e de princípios racionais sem os quais não existem agentes racionais. Dada a ligação entre racionalidade e intencionalidade e esta vagueza da racionalidade é claro que uma caracterização intencional não poderá também ser uma questão de tudo ou nada. Mas essa ideia está, precisamente, de acordo com a TSI e o seu anti-representacionismo.

No que diz respeito à evolução do design para a racionalidade, as ideias sem as quais a TSI não funciona, racionalidade instrumental e algo como racionalidade mínima, dão-se tão bem com o conceito de *satisficing* como com o conceito de optimização. Nenhuma requer (pelo contrário ambas parecem excluir) uma racionalidade óptima ou perfeita. Evidentemente, o que acontece nestas circunstâncias é que a qualidade intrínseca dos produtos e processos de uma racionalidade não óptima nem perfeita não fica garantida. Mas nada comprometia à partida a TSI com semelhante ideia. Tudo o que há na proposta da TSI como teoria da interpretação supondo a racionalidade é a afirmação de que a racionalidade é uma *condição* da caracterização intencional. Nada na TSI obriga a que a racionalidade *perfeita* seja uma condição *necessária* da caracterização intencional.

É evidente que a suposição de racionalidade das teorias quineanas do mental pode ser problemática quando considerada conjuntamente com os estudos empíricos da racionalidade e levando em conta as implicações negras e pessimistas destes. Pior: na medida em que as teorias quineanas do mental representam um argumento a favor da impossibilidade de racionalidade de todo o sistema intencionalmente interpretável, parece desaparecer qualquer espaço para uma epistemologia correctiva, para a legitimidade da mera afirmação de que um sujeito está a raciocinar mal. Quanto à primeira questão, ela é, como Davidson viu claramente, recondutível ao que se entende por comprovação ou infirmação empírica da racionalidade. A comprovação seria supostamente obtida através dos estudos acerca de racionalidade de feitos por exemplo por psicólogos[843]. No entanto, não é razoável admitir que estudos do género provam que as pessoas são irracionais num sentido importante para a questão em causa (da ligação conceptual entre intencionalidade e racionalidade). Davidson, enfrentan-

[843] Cf. TVERSKY & KAHNEMAN 1981, TVERSKY & KAHNEMAN 1982, TVERSKY & KAHNEMAN 1993, SUTHERLAND 1992. Mesmo que se trate de sujeitos academicamente treinados em lógica ou teoria das probabilidades há certas irracionalidades recorrentes, dominantes (dir-se-ia na terminologia da teoria dos jogos que as estratégias dominantes são irracionais...).

do a questão, considera que experiências empíricas acerca de racionalidade não podem infirmar as caracterizações que a teoria da racionalidade que ele toma como definindo a verdadeira natureza da racionalidade (a teoria da decisão) faz da estrutura do agente racional: não saberíamos imaginar o que é para um agente ser racional (ou irracional) sem o quadro teórico da teoria da decisão. Noutras palavras, para Davidson a caracterização do agente (no caso feita pela teoria da decisão) é constitutiva daquilo que é ser racional e agir racionalmente. É óbvio que há uma espécie de circularidade na posição de Davidson: a racionalidade constitutiva do agente é encontrada porque é "lá posta". Na terminologia de A.Zilhão e forçando um pouco os termos do próprio Davidson, isto significa que os axiomas da teoria da decisão são verdades sintéticas a priori acerca de seres racionais quaisquer[844]. Essa é a razão pela qual Davidson pensa que os elementos da teoria da decisão (por exemplo as considerações acerca da transitividade das preferências do agente) têm que ser preservados face a resultados como os de Tversky e Kahneman que mostram como é comum a inversão de preferências em agentes reais, bastando que existam variações do enquadramento da decisão em causa[845]. Davidson conhece os trabalhos de A. Tversky e dos seus colaboradores e esteve ele próprio envolvido em investigações experimentais acerca de racionalidade. Simplesmente, pensa que é ilegítimo supor que através de tais investigações se confirma ou infirma empiricamente a racionalidade. Quando se enfrenta o problema da adequação empírica de caracterizações idealizantes do agente racional, que são essenciais em propostas como as de Davidson e Dennett, é preciso levar em conta que nenhuma interpretação de factos experimentalmente obtidos pode ser feita sem utilizar alguma caracterização daquilo que é um agente racional. Pura e simplesmente não sabemos nem podemos pensar de outra maneira: não podemos deixar de pensar que, se pensamos, pensamos (ou qualquer outra pessoa pensa) geralmente de forma racional (a não ser deixando de pensar que pensamos, por exemplo atribuindo erros a irrupções de distúrbios ao nível do *hardware* cognitivo). Substitua-se a teoria da decisão por uma caracterização sistematicamente pré-teórica da racionalidade e obtém-se uma posição análoga – a que se poderia chamar uma obrigação à idealização mas agora sem obrigação a uma teoria da verdadeira natureza do objecto idealizado – na TSI. Não há

[844] Cf. ZILHÃO 1998/1999 e DAVIDSON 1980 c: 261, *Hempel on Explaining Action.*
[845] TVERSKY & KAHNEMAN 1981.

como negar que se está perante circularidade na tentativa de justificar ou fundamentar a racionalidade. Mas esta é uma situação incontornável em teorias da racionalidade de qualquer género, provavelmente relacionada com o tipo de seres que as fazem, seres instrumentalmente e minimamente racionais, e também com a mera inexistência de alternativa. De facto, como senão através de actividade cognitiva se avaliaria a actividade cognitiva?

É evidentemente também um problema, dada uma tal circularidade, o facto de a racionalidade ser a chave para a psicologia e, do ponto de vista ético, para a pessoalidade. Se a dimensão moral e metafísica da pessoalidade, analisada no Capítulo 4, não parece possível a não ser com apoio numa mecânica do controlo, da centralidade e da auto-representação, relativamente à qual o MEM apresenta uma proposta, ela depende em última análise de uma teoria normativa que precisamente supõe a racionalidade. A situação é agravada pelo facto de Dennett, ao contrário de Davidson, pensar que essa chave, a racionalidade, não tem uma verdadeira natureza. Repare-se que Davidson não apenas pensa que a verdadeira natureza da racionalidade é dada por uma teoria formal, a teoria da decisão, como também pensa que apenas os humanos são verdadeiramente racionais. A TSI não tem que fazer restrições idênticas.

Poder-se-ia sugerir que a resposta ao problema da circularidade já está dada através da análise do estatuto da EI feita nos pontos 6.1 e 6.2 do presente capítulo. Se é certo que em termos de racionalidade se encontra apenas aquilo que se coloca a partir do ponto de vista do intérprete, é certo também que o ponto de vista do intérprete é auto-constitutivo. Por trivial que pareça esta observação, ela significa por exemplo que a racionalidade não é uma característica do funcionamento neurofisiológico dos neurónios ou de outro *hardware* qualquer do agente, nem de processos computacionais definidos sobre entidades intrinsecamente representacionais. O seu estatuto é o do artificial de H. Simon, o interface, a adaptação, no caso na relação do pensamento ao próprio pensamento. A racionalidade supõe um ponto de vista psicológico, um intérprete que o reconheça: esse é o lugar da psicologia e o lugar da pessoalidade possível.

Aquilo a que pretendo chegar é o seguinte: a introdução de uma suposição de racionalidade na teoria do mental, comum às teorias quineanas, não é necessariamente acompanhada pela convicção de que teorias formais da "racionalidade" (como a teoria da decisão ou mesmo a lógica) eventualmente exibirão a verdadeira natureza da

racionalidade. É possível manter coerentemente a ideia para a qual a TSI aponta, segundo a qual a racionalidade não tem uma verdadeira natureza (não consiste em fechamento dedutivo, capacidade inferencial perfeita ou coerência total no corpo das crenças). O ponto de referência da racionalidade é apenas a adaptação suficientemente boa ao ambiente do comportamento, guiado por representações, de agentes cognitivos físicos dotados de uma estrutura de finalidades. Evidentemente, nestas condições, parte do trabalho de uma teoria filosófica da racionalidade consiste em justificar o seu próprio estatuto de "equilíbrio reflectido" entre práticas de racionalidade, formalizadas e comuns, e respectivas intuições, por um lado e formalização de princípios e regras por outro.

O mínimo que se pode concluir de um tal estatuto de imperfeição constitutiva é que não é permitido falar de Uma Racionalidade. Assim, por exemplo a possibilidade de a força das nossas intuições a favor de certos "princípios de racionalidade" (por exemplo princípios lógicos) não ser prova definitiva de nada deve ser encarada. A aparente auto-evidência de conexões de conteúdos pensados, a força do "segue-se de..", não será, em tais circunstâncias, garantia alguma de que uma Conexão se sustente. Evidentemente, a inexistência de qualquer ponto de apoio exterior ao funcionamento desses mesmos princípios justifica por exemplo o facto de a investigação de leis mais profundas da racionalidade (por exemplo leis mais profundas da lógica) não continuar do mesmo modo que continua a investigação de leis mais profundas da física[846].

Há uma conclusão de tudo isto, bem mais alargada do que qualquer conclusão que Dennett arrisca extrair mas que por exemplo S. Stich formula em *The Fragmentation of Reason*. A ideia é que *não há possibilidade de formular constrangimentos a priori para todos os agentes racionais possíveis*. Não existe A Racionalidade, o Agente Racional. Dennett não extrai uma tal conclusão, no entanto, a sua ideia de uma caracterização sistematicamente pré-teórica de racionalidade na TSI, bem como a pluralidade dos mundos nocionais aos quais o comportamento dos agentes se adequa, aponta numa direcção idêntica (i.e. para aquilo a que Stich chama a fragmentação da razão[847]).

[846] Ideia explorada (ou melhor, problema caracterizado) em NOZICK 1993.

[847] Que de resto Stich pensa poder ser acompanhada por um pluralismo cognitivo normativo, o que significa que embora não exista uma relação entre O bom raciocínio e a Verdade, existem relações entre bom pensamento e verdade. I.e., não se abdica da avaliação de estratégias cognitivas, mas não se ignora a sua dimensão de estratégia (é isto o *pragmatismo*). Apenas uma posição semelhante

Apesar da conclusão a que o argumento quineano da impossibilidade de irracionalidade acaba por conduzir, a fiabilidade do desenho desenvolvido nas condições evolutivas referidas, poderia mantê-lo e garanti-lo no presente. Esta situação pode perfeitamente conviver com o já aludido equilíbrio reflectido (como hipótese acerca do estatuto de uma teoria explícita da racionalidade). Recorde-se que a aplicação do equilíbrio reflexivo às teorias explícitas da racionalidade produz o seguinte: aceita-se uma dada norma de inferência porque ela produz inferências que consideramos intuitivamente válidas e por outro lado considera-se válidas as inferências que sejam produto das regras que tiverem sido aceites. Um equilíbrio não definitivo entre intuição e regras é a única alternativa de justificação da racionalidade numa situação como aquela que é caracterizada pela TSI em que existem mentes "mal desenhadas" e não existem fundamentos únicos e reais nem natureza profunda da racionalidade. Nada disto impede que se continue a afirmar que estudos empíricos não podem provar que as pessoas são irracionais na grande maioria dos casos ou sempre. Sem racionalidade nenhuma entidade é concebível como irracional ou sequer como mental. Nada disso obriga no entanto a afirmar que existe Uma só racionalidade definidora do próprio conceito de agência, daquilo que se entende por intencionalidade, por psicologia. A racionalidade suposta é uma (particular) idealização, uma idealização de alguma forma reportada aos funcionamentos imperfeitos que são os funcionamentos reais dos agentes que nós somos.

A relação da racionalidade com a pessoalidade foi considerada no Capítulo 4 nos seus dois aspectos de auto-avaliação e compromisso com a racionalidade própria. Foi assim definido um espaço de manobra para o controlo e representação de si e para o controlo e representação do pensamento próprio possibilitando a ligação entre auto-consciência e auto-agenciamento e consequentemente um novo âmbito para o pensamento racional voluntário. Note-se que é aqui que nasce não apenas a questão moral, a possibilidade de pessoas e acções e a constituição de pessoas através de acções,

leva a sério, de resto, o facto de o núcleo mínimo de racionalidade ser um núcleo *instrumental*. Uma teoria do mental como interpretação não pode deixar de considerar esse núcleo instrumental, o que significa que ela pressupõe uma avaliação consequencialista dos sistemas cognitivos. Ora a questão seguinte é saber o que importa aos agentes conseguirem ou atingirem (e porque é que haveria de ser a Verdade? Porque é que crenças verdadeiras seriam preferidas relativamente a processos pragmaticamente sancionados, com uma conexão óbvia com aquilo que o agente deseja?). Esta situação é o princípio do aludido pragmatismo e é desenvolvida em STICH 1993.

através de auto-controlo meta-reflexivo e avaliação forte, como também o problema filosófico da natureza e alcance do pensamento racional e não apenas da cognição. Tudo o que até aqui se explicitou acerca de racionalidade vai no sentido de mostrar por que razões todos os processos que supõem a racionalidade acontecem sempre incompletamente, constituindo mais uma razão para pensar que as noções de pessoa e acção, relativas a seres capazes de pensamento voluntário e não apenas de cognição, não são noções absolutas. Como diria R. Nozick, a história natural da racionalidade mostra um trajecto que vai desde os *free floating rationales*, as razões para funcionamentos e comportamentos totalmente inapercebidas pelos "agentes" que são tão importantes para a TSI, até uma preocupação com razões, um cuidado com o raciocínio e a qualidade deste, que «*now floats free*»[848], existe livremente, sem qualquer propósito instrumental imediato, e que é até agora característica dos humanos e que os configura como entidades racionais conscientes e pessoais. É apenas relativamente a este adquirido tardio na história da espécie que pode ser posta em dúvida a pertinência de uma teoria da racionalidade que não se descole, de alguma forma, do núcleo instrumentalista.

6.4 *A natureza e o seu interior III. Consciência fenomenal ou ilusão do utilizador de uma Máquina Virtual. Epifenomenismo.*

A dependência da mentalidade relativamente a uma abordagem, que é o núcleo da TSI, tem aspectos aceitáveis e inaceitáveis. Os primeiros foram aprofundados nos pontos anteriores do presente capítulo com as sugestões relativas a intencionalidade e racionalidade. Quando se trata de consciência, no entanto, é especialmente difícil aceitar que se fale apenas de *interpretação* de sistemas, devendo pelo menos ser considerada a hipótese de um facto bruto para o próprio sistema a que se tem chamado "sentir-se ser". Ora, enquanto meta-teoria da cognição, a TSI tem como objectivo ao mesmo tempo estabelecer a dependência da mentalidade relativamente a uma interpretação e eliminar o observador (substituído por agentes, estes seriam eventualmente "dispensados" – é esse o propósito dos modelos funcionalistas). A grande justificação para eliminar o

[848] NOZICK 1993: 114.

observador é o facto de não haver mais ninguém, uma vez deixado para trás o agente cognitivo global (apenas este tem um "ponto de vista"). Se a observação colhe enquanto relativa à teoria cognitiva sub-pessoal não é no entanto "humanamente possível" deixar totalmente para trás o ponto de vista numa teoria geral da mente. Já se viu porquê, num primeiro sentido relativo a intencionalidade e racionalidade. Há um segundo sentido em que o ponto de vista não pode ser deixado para trás e que diz respeito a acepções de "consciência" que não ficam cobertas com o que foi até aqui afirmado acerca de intencionalidade e racionalidade.

Acontece que as motivações de Dennett como teórico da cognição (ser anti-representacionista, ser quineano, eliminar o observador) por vezes se misturam e o ponto em que se misturam pior é o ponto em que uma teoria propositadamente anti-cartesiana da consciência (que é não apenas anti-dualista mas inclusivamente afasta a "diferença de género" entre consciência e não-consciência) rejeita, por motivos que são precisamente (e erroneamente) cartesianos, certas teses que deveria incluir. Assiste-se assim na teoria dennettiana da consciência, a um retorno, com maus resultados, do cartesianismo "recalcado". É isso antes de mais que pretendo aqui mostrar. Observou-se já que se Dennett, ao contrário de, por exemplo, Fodor e Dretske, se atem ao interpretativismo quando considera a intencionalidade é em grande medida porque se recusa a admitir um certo entendimento genuíno (ou, noutras palavras, a diferença entre mentalidade genuína e mentalidade atribuída). Ora esta recusa é função de um preconceito cartesiano inexplicitado de acordo com o qual ou a mente está ligada a um centro ou unidade naturais, dados, ou então não será mente. De facto, se, segundo Dennett, não existe entendimento genuíno por oposição a mentalidade atribuída, é unicamente porque não existe "um lugar funcional central" onde tudo se põe conjuntamente para o entendedor entender. Viu-se no ponto 6.1 que o argumento contra a intencionalidade genuína não colhe na medida em que não é uma unidade desse tipo que está em causa: o *quê* do nível pessoal (a unidade) não tem por que ser confundido com o *como* do nível sub-pessoal, i.e. o nível dos veículos, que o assegura (o processamento paralelo e distribuído). De resto e ao nível da explicação do *como*, o próprio MEM é uma teoria da *unidade constituída*. Ora, é de um argumento semelhante, que parte da inexistência de um lugar funcional central e único, ao qual se junta uma crítica linguística de propriedades cria-

das pelo próprio Dennett, os *qualia*-segundo-Dennett, bem como o pressuposto de uma "ligação conceptual" entre consciência e memória, que depende a rasura da noção de consciência da acepção de "sentir-se ser" para o próprio sistema. Mas do mesmo modo que tal argumentação não é suficiente para afastar o entendimento genuíno, ela não é suficiente para afastar a possibilidade de um sentir-se ser não identificável nem com um lugar de unidade funcional nem com características da posse epistémica dos conteúdos apercebidos nos próprios sistemas cognitivos teorizados pelo MEM.

Recapitular-se-á em seguida brevemente aquilo que neste trabalho foi estabelecido acerca da consciência pretendendo mostrar que o problema da teoria dennettiana não é ser uma negação verificacionista desta, mas o facto de envolver alguns equívocos importantes quanto àquilo de que se fala. É um facto que a teoria dennettiana da consciência faz colapsar a consciência sobre o conteúdo. Mas um tal continuismo não é o principal factor de perturbação: Dennett avança boas razões para o sustentar. O problema é que o continuismo não tem por que conduzir à negação da especificidade do "sentir-se ser" do próprio sistema consciente.

Recorde-se antes de mais que a passagem da teorização dos SI a partir da EI para uma teoria do interior dos sistemas cognitivos é, na obra de Dennett, justificada pelo seguinte. A suposição de racionalidade (ou interpretação) pressupõe a *unidade* do agente globalmente considerado. Há no entanto aí alguma coisa que é tomada como dada – que sistemas cognitivos são *unidades* — que terá que ser "explicada". A teoria dennettiana da consciência pretende oferecer tal explicação. A maneira de conceber o mental depende, assim, do ponto de referência escolhido: ou bem se visa entidades globais comportando-se no ambiente de forma adaptada e aí tem-se a intencionalidade e a racionalidade, constantemente reportadas a um macro-nível, ou bem se visa o interior de sistemas, onde se tem multiplicidade, competição e unidade e centralidade constituídas. Neste quadro, sob o título de consciência entendeu-se, olhando para a obra de Dennett na sua totalidade, um conjunto de problemas que engloba (i) o estatuto epistémico de relatos que sistemas cognitivos fazem acerca de estados mentais próprios, (ii) a unificação do fluxo da consciência a partir das produções múltiplas de agentes e (iii) a virtualidade da centralidade, controlo e presencialidade do mental. As soluções propostas foram referidas no Capítulo 3 sob os títulos, respectivamente, de (sentimento de) incorrigibilidade por razões funcionalistas[849], ilusão do utilizador de uma Máquina Virtual serial[850], virtualidade da centralidade e do controlo[851], ilusão da imanência[852] e princípio da presença virtual[853]. Uma vez identificado o problema da consciência com os referidos problemas, toda a teoria dennetiana da consciência é inevitavelmente uma teoria dos *conteúdos* fenomenológicos. Até aqui, nada a objectar: em grande medida é isso que a consciência é. O problema da teoria é um problema argumentativo, nomeadamente a utilização de pontos que são cognitivos e epistemológicos como dispositivo de eliminação da fenomenologia. Noutras palavras, a questão da consciência só assume a forma que assume devido a uma presunção intelectualista, uma identificação do perceber com perceber-que, a qual se associará ao verificacionismo e a um estranho cartesianismo obsessivo. Vejamos como é que isso se passa ao longo da obra de Dennett.

Em C&C, o primeiro ganho permitido pela atenção (anti-witttgensteiniana e anti-ryleana) às teorias cognitivas sub-pessoais é a conclusão de que aquilo que se evoca sob o nome de "consciência" não é um traço único mas vários traços, ligados nomeadamente ao controlo do comportamento e à reportabilidade. Embora se admitisse em C&C essa variedade, ela coexistia ainda com a ideia de uma linha divisória entre consciência e não-consciência, i.e. com uma diferença clara entre consciência e não consciência. Mas o que interessa sobretudo verificar é que a colocação do problema da consciência em termos de vários traços sub-pessoais permite defender que para além do apercebimento ligado ao controlo e à reportabilidade não há uma coisa a mais, que seria um apercebimento sentido ou fenomenal. Toda a problemática da consciência é "intencionalizada", i.e. transposta para o âmbito do conteúdo, e – esse é o ponto que verdadeiramente importa – ao mesmo tempo para um âmbito epistemológico, focalizado na questão dos relatos acerca dos conteúdos mentais próprios. "Centro da fala" era o nome proposto em C&C para a estrutura cognitiva que permitiria ao sistema fazer relatos linguísticos. O centro de fala era essencial pela ligação da unidade à linguagem e portanto da consciência à reportabilida-

[849] Cf.Capítulo 3, especialmente 3.3.3, 3.3.5 e 3.3.6.2.
[850] Cf. Capítulo3, especialmente 3.3.2, 3.3.3, 3.3.4 e 3.3.5.
[851] Cf. Capítulo 3, especialmente 3.3.5.1, 3.3.5.2 e 3.3.6.
[852] Cf. Capítulo 3, especialmente 3.3.6.4.
[853] Cf. Capítulo 3, especialmente 3.3.6.4 e 3.3.6.5.

de. Ora, embora a linha do apercebimento, e correlativa cisão entre consciência e não-consciência, venha a ser afastada e substituída pelos esboços múltiplos, Dennett continuará a necessitar desse centro fantasmático na teoria da consciência: a lamentação pelo centro que não está lá naturalmente é o que lhe permitirá eliminar a especificidade do "sentir-se ser".

É certo que ao longo do tratamento da consciência exposto neste trabalho Dennett faz mais do que argumentar acerca de incorrigibilidade e reportabilidade: ele apresenta modelos funcionalistas sub-pessoais que dão corpo aos princípios mais gerais da teoria da consciência, a que se chamou fisicalismo funcionalista e superveniência lógica. O problema é o movimento estratégico seguinte: o intelectualismo apoia o estabelecimento de uma relação determinada entre os argumentos acerca de incorrigibilidade e reportabilidade e os modelos funcionalistas da cognição. Esse problema não se coloca em geral: a ligação do intelectualismo à presença virtual nos modelos da cognição é pertinente e provavelmente Dennett aponta os princípios responsáveis pelo mascarar das faltas e interrupções que "deveriam" caracterizar a existência dos conteúdos em vidas mentais baseadas em suportes físicos como cérebros. Mas quando o intelectualismo se espalha à totalidade da vida mental e do auto-apercebimento a questão é outra.

O elo de que Dennett se serve para estabelecer a correspondência intelectualista entre incorrigibilidade e modelos funcionalistas é a "relação conceptual" entre experiência e memória, absolutamente central nos modelos de BS e CE. Recapitulando os princípios subjacentes aos modelos obtem-se fundamentalmente as seguintes ideias: (i) o apercebimento de um sistema cognitivo é mais alargado do que a "experiência consciente", (ii) a existência de apercebimento é identificável com a disponibilidade de um saber-que na *memória* do sistema. A expressão ou "publicação" de qualquer conteúdo, identificada com a consciência, envolve um acesso subpessoal entre memória e mecanismos cognitivos para a reportabilidade. Sendo o conteúdo do apercebimento tudo aquilo que é registado na memória, esse conteúdo pode degradar-se ou sofrer interferências antes de ser recuperado para acesso público (e só a esse nível se poria a questão da convicção subjectiva). É por contraste com os acessos sub-pessoais, que o acesso pessoal ou reportabilidade de si para si é identificado com a *consciência*. Nestas circunstâncias, a autoridade introspectiva é apenas um sentimento (e não

uma garantia) que advém da forma como se dão as intenções semânticas (que têm como base uma ligação entre controlo e reportabilidade), sendo a explicação de toda a situação relativa a mecanismos cognitivos sub-pessoais. "Consciência" não é assim um género natural: não há autoridade subjectiva quanto a géneros naturais, os géneros naturais são da competência da investigação empírica, nada garantindo a priori que a consciência seja um. A teoria da consciência só pode ser uma *teoria do apercebimento-memória*, da forma como os *conteúdos* estão presentes, se dão unificadamente no acesso de si a si de um sistema cognitivo e são *expressos*. A agenda subjacente ao modelo de CE é evidentemente idêntica, daí que a apresentação do modelo desde logo se erga contra a categoria, considerada bizarra, do "objectivamente subjectivo" (que faria parte da experiência do sujeito mesmo que este não apercebesse-que). A partir do princípio intelectualista e da relação conceptual entre consciência e memória, são apresentadas, através da interpretação dos casos de anomalias temporais, novas razões para o verificacionismo defendido a propósito das fenomenologias nucleares em BS (sonhos, imagens e dor). Recorde-se que de acordo com os princípios claramente enunciados pela primeira vez a propósito dos sonhos em BS, é impossível *por princípio* a consciência de um estímulo na ausência de crença nessa consciência. Não se poderia exigir pressupostos mais cartesianos, i.e. preconceituosamente centralistas e intelectualistas, para a captura daquilo em que consiste a consciência e são estes pressupostos que perturbam o continuismo da teoria dennettiana da consciência. Quando Dennett pensa na consciência, aquilo que tem em mente é um sistema cognitivo em que tudo vai ter a um centro e ao que o centro sabe e exprime. O problema é saber se é apenas isso que está em causa. Está sem dúvida em causa, a certo ponto, quando se trata de consciência, a *unidade* de uma mente, e essa questão é tratada no âmbito do MEM como relativa à instauração de uma Máquina Virtual como centro de controlo num dado *hardware* físico. A natureza de tal unidade não é, no entanto, sequer estritamente identificável com a unificação pela intencionalidade de que a TSI trata. A unidade de que o MEM trata mais legitimamente é, relativamente aos humanos e aos cérebros destes, a unidade do sujeito psicológico, a unidade do eu pessoal, empírico, possibilitada pelo funcionamento dos mecanismos sub-pessoais da cognição. Ora, parte dos equívocos que acompanham o MEM dependem da suposição de Dennett de que o que está em

causa com as questões do Eu, do centro, da unificação, é algo mais do que apenas essa *unidade de forma*. De qualquer modo, o MEM *reconstitui* a unidade da consciência, não a elimina de forma alguma. Mais: o MEM é, apesar de toda a oposição ao Teatro Cartesiano, uma teoria da *percepção interior*, que dá mais atenção do que aquilo que é habitual nas teorias da percepção interior estritamente filosóficas a características mecânicas sub-pessoais (nomeadamente põe em relevo a ilusão do utilizador de uma Máquina Virtual, cuja operação explicaria as características da fenomenologia, pretendendo assim preencher o abismo nageliano entre fisiologia e fenomenologia). Em concreto, o estatuto de ilusão do utilizador de uma Máquina Virtual instalada num *hardware* de processamento paralelo é o estatuto do *controlo* e da *serialidade* e portanto da *unificação* de um decurso de conteúdos no tempo, características de forma associadas à consciência nos humanos. Para além da percepção interior e correlativa ilusão do utilizador, o modelo inclui ainda uma teoria do Eu como "centro de gravidade" de uma representação narrativa de si feita pelo sistema. Esse Eu é associado quer à unificação e centralização do controlo e da intenção, quer à autoria da fenomenologia. Todos os "efeitos" mencionados – centro, controlo, serialidade, autoria, representação narrativa de si – são resultado da competição entre agentes que caracteriza o nível sub-pessoal do sistema e é precisamente a caracterização do interior do sistema cognitivo, a que se chamou no Capítulo 3 um *close up* do observador, que mostra que é impossível que a unidade do "autor" de uma vida mental, a unificação do fluxo fenomenológico e a diferença entre consciência e não-consciência, sejam consideradas como dadas.

Até aqui nada a objectar. O problema é que acompanhando esta descrição cognitiva está *um equívoco na concepção de percepção interior*: Dennett tanto considera esta como auto-monitorização sub-pessoal como a considera ao nível do conteúdo. É em termos de conteúdo que se coloca por exemplo a pretensão acerca da não existência uma correspondência natural entre tempo fenomenológico e tempo físico do processamento. De acordo com Dennett, só existiria uma tal correspondência se houvesse um lugar no sistema que fosse uma meta – este é o obsessivo centro cartesiano, em cuja falta se insiste, como se ele tivesse que ser criticado por não estar lá. Se Dennett não fundisse a auto-monitorização mecânica com a auto-monitorização ao nível do conteúdo, perceber-se-ia mais claramente a utilização de algo que ainda não foi justificado: o facto de

se estar a conceber a percepção interior intencional (e não apenas a sub-pessoal) de forma intelectualista.

O intelectualismo é antes de mais visível na teoria dos *qualia* e recordando as origens da teoria dennettiana, não deixa de ser irónico que, apesar do professado wittgensteinianismo, para eliminar a consciência fenomenal Dennett precisamente *reifique* os *qualia*, supostos objectos do reportar interior. A noção dennettiana de *qualia* não é uma boa noção, nem o ponto mais interessante da teoria da consciência apresentada em CE. Não passa uma invenção do próprio Dennett, contra todos os seus professados princípios wittgensteinianos, criada para ser destruída. O que está a ser discutido é o mundo fenomenal, o fluxo fenomenológico e como I. Fox afirma[854] este é o "fim da linha", não requer observador ou conhecimento alguns, e só se os requeresse é que os *qualia* poderiam ser considerados com uma questão relativa a incorrigibilidade.

Uma avaliação final do tratamento da consciência como uma questão de conteúdo e de percepção interior no MEM coloca a questão da seguinte maneira. A percepção interior "intencional" apoia-se, ao nível sub-pessoal, numa ilusão do utilizador. Ora, nada aí obriga a identificar o sentir-se ser ao nível pessoal com uma ilusão do utilizador. O *sentir-se ser* e a *ilusão do utilizador* são questões de âmbito diferente: a ilusão do utilizador, como a virtualidade, é um princípio cognitivo, mecânico, uma hipótese acerca de *como* acontece a vida mental unificada num sistema consciente, não uma hipótese acerca do *que* ela é. Quanto ao *que* ela é, nada obriga a considerar ilusória a experiência de pensar e sentir. A única ilusão seria supor o eu adstrito à experiência como substancial e separado, fora do mundo. No entanto, nada de semelhante decorre do que no início do presente capítulo se afirma acerca de intencionalidade intrínseca, como nada de semelhante decorre da realidade de um sentimento de si. Em suma, o grande desafio que se coloca a quem deseje partir do que está correcto nas análises de Dennett *é não negar o sentir-se ser*. Os argumentos que conduzem à eliminação dos *qualia*, à declaração da impossibilidade de *zombies* e de inversão dos *qualia*, a declarar que o caso de Mary é inimaginável, que a visão cega diz respeito à riqueza da informação disponível, e portanto finalmente ao encerramento do Teatro Cartesiano, não são suficientes para fazer desaparecer um interior de si dos sistemas conscientes. No estrito

[854] FOX 1994.

âmbito do MEM, tais argumentos não afastam sequer um Teatro Cartesiano impreciso. A verdadeira questão é saber se esse interior pode legitimamente ser concebido, ao nível intencional, da forma totalmente intelectualista que Dennett propõe, ou se pelo contrário se deve começar por admitir (utilizando os termos de Dretske) uma *distinção entre percepção epistémica e não epistémica* para conceber o estatuto da percepção conteúdada interna.

Aquilo que substitui na interioridade tal como Dennett a caracteriza a apresentação ou aparição usualmente correspondente à consciência fenomenal e aos *qualia* é a memória de conteúdos, a auto-referência e os relatos acerca do funcionamento cognitivo próprio. De acordo com o MEM, é essa e apenas essa a natureza da nossa interioridade. Isto significa que o apercebimento da interioridade própria em cada entidade consciente seria uma questão imediatamente cognitiva e conceptual. Como se viu no Capítulo 1, já R. Rorty criticava à primeira abordagem da consciência feita por Dennett em C&C o facto de esta amalgamar a capacidade humana de aperceber linguisticamente com a capacidade de fazer auto-relatos incorrigíveis, afirmando em seguida que assim explicava a consciência. Ora também o MEM identifica a consciência com a introspecção linguística de um sistema que produz relatos do seu mundo nocional. No entanto, tudo o que é acertadamente afirmado através da ideia de mundo nocional na teoria do conteúdo (a oposição à ideia de Racionalismo do Significado, ao sentencialismo fodoriano, à possibilidade de medidas proposicionais reais e determinadas de conteúdo, à distinção entre crenças *de re* e *de dicto*[855]) é enganosamente transposto para a teoria da consciência sob a forma de uma espécie de intelectualismo cientista.

É este aspecto do rebatimento do problema da consciência sobre o problema do *conteúdo* que converte o MEM num injustificado absolutismo da terceira pessoa, restringido aos problemas simples da consciência, na terminologia de Chalmers. Este rebatimento dá-se não no sentido em que o MEM pressupõe a TSI, o que é perfeitamente admissível, mas no sentido em que a "intencionalização" da consciência se transforma em intelectualismo. O mundo nocional ou heterofenomenológico que é o *explanandum* da teoria dennettiana da consciência não tem que ser concebido como constituído pelas crenças explícitas do sistema. A "deficiência" epistemológica das aportações do acesso privilegiado é perfeitamente indiferente, não anula a possibilidade de a percepção interior "dar" uma percepção não epistémica, um sentir-se ser .

[855] Cf. Capítulo 2, 2.2.2.

Temos portanto até aqui ideias funcionalistas deflacionárias, que nada obriga que sejam acompanhadas por um desaparecimento da experiência interior ao nível pessoal. Uma outra questão é o facto de a virtualidade da "forma da consciência" ser perfeitamente compatível com (e até apoiar) uma posição epifenomenista. De facto, o que separa a teoria da consciência de Dennett do epifenomenismo a não ser um rápido afastamento devido ao absurdo da posição[856]? A menorização do papel da consciência no controlo da acção está já presente no modelo de BS: a consciência, componente responsável pela "publicitação", não passa de um "porta-voz", fora do "círculo de decisão". No MEM este estatuto é ainda mais claro. Não se entrará no entanto aqui numa discussão acerca de epifenomenismo uma vez que não estão em jogo elementos suficientes. Se no entanto se retiver um sentido mínimo de epifenomenismo como não "intervenção" causal da consciência na efectuação do controlo por um centro unificado num sistema cognitivo, o MEM tem qualquer coisa a dizer, e aquilo que o MEM tem a dizer vai no sentido do epifenomenismo.

Sublinhe-se de novo, para finalizar, que as contribuições mais importantes do MEM para a teoria da subjectividade são ideias do âmbito da teoria da cognição, ideias acerca de agência, de intenção, de unificação, de auto-referência e auto-apercebimento, de realidade e presença virtual de conteúdos mentais no contexto do processamento paralelo e distribuído e não acerca do nível pessoal numa teoria intencionalizante da consciência. Procurar-se-á listar esquematicamente essas contribuições:

(1) Uma ideia acerca das condições preparatórias da criação de uma unidade representada explícita num sistema cognitivo físico: estas correspondem à necessidade de uma melhor comunicação interna de informação, a qual conduz à criação de hábitos de auto-manipulação e auto-exortação no sistema.

(2) Uma ideia acerca do que cria, em parte, o sentido de posse da fenomenologia pelo Eu e da intenção e iniciação unificadas: a existência de uma arena centralizada virtual e temporária para o con-

[856] De facto, Dennett não argumenta contra o epifenomenismo em CE, apenas o considera uma espécie de dualismo imaterialista intestável, insuficientemente sério de um ponto de vista naturalista.

trolo, num sistema "naturalmente" múltiplo ao nível sub-pessoal. O Componente Controlo, central no modelo de consciência de BS é substituído no MEM por agenciamentos mecânicos semi-independentes e semi-inteligentes que poderão depois ser dispensados. Assim, apenas o apoio na unidade virtual do Eu possibilita a existência de um sentido de controlo, autoria, produção.

(3) Uma ideia acerca do estuto da unificação do fluxo fenomenológico: o fluxo fenomenológico poderá consistir na ilusão do utilizador de uma Máquina Virtual instalada num cérebro, o cérebro do corpo a que o sistema cognitivo "chama" seu. Isto significa não que a consciência é ilusória mas que a forma do fluxo, nomeadamente a sua unificação, é virtual.

(4) Uma ideia acerca da ficcionalidade da Identidade Pessoal e da des-unidade sempre subjacente à representação de unidade instituída que é o Eu. As várias desunificações do Eu que foram sendo exploradas pelos teóricos da mente (desde o cérebro dividido até à Desordem de Personalidade Múltipla, para não mencionar os casos mais clássicos de esquizofrenias, paranóias, etc) deixam evidentemente em aberto o problema das funções morais e metafísicas associadas ao conceito tradicional de um Eu unificado prévio, uma vez que elas parecem ficar de algum modo estilhaçadas e impedidas por considerações funcionais.

(5) Uma ideia acerca de experiência e negligência, da importância do mascarar informacional de interrupções. O princípio geral é que quando não existem "agentes ansiosos por informação" mesmo que a informação não chegue não há "reclamação". Se, como dizia Dennett, o ponto cego, na visão, é um caso de negligência sem problemas, um caso em que todas as pessoas com visão normal sofrem de anosognosia, i.e. não se apercebem do seu próprio défice, o princípios da presença virtual e do intelectualismo generalizam uma tal "negligência" ou anosognosia à forma como os conteúdos mentais são conscientes. O MEM tem, sublinhe-se, absoluta necessidade desta ideia de negligência que acompanha a substituição da Testemunha ou Observador pelos especialistas ou agentes que procuram apenas o seu próprio objecto e nada mais.

(6) Uma ideia acerca do estatuto da introspecção de que somos capazes: a Ilusão da imanência. Nós, os "utilizadores" dos nossos próprios cérebros, não somos capazes de distinguir por introspecção entre "o que sempre esteve lá" (presencialidade preenchida dos conteúdos "actualmente" conscientes, memórias estabelecidas e preen-

chidas quando são "chamadas") e "o que é preciso ir buscar" quando uma "pergunta" é feita aos agentes. Pensar que seríamos capazes de tal distinção é aquilo a que M. Minsky chama a ilusão da imanência. Dennett generaliza este estatuto à consciência de conteúdos.

Se estas são contribuições importantes, no entanto e como acontece com a maioria dos modelos em ciência cognitiva, o MEM não inclui uma explicação das razões por que os processos descritos deveriam originar subjectividade nem uma análise da natureza especial dessa subjectividade ou do seu lugar no mundo[857]. O autor do MEM consideraria, é certo, que tais objecções incorrem numa petição de princípio.

Ao criticar alguns aspectos do rebatimento da teoria da consciência num modelo cognitivo e numa teoria do conteúdo não se pretende negar que inevitavelmente qualquer teoria da consciência encontre as questões do conteúdo. No entanto, na obra de Dennett esse encontro foi exagerado nas suas dimensões, de uma forma que, ironicamente, quase faz recair a teoria dennettiana do conteúdo na teoria fodoriana. De facto, se o mundo heterofenomenológico do MEM é o mundo nocional da TSI, no âmbito do MEM prevê-se algo de muito estranho no quadro da TSI: a possibilidade de um teste do mundo heterofenomenológico por mapeamento cerebral (*brain mapping*) e a possibilidade de que pareça existir fenomenologia sem que exista fenomenologia alguma, por uma falta de correspondência de elementos do mundo heterofenomenológico com eventos no cérebro. Esta ideia parece totalmente estranha à teorização do conteúdo feita a partir da EI, a qual culmina em princípios a que se só pode chamar fenomenológicos e que distinguem de forma clara os veículos da representação daquilo que é "representado" na representação. Aliás, a partir do que até aqui foi dito é claro por que a teoria do Eu deve ser considerada central na teoria dennettiana da consciência. Ora, o Eu só pode ser concebido, de acordo com a própria TSI, como uma noção menos fundamental do que as noções de intencionalidade e intérprete. Aliás, várias noções essenciais do MEM (tais como agentes, edições, revisões, fixações de conteúdo) dependem tanto da EI como as representações e computações de que se falava no Capítulo 2. Em última análise todo o funcionalismo depende da EI e portanto do intérprete

[857] Não se pretende de modo algum afirmar que tais ausências constituem uma deficiência dos modelos cognitivos da consciência. Apenas se admite que existe um problema para além da elaboração dos modelos cognitivos que é aquele a que se tem chamado o problema metafísico da consciência.

e da subjectividade, no sentido explorado no primeiro ponto do presente capítulo.

Na teoria da consciência como na teoria do conteúdo aquilo que está finalmente em causa é a eliminabilidade ou não eliminabilidade do subjectivo. Enquanto teórico do conteúdo Dennett aproxima-se muitas vezes do reconhecimento explícito da ineliminabilidade do subjectivo, nomeadamente nas críticas que endereça a outros autores. Fá-lo por exemplo quando considera o que está certo e o que está errado no materialismo eliminativo, a posição defendida não apenas pelos Churchland mas também, em tempos, por um autor cujas ideias acerca de incorrigibilidade, como Dennett reconhece hoje, tiveram uma grande influência no seu pensamento acerca da consciência, Richard Rorty. Como se referiu no Capítulo 1, Rorty defendeu num conjunto de artigos dos anos 70 que o que torna uma entidade mental não é o facto de ela ser explicativa do comportamento nem o facto de ser ou não ser propriedade de uma entidade física mas apenas o facto de certos relatos acerca da sua existência terem um estatuto de incorrigibilidade. Rorty notava ao mesmo tempo que essa incorrigibilidade não era infalibilidade, que os relatos [de um dado sistema cognitivo] acerca do que se passa consigo podem perfeitamente estar errados, simplesmente as pretensões ao conhecimento não podem ser superadas (*overridden*). Ora, Rorty passa desta caracterização da incorrigibilidade para um materialismo eliminativo acerca do mental de acordo com o qual o conhecimento em terceira pessoa pode sobrepôr-se ao conhecimento em primeira pessoa dos estados mentais próprios. É precisamente uma tal posição, a que chama "*the Churchlandish alternative*"[858], que Dennett não aceita. Pensar que as pessoas poderiam explicar-se e ao seu comportamento referindo apenas estados cerebrais permitiria que a linguagem mental caisse em desuso. Isso é inconcebível para Dennett, que pensa que não poderá haver pessoas (com toda a abertura do conceito quanto aos "materiais" físicos subjacentes) tais que tenham deixado de utilizar linguagem mentalista[859]. Precisamente, o materialismo eliminativo elimina totalmente a subjectividade e Dennett recusa-se a admitir que considerações acerca de incorrigibilidade possam conduzir a

[858] DENNETT 2000.

[859] Para o desenvolvimento desta ideia sob a forma de uma história de seres que são em tudo semelhantes aos humanos mas que não sabem, ou melhor, que não pensam, que têm mentes, cf. RORTY 1988, Capítulo 2, Pessoas sem Mentes.

uma tal eliminação. Paradoxalmente, um equívoco análogo aparece na sua teoria da consciência[860]. J. Searle[861] faz uma observação especialmente certeira quando nota que a posição verificacionista da qual Dennett nunca abdica a partir do momento em que faz surgir a teoria da consciência na continuidade da teoria do conteúdo é um erro grave sobretudo na medida em que o conduz a *confundir o sentido epistemológico de subjectividade com o sentido ontológico*. Quando se trata de "sentir-se ser" trata-se de qualquer coisa de ontologicamente subjectivo, mesmo que na continuidade de apercebimentos conteúdados não naturalmente centrados ou unificados, e essa é uma questão diferente da questão relativa ao conhecimento mais ou menos objectivo que dessa coisa se pode ter. Toda a parte desconstrutiva da teoria dennettiana da consciência ao longo do período analisado neste trabalho assenta na confusão localizada por Searle entre subjectividade no sentido epistemológico e subjectividade no sentido ontológico. Só assim, sendo o auto-apercebimento concebido como uma questão de conhecimento e de justificação, o verificacionismo, entendido como o princípio segundo o qual onde não existe *evidência* que permita decidir uma questão, não existe questão, pode operar. Ora Dennett pode perfeitamente estar certo quando defende que não existe aparição pura, a-conceptual, nem momento absoluto nem lugar único da consciência no cérebro, nem distinção nítida entre consciência e não-consciência, entre estar ou não estar no palco do Teatro Cartesiano, ser ou não conteúdo "aparecido", sem que daí decorra a inexistência de algo de ontológicamente subjectivo e de entidades que se sentem ser. Mais: sentir-se ser não é uma característica tal que qualquer entidade com essa característica deva provar a si própria que é assim consigo. Noutras palavras, a questão acerca da consciência própria não é uma questão acerca da justificação de um saber. Não há razão para excluir uma aportação do sentido interno ou um tipo de saber introspectivo que não permita justificação, que seja algo como uma percepção não epistémica. Quando uma entidade se sente ser trata-se mais propriamente de *ter a certeza* do que de saber algo de específico. Nos termos utilizados na listagem de argumentos (apresentada no ponto 3.1.3 do presente trabalho) que sustentam o caso oposto ao de Dennett, trata-se de admitir a *assimetria epistémica* entre

[860] O erro de Dennett na sua teoria dos *qualia* pode ser assim expresso: «La critique épistémologique du cogito ne s'étend pas à la conscience phénoménologique de soi» (GIL 2000: 33).

[861] SEARLE 1992.

o conhecimento do facto de se ser consciente e o conhecimento de todos os outros factos e de admitir a *impossibilidade de uma análise não circular* daquilo que é "experiência".

Uma vez que apesar de tudo os modelos dennettianos de consciência são, tal como outros modelos psicológicos, modelos funcionalistas, terminar-se-á apontando duas salvaguardas quanto ao funcionalismo que Dennett não deixa de considerar. Por um lado, e embora continue a considerar a caixologia (*boxology*) funcionalista (i.e. a decomposição por funções independentemente do significado anatómico ou físico destas) como uma boa táctica para a teoria da cognição, Dennett nomeia o perigo dessa "caixologia funcionalista". Ela arrisca-se a ocultar decomposições alternativas de funções bem como a existência de funções múltiplas sobrepostas. Por outro lado, a possibilidade de os materiais importarem para a consciência de um sistema põe em causa a "pureza" do funcionalismo: para Dennett é claro que não há mente cognitivista pura tal que receba input do corpo e forneça output para o corpo preservando a independência e a separação de uma natureza simbólica, sintática. De resto, o próprio anti-representacionismo da TSI aponta no sentido da incorporação (*embodiment*) e do embebimento (*embeddedness*), por oposição a um nível simbólico autónomo para a explicação da mentalidade de um sistema. Quer o anti-representacionismo quer a continuidade entre *design* e representação (que é o núcleo do teleofuncionalismo considerado no Capítulo 2 deste trabalho) conduzem a uma ideia de inteligência ou mentalidade *incorporada*. Será de resto porque a inteligência está incorporada no desenho das partes de sistemas (globalmente) inteligentes que não tem que haver um centro no qual a iniciação da inteligência ocorra. É certo que com este tipo de afirmações, notórias por exemplo em *Kinds of Minds*, Dennett parece estar a concordar com a grande crítica que A. Damásio faz ao MEM, segundo a qual este seria a teoria de uma *selfless mind*, visando apenas um dos dois aspectos envolvidos no problema da consciência, aquele a que Damásio chama o problema do filme dentro da cabeça, i.e. o fluxo fenomenológico, deixando de fora o sentido de pertença do filme, a "propriedade" da fenomenologia, que seria relativa à posse de si mental por um corpo particular no mundo. Fica em aberto a possibilidade de a teoria de uma *selfless mind* ser tudo o que princípos funcionalistas (em última análise dualistas, recorde-se...) alguma vez poderão atingir.

Uma observação quanto às diferenças entre a consciência humana e a mentalidade animal impõe-se: não restam dúvidas de que os funcionamentos específicos ligados à consciência descritos no MEM dizem respeito apenas aos humanos (ou ao que quer que seja de identicamente sofisticado do ponto de vista funcional). Segundo Dennett animais não humanos não estão sujeitos à ilusão do Teatro Cartesiano. A ilusão do utilizador não existe noutras espécies e nem mesmo nos recém-nascidos humanos. Sendo todas estas mentes muito mais descentralizadas do que as mentes humanas adultas, elas não são (auto-)conscientes no sentido do MEM. Dennett defende que aquilo que tem que existir para além da mera discriminação para contar como consciência não é ubíquo nos organismos usualmente contados como `sensientes´. Como afirmei anteriormente, existe uma agenda oculta na aproximação de Dennett ao problema das mentes animais e nomeadamente à etologia cognitiva: o caso das mentes animais reforça o anti-sentencialismo e a secundaridade da linguagem preconizados na TSI bem como a indispensabilidade da linguagem à consciência. Assim, aquilo a que o MEM chama consciência é apenas humano, linguisticamente possibilitado e explicitamente auto-referencial.

Conclusão

Vale a pena fazer filosofia da mente? Um retorno às origens.

Terminar-se-á este trabalho retomando a questão das origens e procurando explicitá-la. O retorno às origens tem vários sentidos. Um primeiro sentido é o retorno às questões históricas e sociológicas relativas a uma ciência do mental, aí incluindo o problema do posicionamento da filosofia da mente no seio da filosofia e da filosofia analítica no seio da filosofia contemporânea. Um segundo sentido é o retorno às origens da TSI, cobrindo o campo da filosofia em geral, embora seleccionando os dois filósofos, Quine e Putnam, cujo pensamento mais directamente influenciou Dennett. É esta a altura de verificar, nomeadamente, em que sentido se pode afirmar, como se fez na Introdução, que a filosofia da mente traz consigo uma renovação de problemas clássicos quanto à natureza do mundo e ao lugar do espírito no mundo[862]. Tais problemas foram extremamente prementes na filosofia moderna, tendo sido objecto de discussão acesa sob a forma de procura de alternativas para o imaterialismo cartesiano. Tratava-se então não apenas de problemas ontológicos e gnoseológicos como também de problemas metodológicos (pense-se na alternativa quanto à forma de conhecer o mental que as propostas de Descartes e de Hume quanto à teoria da mente representaram: de um lado a intuição, do outro o rumo naturalista). Nesta altura não restam dúvidas quanto à dominância do rumo naturalista em grande parte da filosofia da mente contemporânea. No entanto, muitas das discussões dos filósofos modernos persistem no interior do próprio rumo naturalista e especificamente nas teorias do conteúdo e da consciência. São disso

[862] Utilizar-se-á a partir deste momento o termo "espírito" nos contextos em que se tem utilizado os termos "mente" e "mental" de modo a mais facilmente pôr em relevo a semelhança dos problemas contemporâneos com os problemas clássicos.

exemplo o confronto entre "racionalismo" e "empirismo" na discussão acerca do conteúdo do espírito ou o confronto entre aqueles que defendem que a actividade mental é intrínsecamente consciente de si mesma, ligando todo o mental à auto-consciência (como fazia Descartes, como faz hoje por exemplo J. Searle) e aqueles que não estabelecem um corte absoluto entre a auto-consciência e outros tipos de apercebimento (era por exemplo o caso de Leibniz, é hoje o caso de Dennett). Outros exemplos da persistência dos problemas clássicos são a questão da intervenção do espírito na matéria através da acção, a diferença (kantiana) entre sujeito transcendental e sujeito empírico, a diferença entre a unidade pela intencionalidade e a intuição do sujeito enquanto objecto, a questão, também ela por exemplo kantiana, da realidade de funções ou finalidades na natureza, ou a questão do sentido totalizante da ideia de mundo (nomeadamente quando se discute posições metafísicas acerca da natureza do mundo). De facto, e não olhando às naturais diferenças terminológicas, todos os problemas mencionados foram abordados no presente trabalho.

Um terceiro sentido do retorno às origens é o retorno à nossa própria vida mental. A filosofia da mente visa questões elementares acerca da vida mental de cada um de nós, muito embora, na medida em que ela tem a pretensão de ser filosofia da ciência, metateoria das explicações sub-pessoais (neurobiológicas, psicológicas, etc), as suas respostas não devam ser consideradas como apenas "análises conceptuais". Essas perguntas elementares são por exemplo: o que é pensar, recordar, desejar, sentir-se ser? O que é experimentar emoções, ser activo e passivo? O que é agir, decidir, ser uma pessoa? O que é continuar a ser uma mesma pessoa e como se sabe tal coisa? O que é fazer aquilo que se quer? O que é ser racional? Nem todas estas questões foram tratadas no presente trabalho mas todas são tratadas na filosofia da mente. O retorno às origens significa assim um retorno ao interesse que temos pela nossa própria natureza, obrigando-nos a encarar a possível existência de limites à revisão da nossa auto-concepção enquanto seres mentais com determinadas características (seres com uma vida mental própria, centrada, e sentida como presencial, contínua e determinada). Está em causa saber se alguma vez será possível combinar a *perspectiva a partir de dentro que cada um de nós tem de si*, com o sentimento e a compreensão inerentes a essa perspectiva, *com uma perspectiva exteriorista*, susceptível de nos afirmar "tu és isto", por exemplo numa situação

em que olhamos para um écran de computador e vemos uma imagem do nosso próprio cérebro. Em última análise está em causa saber se, quando se trata da nossa natureza mental, alguma vez poderemos conhecer tudo o que é "real".

A partir da exploração destes pontos procurar-se-á sugerir que são várias as razões por que vale a pena fazer filosofia da mente.

Contexto histórico e sociológico da filosofia da mente e da ciência cognitiva.

Olhando para o contexto histórico e sociológico das relações da filosofia da mente com a ciência cognitiva nos últimos trinta e cinco anos e tomando as discussões apresentadas neste trabalho como um sintoma, verifica-se que a filosofia da mente representa um retorno da filosofia epistemologicamente vocacionada e ontológica por contraste com percursos culturalistas mais ou menos relativistas, niilistas e estetizantes, que pelo menos na filosofia a que os filósofos analíticos chamam "continental" tiveram uma enorme importância no mesmo período. No espaço filosófico americano ao qual este trabalho em grande medida se reporta tais discursos parecem ter encontrado hoje o seu lugar académico nos departamentos de estudos culturais e literários mais do que na filosofia. Por outro lado, no interior da própria filosofia analítica, o campo tratado neste trabalho é ele mesmo sintoma de uma viragem, mais especificamente uma viragem da filosofia da linguagem para a filosofia da mente[863]. As razões para essa viragem foram várias, mas não foi certamente indiferente a ascensão académica da ciência cognitiva e a relevância teórica dos progressos desta. Ora, curiosamente, dada a aversão à metafísica de tanta filosofia analítica nas primeiras décadas do século vinte, continuada em muitas das práticas da filosofia da linguagem mais próximas no tempo, a filosofia da mente trouxe consigo questões ontológicas e metafísicas. Os problemas da filosofia da mente não são em primeira ou em última análise problemas de linguagem, terapeuticamente sanáveis, mas essencialmente

[863] «Gradually but unmistakably in the latter part of the 1970s the philosophy of language lost its place as the dominant starting point for philosophical activity. No other area of philosophy assumed quite the status that the philosophy of language had had since the 1950s. But the degree of interest in relatively pure philosophy of language has certainly diminished. Moreover, there has been a perceptible shift of ferment toward issues in the philosophy of mind» (BURGE 1992: 27).

problemas acerca do posicionamento do espírito no mundo, tratados a partir da ideia mínima da metafísica materialista, a ideia da superveniência do mental ao físico, segundo a qual não há diferenças mentais sem diferenças físicas[864]. Esta ideia é precisamente metafísica e pode ser desenvolvida de muitas maneiras. A TSI é apenas uma delas.

Como se afirmou na Introdução, este trabalho situa-se em grande medida no âmbito da filosofia americana pós-quineana, na qual o imperativo da naturalização da epistemologia, entendido antes de mais como o preconizar de uma continuidade científico-filosófica do inquérito e como oposição ao apriorismo, se generalizou. Ora, a naturalização da epistemologia conduz a um tipo de filosofia que não pode ser estritamente *analítico*: o primeiro sentido, metodológico, de "analítico" diz respeito ao esclarecimento conceptual apriorista do uso de linguagem, pressupondo que uma estrutura profunda e verdadeira da linguagem e da realidade seria alcançadas mediante tal análise ou então tomando o uso comum como referência e solo último. A naturalização da epistemologia opõe-se a este tipo de metodologia conceptual, com o compromisso apriorista que ela envolve. De acordo com uma leitura possível da epistemologia naturalizada, nem o uso comum da linguagem é uma referência importante do inquérito nem existem domínios teóricos imunes à revisão. Do ponto de vista de um defensor da epistemologia naturalizada, a análise impõe, de forma estéril, fronteiras artificiais ao inquérito. Por outro lado e ainda relativamente à naturalização da epistemologia, no caso da TSI e de um modo não muito quineano, ela conduz à conclusão segundo a qual é impossível uma teoria exclusivamente naturalista, causal, não histórica, da subjectividade e da representação, devido à presença de uma dimensão ontológica de constituição. Esta conclusão não pretende ser de modo nenhum o princípio de um argumento no sentido de afastar, ou considerar insignificante na abordagem da cognição, o naturalismo metodológico proposto por Quine. Na Introdução afirmava-se que o naturalismo metodológico acompanhava naturalmente a nova inocência ganha após um período de divórcio entre a filosofia e as ciências naturais. Mas talvez ele seja mais do que isso: talvez seja um dos responsáveis por essa nova inocência bem como por uma grande

[864] Cf. PINTO 1999: 59 para a defesa da ideia segundo a qual é a utilização da superveniência para a especificação do materialismo que o torna uma tese de natureza *lógica*, não pelo facto de ser uma tese deduzida mas pelo facto de não recorrer a leis ou entidades específicas de uma particular ciência.

difusão de um (bom) espírito científico em filosofia, entendido como prática do inquérito racional em comunidade, tornando as discussões obviamente técnicas mas também controladas.

Apesar de tudo, por várias razões, e mesmo após os enormes progressos recentes da ciência cognitiva, uma explicação científica da subjectividade continua a parecer impossível. A questão da "explicação científica da subjectividade" tem aliás todo o aspecto de estar mal colocada, antes de mais porque acarreta a colocação da subjectividade cognoscente numa posição a que esta precisamente não tem direito, uma posição de domínio e de supervisão de si própria. É certo que qualquer explicação passa pela subjectividade. Na ciência da cognição a subjectividade explica e eventualmente explica-se a si própria sem no entanto se erguer em supra-subjectividade. Por isso mesmo, o mínimo que se pode dizer acerca do temor "humanista" da explicação da subjectividade é que há nele uma ilusão de poder. Se por explicação científica da subjectividade se entende a ciência natural dos suportes físicos da cognição (da racionalidade, da consciência, etc) esta já está em marcha há muito tempo e conduzirá certamente à possibilidade de controlo sobre aspectos de seres mentais e de vidas mentais. Em função de quê se negaria isso? Pensando no suporte físico da cognição humana, o sistema nervoso, este não é diferente de outras partes do mundo natural e conhecer é poder controlar (além de ser, em última análise, poder de reconstruir um domínio de objectos). De resto, a própria possibilidade de ser pessoal e moral foi neste trabalho relacionada com o exercício de acção controladora sobre si de determinados sistemas cognitivos físicos. De facto, isso a que se chama subjectividade não está nem nunca esteve desligado da questão do controlo, pelo contrário sempre esteve ligado às variações de actividade e passividade envolvidas no controlo de si. Por outro lado, por tudo o que se afirmou acerca das características do mental e nomeadamente acerca da ligação da intencionalidade à interpretação holista de sistemas, não há razões para esperar que ciências mentais e sociais se desenvolvam de modo exactamente paralelo às ciências físicas ou, mais especificamente, que um género idêntico de previsão venha alguma vez a ser possível[865].

[865] No artigo *Psychology as Philosophy* D. Davidson defende essa ideia: «This does not mean that there are any events that are in themselves undetermined or unpredictable; it is only events described in the vocabulary of thought and action that resist incorporation in a closely deterministic

Assumiu-se desde o início deste trabalho que a questão de uma ciência do mental passava pelas relações da filosofia com a psicologia e pelo esclarecimento da importância da psicologia para a filosofia. A avaliação final possível é a seguinte. A mera decisão de entender como "psicologia" o que todas as disciplinas da cognição fazem é uma decisão filosófica e, como se dizia na Introdução, uma tentativa de resposta à necessidade de esclarecimento conceptual que a convergência de áreas de estudo muito díspares gera, entendendo-as como contribuições para uma teoria da subjectividade, que pelo menos em parte consiste numa análise das condições para atribuir mentalidade a sistemas físicos. Aliás, embora se considere frequentemente que o funcionalismo cognitivista elimina a subjectividade do domínio mental – fazer uma ciência do mental seria afugentar o sujeito da sua própria casa – assistiu-se neste trabalho ao desenvolvimento de uma teoria da subjectividade e nesse sentido a uma reinstauração da subjectividade. Esta reinstauração fez-se, é certo, a partir de uma distinção de dimensões e nomeadamente de uma distinção entre grupos de questões – relativas a interpretação, intencionalidade e racionalidade, relativas a fenomenalidade, apresentação e unificação, relativas a *self*, controlo, identidade pessoal e voluntariedade – deixando aparentemente o sujeito partido (teoricamente) em pedaços. Mas o sujeito desconstruído nunca desapareceu. O quadro traçado acerca das relações entre filosofia e psicologia foi o seguinte: se a filosofia for concebida como tendo como uma das suas tarefas uma metafísica da representação, e se se considerar que a implementação material dos interfaces da representação é de certo modo desprezável na tarefa descritiva dessa metafísica da representação, a psicologia, considerada como engenharia cognitiva, não é demasiado importante para a filosofia. Defender a realidade do mental de um ponto de vista ontológico e analisar o seu estatuto no mundo não é o mesmo que remeter a filosofia à engenharia dos sistemas cognitivos. Nestas circunstâncias, a psicologia não é demasiado importante para a filosofia, ou pelo menos não o é mais do que outros inquéritos científicos disciplinares: a investigação de modelos de mecanismos e veículos da cognição não é necessariamente o melhor caminho para a investigação da natureza do mundo, da possibilidade de verdade no

system. These same events, described in appropriate physical terms, are as amenable to prediction and explanation as any» (DAVIDSON 1980 e: 231).

pensamento e do lugar da (por exemplo nossa) natureza mental no mundo, com as implicações morais e metafísicas inerentes. Segundo a TSI, não ser cientista e psicologista ao fazer filosofia da psicologia traduz-se em encarar a questão da constituição da objectividade para tratar o lugar do mental na natureza, a natureza de interface da representação, por oposição a lidar com categorias como representações e computações como referindo géneros naturais. Este trabalho pretendeu assim ser acerca de psicologia sem ser psicologista. Não se tratou de modelos nem de terapia de indivíduos mas sobretudo de uma investigação acerca daquilo que faz de um sistema físico "psicológico", ou, melhor, mental, e das implicações ontológicas, metafísicas e morais de uma tal natureza. Concluiu-se, é certo, a partir da TSI e do seu anti-representacionismo, que não há géneros naturais específicos que seriam o objecto de uma ciência que seria a psicologia. É evidente, no entanto, que a divisão académica do trabalho continuará sem problemas, bem como a nossa natureza impura, porque psicológica.

A tradição filosófica.

Viu-se no Capítulo 1 que nas origens próximas da TSI na filosofia estão o funcionalismo de Putnam e o fisicalismo e naturalismo de Quine. De facto, curiosamente, foi a partir das posições exterioristas e de certo modo objectivistas sobre o mental defendidas por ambos os autores nos anos 60 que surgiu uma teoria, a TSI, cuja interpretação ontológica se distancia bastante de tais pressupostos. Muito embora no Capítulo 1 tenham sido apontados outros autores (Wittgenstein, Ryle, Anscombe e Taylor, nomeadamente) como estando nas origens da TSI, Putnam e Quine são os mais directos inspiradores do teor epistemológico e ontológico da TSI pelo facto de ambos colocarem a questão do mental no contexto da teoria da ciência. Se é em parte por influência do fisicalismo professado por Quine que Dennett se declara desde o início fisicalista, o que caracteriza o fisicalismo de Quine e o que o distingue do fisicalismo da TSI? Antes de mais, é preciso não esquecer que o fisicalismo de Quine é algo que parece estar "em erosão" nos escritos mais recentes. Isso acontece precisamente devido a problemas com o lugar privilegiado concedido às entidades postas pela física, uma

vez considerada a linguagem matemática imprescindível a esta. Segundo B. Stroud[866], do fisicalismo de Quine só restariam hoje em última análise classes...e nada de material. É o próprio Quine no entanto que considera que quando evoca o fisicalismo o faz apenas como modo de se dissociar de algo cuja entrada não permite na sua filosofia: o dualismo e a semântica mentalista. De facto, em concreto, o ponto fixo da epistemologia quineana é o naturalismo, e é o naturalismo que faz com que «mesmo essa dissociação [o afastamento do dualismo e da semântica mentalista] seja uma questão de ciência falível e não um dogma a priori»[867]. De resto, as mais recentes afirmações quineanas acerca do estatuto do mental numa posição fisicalista vão no sentido de defender o monismo anómalo de Davidson[868] e de manter a ideia de *double standard* proposta em *Word and Object*, a ideia que, como se viu no Capítulo 1, precisamente gerou a TSI. Putnam diria que se trata aqui de mais um sintoma do esquecimento do transcendental na filosofia analítica[869] (entendendo por transcendental a concepção da subjectividade como função de unificação explorada no início do Capítulo 6), da esperança de que seja possível afastar indefinidamente a questão da subjectividade. Mas, curiosamente, esta relação constitutiva entrou de outro modo na TSI e através do próprio Quine. De facto, observou-se neste trabalho que foi a partir da ideia quineana segundo a qual uma teoria da mente só pode ser uma teoria da interpretação que Dennett construiu a TSI acabando por se ver confrontado com o pro-

[866] STROUD 1990.

[867] QUINE 1990: 334.

[868] QUINE 1990: 87. De acordo com um curtíssimo texto relativamente recente sobre estados mentais, o dualismo é obviamente uma «idle redundancy» (QUINE 1985: 5), os estados mentais devem ser adscritos ao corpo, e «the mind goes by the board; it will not be missed» (QUINE 1960: 5). Segundo Quine, não se deve utilizar uma linguagem de acompanhamento dos estados mentais por estados físicos, antes os termos mentalistas devem ser concebidos (*construed*) como referindo estados corporais. No caso de este fisicalismo pouco esforçado ou sem esforço (*effortless physicalism*) parecer demasiado simples e por isso pecaminoso, Quine sugere que convém recordar que os termos mentalistas foram aprendidos pressupondo a referência externa e sintomas publicamente observáveis (embora seja claro que nem todos os estados mentais se manifestam no comportamento, sendo por isso realmente conveniente afirmar que eles são estados nervosos (*states of the nerves*)). No entanto, não faz nenhuma diferença especial identificar estados mentais e estados dos nervos ou simplesmente repudiar estados mentais. De qualquer modo trata-se de conceber o mental como neuronal. Claro que nada do que foi dito pressupõe que seja possível a tradução dos idiomas mentalistas em terminologia neurológica, nomeadamente porque a terminologia mentalista continua a ser «disarmingly uniform» (QUINE 1960: 7) quando comparada com a extrema heterogeneidade dos mecanismos fisiológicos e com o conteúdo empírico. Por todas estas razões, Quine recomenda que se faça o melhor uso possível do monismo anómalo davidsoniano.

[869] PUTNAM 1983: 226.

blema a que Putnam chama o problema do transcendental e que o funcionalismo supostamente afasta[870].

Note-se ainda que, voltando à questão do fisicalismo de Quine, nada nesse fisicalismo garante algo como uma fundamentalidade concedida a algum tipo específico de entidades colocadas pela física. De resto, é preciso reconhecer que foi fundamentalmente o estatuto das *previsões* possibilitadas pelo ponto de vista da física, o alcance geral destas, que conduziu Dennett a conceder uma posição de privilégio à EF na TSI. Se no caso de Quine o que perturba uma possível situação de entidades físicas como "primeiros conceptuais" é o estatuto da matemática, um certo afastamento da referência exclusiva ao conteúdo de teorias físicas específicas na elaboração de posições fisicalistas é visível em vários outros autores que exploram o problema metafísico do fisicalismo, e nomeadamente nos autores a cujas posições se aludiu no Capítulo 3. De facto, a metafísica fisicalista ou materialista não envolve o compromisso com a ideia de que qualquer explicação possível deva ser dada no vocabulário da física[871]. Se a física é considerada fundamental é por nela ocorrer um determinado encerramento explicativo, i.e. por se encontrarem sempre como antecedentes de fenómenos físicos outros fenómenos físicos, não se podendo afirmar o mesmo acerca de fenómenos metereológicos ou, no caso que aqui importa, psicológicos. Se na formulação de uma metafísica fisicalista não se trata de «uma particularmente moderna concepção de matéria, proveniente da investigação mais actualizada em Física Fundamental»[872] é precisamente porque se considera que o problema para a metafísica fisicalista é o problema da superveniência (lógica, natural) dos outros aspectos do mundo sobre aquilo cuja natureza se supõe ser uma e fundamentalmente física. Sem prolongamentos por esta discussão que não constitui o âmbito do presente trabalho, diga-se pelo menos que não é verdade que a única forma de defender uma posição fisicalista acerca do mental corresponda a afirmar que o que existe basicamente é o que a física diz que existe

[870] Putnam interroga-se a propósito: «Why should we expend our mental energy in convincing ourselves that we aren't thinkers (...) that there is no sense in which thought is right or wrong?» (PUTNAM 1983: 246).

[871] Cf. PINTO 1999: 36 e, em geral, o Capítulo 2, *Materialismo e Superveniência I*: listas, reduções e emergências (O materialismo, a Física e o físico, Das reduções à superveniência mereológica, Superveniência e Emergência, Primeira aproximação à conjugação entre o materialismo e a superveniência) e o Capítulo 4, *Materialismo e Superveniência III*: a variedade global.

[872] PINTO 1999: 35.

(muito menos objectos materiais resistentes de tamanho médio) ou que todos os posicionamentos no mundo de aspectos do mundo têm que ser convertidos no vocabulário da física. O núcleo mínimo do fisicalismo é constituído apenas por monismo e fechamento explicativo[873] e praticamente todos os autores aqui tratados, maugrado a diferença das suas posições, o aceitariam. A formulação da superveniência como indiscernibilidade, de acordo com a qual a indiscernibilidade física acarreta indiscernibilidade mental, é compatível com várias posições metafísicas, idênticamente anti-cartesianas[874]. Foi visível neste trabalho que o naturalismo como princípio epistemológico-metodológico é perfeitamente compatível com formulações bastantes diferentes do fisicalismo.

Passando agora a H. Putnam, cuja formulação do funcionalismo foi uma contribuição essencial para a TSI, convém relembrar o que levou Putnam a abandonar e criticar o funcionalismo que ele próprio tinha proposto nos anos 60. Para Putnam a resposta à questão "será que o funcionalismo se sustenta?" depende em grande parte da resposta à questão quineana "será que a epistemologia pode ser naturalizada?"[875]. É por pensar que a epistemologia não pode ser naturalizada que Putnam considera que o funcionalismo não se sustenta. Isto não significa que as ideias de realizabilidade múltipla e de uma certa independência do nível cognitivo relativamente ao substracto sejam renegadas por Putnam[876] mas apenas que não é razoável pretender uma teoria do mental feita exclusivamente em terceira pessoa se se pretende que essa teoria seja abarcante. Para Putnam, as perturbações mais importantes do funcionalismo são as "texturas abertas" das noções de objectos, referência, significado, verdade, justificação e a impossibilidade de conceber essas noções a partir de fora, numa *view from nowhere,* utilizando a expressão nageliana. O problema do funcionalismo filosófico inicial, segundo o

[873] Este último será a contraparte do "fechamento causal do mundo", pelo qual se excluem entidades não-físicas dotadas de poderes causais. Cf. PINTO 1999: 39. Cf. KIM 1996: 12, para o fisicalismo mínimo.

[874] KIM 1996: 10. Cabe notar que a formulação conversa da formulação da superveniência ("não há diferenças mentais sem diferenças físicas") é falsa: criaturas psicologicamente idênticas não têm que ser fisicamente idênticas, com a superveniência apenas se afirma que criaturas não podem ser psicologicamente diferentes e fisicamente idênticas.

[875] PUTNAM 1983 a, *Why Reason Can't Be Naturalized.*

[876] «This much» diz Putnam «still seems to me to be as true and as important as it ever did» (PUTNAM 1988: xii). Como Putnam recorda, tratava-se, na formulação do funcionalismo, de «a reaction against the idea that our matter is more important than our function, than our what is more important than our how» (PUTNAM 1988: xii).

próprio Putnam, foi o facto de este ter sido uma tentativa paracientífica de conceber os fenómenos da intencionalidade, quando esse fenómenos não são causais e dependem da interpretação. Note-se no entanto que a interpretação da TSI feita no Capítulo 6 do presente trabalho se sustenta perante as críticas de Putnam, que apontam sobretudo a impossibilidade de uma *God's Eyeview* sobre os fenómenos da intencionalidade (mesmo uma *God's Eyeview* aparentemente tão neutra como a envolvida na TRM de Fodor). Inclusive responde-lhes admitindo que parte de uma teoria da cognição será uma teoria da constituição da objectividade sendo obviamente esse o contexto no qual as noções-problema de Putnam devem ser retomadas.

A importância das interpretações da TSI com que se iniciou o Capítulo 6 reside no facto de elas justificarem, de um modo mais profundo do que Dennett, ele próprio, o faz, a in-eliminabilidade das descrições intencionais. Em última análise, mais do que uma questão acerca do estatuto das representações mentais no interior dos sistemas cognitivos trata-se, na discussão de tal in-eliminabilidade, da nossa própria realidade enquanto sujeitos e objectos de descrições intencionais. Como diria Nagel, é impossível deixarmos totalmente para trás o nosso ponto de vista sem que deixemos de existir. Se esta é uma observação epistemológica geral, mesmo acerca das áreas que mais se distanciam do "prendimento a si" subjectivo com que todo o pensamento começa, ela torna-se ainda mais óbvia como afirmação epistemológica acerca do estatuto da cognição, observação que de resto a TSI assimila correctamente.

Putnam retoma aliás nas suas críticas actuais aos programas funcionalistas muitos dos pontos defendidos por Dennett desde os anos 60, nomeadamente a ideia segundo a qual o único modo de capturar a noção de "mesma crença" é a prática interpretativa[877] – que é afinal aquilo a que neste trabalho se chamou um funcionalismo enfraquecido por oposição ao funcionalismo fodoriano – devido ao facto de os "estados computacionais" serem tão susceptíveis à objecção da realizabilidade múltipla[878] como os estados mentais

[877] «The only handle we actually have on the notion of same belief is interpretative practice» (PUTNAM 1988: 104).

[878] «What about physical states? The reason for introducing functionalism in the first place was precisely the realization that we are not going to find any physical state (...) that all physically possible believers have to be in to have a given belief, or whatever. But now it emerges that the same is true of computational states. (...) Physically possible sensient beings just come in too many designs» (PUTNAM 1988: 84).

semanticamente caracterizados (esta foi como se sabe a objecção que conduziu o próprio Putnam a propor o funcionalismo como alternativa à teoria da identidade psicofísica).

A partir de Quine e de Putnam chegou-se portanto a uma teoria da mente, a de Dennett, a que se pode chamar uma teoria transcendental e que é ao mesmo tempo um auto-professado "empirismo". Como é isto possível? De facto, sendo acompanhado por um transcendentalismo, o empirismo da TSI é em certa medida "idealista". O que se procurou esclarecer no Capítulo 6 e especificamente nas conclusões acerca de intencionalidade e de racionalidade foi a forma deste "idealismo" a que se chega por um imperativo de naturalização da epistemologia e com um compromisso empirista na teoria da mente. Se o idealismo presume a primazia do espírito e toma o resto como objecto do espírito – como se o espírito "jorrasse" imaterial – não é decididamente disso que se trata. Mas existe outra forma de admitir que o espírito é incontornável, sem lhe conceder tal primazia, uma forma que merece o nome de "fenomenológica".

Se de acordo com o "idealismo" da TSI a mentalidade de sistemas físicos depende de interpretação, onde está o espírito necessário para uma tal interpretação? Em nós? Se se pensar duas vezes, a resposta a esta questão que a TSI possibilita é um monismo anómalo reinventado. O monismo é evidentemente uma aposta metafísica na unidade daquilo que existe: não há em última análise senão acontecimentos físicos. A anomalia diz respeito à explicação, à inexistência de leis e causação psicológicas. Acontecimentos sob descrições semânticas e mentalistas não caem sob leis físicas. Sem presunção de originalidades[879], as concordâncias entre a TSI de Dennett e o monismo anómalo de Davidson (bem como a recomendação deste por Quine) no que respeita à locação do mental no mundo acentuam da importância do *als ob*, do como se, na teoria da mente feita por seres do nosso tipo. É assim perfeitamente possível defender uma posição fisicalista acerca do mental e do significado não olhando para algo como um nível último e fundamental de realidade e sim para uma "macro-física qualitativa" cuja linguagem será a EI, a interpretação. As únicas *explicações* possíveis são causais. No entanto, perante algumas manifestações complexas da causalidade natural, nomeadamente perante "formas aspectuais" como as mentes, são inevitáveis juízos teleológicos, obviamente heurísticos, não explicativos, anómalos, interpretativos. Dir-se-ia que para seres do nosso tipo o dualismo permanece na prática inultrapassável.

Mas, pelo menos, ser mental não é uma diferença específica dos humanos relativamente ao resto do mundo natural. Não há um corte entre a "alma" e o "corpo". Não se trata de duas substâncias de natureza essencialmente diferente. Não é por sermos uma alma simples e imaterial distinta deste mundo que cada um de nós é inteligente, consciente, racional. O problema é que se a única concepção clara e distinta possível da nossa natureza era a concepção de uma natureza pensante, a partir do momento em que se admite que essa natureza pensante não é essencialmente separável deve igualmente admitir-se que não temos qualquer concepção clara e distinta da nossa natureza. Imediatamente, o "idealismo" suposto pela interpretação fica desligado de conotações aprioristas e do conhecimento incorrigível e intuitivo do espírito pelo espírito.

Negar que mentes sejam substâncias simples e imateriais não é negar que sejamos – nós e muitas outras entidades – mentais. O monismo fisicalista e a atitude ontológica natural conduzem apenas a crer que não existem abismos metafísicos entre consciência e não-consciência, entre intencional e não intencional: mesmo numa disputa como a de Dennett com Fodor acerca da existência ou não existência de géneros naturais intencionais está-se dentro da atitu-

[879] Que estariam muito mais do lado de Davidson do que do lado de Dennett. De facto, do mesmo modo que para explicar as origens da TSI Quine e Putnam se impuseram como os dois autores incontornáveis, à medida que este trabalho progredia a pura necessidade dos temas da filosofia de Davidson impôs-se. Na obra de Davidson aparecem a nu, em grande parte devido ao ascetismo de estilo, os temas filosóficos que o presente trabalho visa através de Dennett e das questões mais práticas que ocupam este, nomeadamente as relações entre filosofia da mente, ontologia e teoria da acção. Davidson nomeia claramente os problemas relativos às articulações dos temas dessas áreas, por exemplo o problema colocado pela tradução radical – a chave da teoria da mente – à ontologia. O monismo anómalo, ao contrário da TSI, visa directamente esse problema, em termos de explicação, de superveniência e de ausência de eco na física do holismo que permite o acesso ao mental, indo assim ao âmago da questão que de facto também subjaz à TSI. Os temas "práticos" são igualmente comuns a Davidson e a Dennett, nomeadamente a questão da acção como parte do problema do mental, o estuto das razões como "causas", as relações da filosofia e da psicologia. A vantagem de Dennett é, em contrapartida, a sua maior ligação à ciência cognitiva propriamente dita, em concreto a autores como D. Hofstadter, M. Minsky ou A. Newell, que são uma real inspiração do seu trabalho. Quando comparada com a obra de Davidson, é notório, por isso, que a obra de Dennett visa mais a filosofia da ciência cognitiva do que a filosofia da linguagem e a ontologia, mais a cognição em geral do que a especificidade do pensamento humano, para além de prestar mais atenção às imperfeições não cognitivas da fenomenologia humana. Deste modo, ao longo da elaboração do presente trabalho, Davidson foi aparecendo como uma espécie de *alter ego* de Dennett, mais estritamente filósofo, menos cientista cognitivo. Não convem esquecer no entanto que uma das coisas que se joga no presente trabalho são as razões de uma passagem disciplinar de temas que foram/são filosóficos para a ciência cognitiva, e essa perspectiva lança uma outra luz sobre as vantagens relativas das obras de Davidson e de Dennett.

de ontológica natural, do fisicalismo. O clássico problema mente-corpo volta a colocar-se de forma interessante precisamente neste ponto, quando se pergunta qual é o lugar do mental num mundo essencialmente *físico* (esta última qualificação é metafísica).

Cabe aqui uma outra constatação histórica. Tudo aquilo de que os anti-cartesianos modernos puderam ser acusados (por exemplo de necessitarismo e pampsiquismo, recordando o caso de Espinosa, que tem em comum com grande parte da filosofia da mente contemporânea o seu fundamental monismo) espreita o horizonte do tratamento contemporâneo do problema mente-corpo. É também neste quadro que a oposição também ela moderna entre racionalismo e empirismo ganha uma nova pertinência: por exemplo as conclusões estáveis a que a TSI permite chegar posicionam-na como uma justificação filosófica da corrente anti-representacionista, e portanto "empirista", em ciência cognitiva. O "movimento" ou corrente anti-representacionista em teoria da cognição – que tem como exemplos a abordagem da cognição através de modelos conexionistas, da teoria dos sistemas dinâmicos, ou a robótica de Brooks[880] – opõe-se ao cognitivismo simbólico clássico, cujos princípios foram exemplificados neste trabalho pelo realismo intencional sentencial da TRM. A ideia nuclear do anti-representacionismo é, sublinhe-se de novo, que não há um nível autónomo intrinsecamente representacional nos sistemas físicos.

Vale ou não vale a pena? A nossa natureza mental.

Assim sendo, e verificada a renovação de tantas questões clássicas, talvez demasiado clássicas, tão clássicas que pareceriam ultrapassadas,

[880] Cf. BROOKS 1997: 419: «It is not feasible to identify the seat of intelligence within any system, since intelligence is produced by the interactions of many components. Intelligence can only be determined by the total behavior of the system and how that behavior appears in relation to the environment». Rodney Brooks termina este texto, *Inteligence Without Representation*, um manifesto anti-GOFAI (Good Old Fashioned Artificial Intelligence) e portanto um manifesto da oposição ao cognitivismo simbólico clássico, com uma moral que poderia ser a de Dennett com a TSI: «Intelligence is in the eye of the observer». Se é assim, o que é que está no próprio sistema? Cf. MACHUCO ROSA 2002, para a proposta segundo a qual os sistemas cognitivos são sistemas dinâmicos *interpretáveis* em termos cognitivos. Esta forma de colocar a questão tem consequências práticas directas, nomeadamente relativas aos formalismos matemáticos requeridos pela teoria da cognição. A. Machuco opõe o paradigma simbólico ao conexionismo em função dos seguintes parâmetros: (1) formalismos (equações diferenciais versus estruturas lógico-simbólicas), (2) caracterização da transição entre estados (estatística versus inferência lógica), (3) "situação" das representações no sistema (localização modular versus propriedades distribuídas), etc.

vale ou não vale a pena fazer filosofia da mente? A primeira resposta é que sim, que vale a pena fazer teoria da mente mas ela será em grande parte e indistintamente ciência, ciência cognitiva. A teoria do mental deixou em grande medida de ser filosofia e se se pretender abordar as questões da cognição por meio de um inquérito exclusivamente conceptual corre-se o risco de negar factos. Fazer filosofia da mente é passar pela ciência cognitiva e Dennett – o autor escolhido para guiar o presente trabalho – é disso exemplo. Nesse sentido, a regra de ouro ryleana "Não fazer hipóteses mecânicas sobre o mental" (pelo menos se se é filósofo) caducou. Futuros filósofos ou teóricos do mental trabalharão nas disciplinas da cognição (talvez mesmo exclusivamente, sendo filósofos "ao fim do dia" como os filósofos-cientistas do séculos XVII e XVIII). Há no entanto um certo tipo de trabalho teórico, a que se pode chamar fenomenológico, no sentido atrás nomeado, cuja necessidade na teoria do mental Dennett não deixou de reconhecer e do qual considerou ter encontrado um exemplo em Wittgenstein e em Ryle. Um trabalho fenomenológico não consiste em hipóteses mecânicas sobre os veículos do mental mas sim na descrição de "efeitos", i.e. na descrição, feita a partir de dentro, de como é ter entendimento, compreensão, pensamento, numa postura "recolhedora" ou experimental perante o acontecer do espírito, não visando de resto exclusivamente apenas os aspectos cognitivos deste mas também os aspectos volitivos e afectivos e procurando analisar as relações entre todos eles. Chamo a esse tipo de trabalho teórico acerca do mental um *trabalho fenomenológico* na medida em que ele (1) visa o mental como dando-se, acontecendo, (2) visa o mental como não explicitável intelectualmente de forma exaustiva, (3) visa o mental como sentido, emocional, desejante e não apenas cognitivo. O termo é justificado ainda no sentido científico mais elementar de recolha do ainda não tratado, que não se deixa à partida reconduzir a ou explicar por outra coisa, devendo ser apenas descrito. Este trabalho é anti-cartesiano na medida em que assume que o mental não apenas não é imaterial como não se dá clara e distintamente e quando é levado a cabo na filosofia da mente poderá distanciar-se não apenas do idealismo como da fenomenologia no sentido mais tradicional e escolar do termo, já que grande parte dos filósofos da mente contemporâneos procuram conceber o mental como uma característica geral do mundo, a par de leis e entidades físicas. Perante este tipo de posicionamento, os idealismos e a fenomenologia clássicos estariam demasiado presos à forma humana do mental.

Uma segunda razão por que vale a pena fazer filosofia da mente é a verificação de que não convém utilizar em vão os termos "subjectividade", "racionalidade", "consciência" ou "acção", presumindo saber do que se fala. Chega-se a essa conclusão modesta pelo estrito cumprimento do imperativo quineano da naturalização da epistemologia. Na verdade, este trabalho foi iniciado com a intenção de verificar se a TSI era uma teoria coerente da subjectividade. Ora, o mínimo que se pode concluir após o trajecto é que não é legítimo presumir a priori que se possui A noção de Subjectividade. Ser subjectivo não é condição suficiente para o (auto) conhecimento de algo que se declina de tantas maneiras[881]. Assim como não é legítimo presumir a priori que se possui uma noção precisa de racionalidade não é legítimo presumir que unicamente por sermos de natureza subjectiva possuimos "O conceito de subjectividade".

Uma terceira razão pela qual a filosofia da mente vale a pena, é o facto de, como já se referiu, ela retomar certos problemas clássicos da filosofia, relativos ao lugar do espírito na natureza, que aparecem agora em grande parte sob a forma de problemas de ciência cognitiva (é o caso de problemas como a natureza da racionalidade, da consciência, das imagens mentais, dos conceitos, etc). A filosofia da mente mostra assim que uma determinada concepção aparentemente desactualizada de filosofia, que continuava a incluir esses problemas e a tratar a muito desconstruída subjectividade, não era assim tão extraviada. De facto, ao recusar o tratamento directo de tais problemas, nomeadamente a partir da prática científica, admite-se implicitamente uma solução imaterialista para o problema do espírito, ou pelo menos torna-se menos problemática a aceitação desta, além de se correr o risco de fazer proliferar entidades cuja existência é exaurida pelo aparecimento, entidades ou acontecimentos acerca das quais é impossível afirmar ou saber o que quer que seja, realidades precisamente inefáveis. Repare-se que a época tratada neste trabalho foi uma época em que grande parte da filosofia europeia continental se definia como fraca ou débil, como crítica cultural e civilizacional mais ou menos estetizante, enquanto a filosofia analítica continuou a tratar, por estrita necessidade gerada pelo trabalho em campos científicos (e por um certo alheamento histórico-cultural, é certo, não raro de resto nas práticas científicas), de questões epistemológicas, ontológicas e morais

[881] Cf. GIL 2000, 1 (*Être un sujet*).

fortes. Não é possível sequer afirmar que o fizesse ingenuamente, na ignorância do "fim do homem" e da "consumação da metafísica". Afinal, os tempos continuaram e há aí sociedades cuja especificidade histórica está ligada à informação, ao conhecimento, à realidade virtual, fenómenos que através das ciências cognitivas a filosofia da mente precisamente visa. Essa realidade não desaparece nem as suas especificidades são capturadas com veredictos intelectuais de consumação da metafísica. Nesse sentido poder-se-á perguntar de que lado está o alheamento sócio-histórico-cultural, na filosofia concebida como consistindo quase exclusivamente em crítica cultural e civilizacional, ou no tipo de filosofia de que se tratou neste trabalho. Por exemplo, apenas a falta de atenção da filosofia ao que é conhecido cientificamente sobre a natureza permite considerar esta como pura objectividade, *en soi* massivo, o outro da consciência. Uma tal ideia ilusoriamente objectivista da "natureza" tratada pelas ciências naturais torna-se especialmente estranha perante a ciência cognitiva contemporânea.

Apesar de tudo, algumas das conclusões da filosofia da mente e da filosofia europeia continental recente coincidem, nomeadamente "conclusões quanto à desconstrução da subjectividade", para utilizar a expressão de J.P. Dupuy[882]. Sem a roda livre intelectual frequente na filosofia a que os filósofos analíticos chamam "continental" as ciências cognitivas participam numa desconstrução ou fragmentação do sujeito, na medida em que propõem modelos dos mecanismos físicos que manifestam propriedades de subjectividade, mostrando assim que não há fantasma na máquina, "centro de subjectividade" ou observador desligado, que a singularidade e o controlo centrado do Eu são em grande medida efeitos virtuais, ilusórios. Quando algum Eu pensa não há aí um sujeito homuncular visionando, controlando e iniciando, mas um isso regulado por leis físicas, processos sem sujeito, funcionamentos que têm como um dos seus efeitos esse Eu e os seus pensamentos. A diferença é que a desconstrução aqui em causa foi acompanhada de uma reconstituição mostrando que provavelmente não nos livraremos do sujeito, o que certamente evita as aporias em que os diagnósticos da morte deste normalmente redundam.

Os perigos do cientismo, psicologismo e esterilidade na interpretação filosófica do trabalho científico sobre o mental, apontados

[882] DUPUY 1994 e DUPUY 2000.

na Introdução deste trabalho, evidentemente existem, inclusive nas respostas prontas e nas etiquetas de escola da filosofia americana. Mas também estão lá análises que conduzem a novos problemas específicos que não são de todo estéreis. Poder-se-ia continuar este trabalho (é de regra referi-lo...) retomando e aprofundando as diferenciações feitas no seio do problema da subjectividade. De facto, elas permitem inúmeras questões intersticiais, questões precisas e de minúcia, prolongando o inquérito acerca do sentimento de si e de saber, considerados no sentido fenomenológico referido, e nomeadamente incluindo domínios mais emocionais e afectivos do que aqueles que foram considerados e que de resto a filosofia da mente contemporânea não ignora. Exemplos de tarefas que ficam delineadas e abertas na continuação deste trabalho dizem respeito a questões como a categorização[883], a racionalidade[884], a fenomenologia da agência, do querer, da auto-consciência. Fica aberta a possibilidade de distinguir componentes da subjectividade e da cognição que não "são como nada" a partir de dentro (talvez questões como "como é um estado mental ter conteúdo?", "como é ter o conceito de x?" não tenham resposta ao nível pessoal) dos componentes mais ou menos explicitamente sentidos e reflectidos da subjectividade. Por outro lado, levanta-se a questão das relações entre "aspectos" cognitivos, afectivos e volitivos na subjectividade, incontornável se se pretende elaborar uma teoria da vida mental e do pensamento[885] mais complexa do que aquela, sobretudo cognitiva, visada neste trabalho.

Apesar das semelhanças apontadas entre os problemas da filosofia da mente e certos problemas clássicos da filosofia, a verdade é que a estranha configuração da filosofia da mente é bastante amnésica relativamente à história da filosofia. No entanto o risco

[883] De facto, a questão ficou em suspenso no Capítulo 2, ponto 2.2.3.3. O problema consiste em saber que teoria dos conceitos fará sentido no seio do anti-representacionismo proposto, aceitando as intuições anti-descritivistas acerca da capacidade do espírito humano para formar universais ou ideias. Essas intuições são enquadradas por Fodor na Teoria Representacional da Mente em *Concepts*. É preciso portanto uma obra análoga a *Concepts*, produzida pela corrente anti-representacionista.

[884] Os prolongamentos do "chauvinismo da intencionalidade e da racionalidade" numa teoria do pensamento são tentados por Stich em STICH 1993 (*The Fragmentation of Reason*) e STICH 1996 (*Deconstructing the Mind*).

[885] É isto em grande medida que F. Gil entende como uma *teoria da compreensão*. A necessidade desta surge quando se admite que o entendimento da inteligibilidade, da posse de qualquer conhecimento ou explicação, passa por um inquérito fenomenológico sobre a subjectividade que deve incluir todas as dimensões referidas – cognitivas, volitivas, afectivas – e respectivas relações (cf. GIL 2000, *La Conviction*).

não é grande. A agenda da filosofia da mente é preenchida e específica e, por isso, ou ela revisita a história da filosofia ou a refaz, não por razões de historiografia erudita mas por necessidade do inquérito racional em curso. De qualquer modo, mostra que os temas desta não se tornaram proíbidos por razões epocais de fim da filosofia epistemológica.

Uma quarta e última razão por que vale a pena a filosofia da mente tem simplesmente a ver com o facto de ser importante para nós sabermos de que natureza somos. Nesse sentido, a filosofia da mente traz consigo um renovado interesse pela questão acerca do que é distintamente humano, ao mesmo tempo que estilhaça essa questão. A questão é estilhaçada por um lado porque, como afirma T. Nagel, «Mesmo que as manifestações do mental evidentes para nós sejam locais – elas dependem dos nossos cérebros e de estruturas orgânicas similares – a base geral deste aspecto da realidade não é local, mas deve ser suposta como sendo inerente aos constituintes gerais do universos e às leis que os governam»[886]. Mas as questões analisadas e os critérios analisados e propostos nos Capítulos 4 e 5 deste trabalho dizem respeito ao que torna o humano humano. A partir das diferenciações sugeridas é possível avançar propostas específicas quanto a algumas diferenças existentes por exemplo entre pessoas, cadáveres humanos, animais e robôs, imaginar o que é sobreviver como uma mesma pessoa ou que espécies de continuidade com aquilo que somos nos satisfariam. No entanto o aspecto mais importante das questões abordadas nos Capítulos 4 e 5 é a queda das dicotomias estritas entre natural e artificial[887], pessoas e não-pessoa, acção e não-acção. Questões aparentemente conceptuais como as questões relativas à natureza da subjectividade, racionalidade e pessoalidade dependem de descobertas empíricas e de avanços tecnológicos (o que mais uma vez confirma o ponto metodológico segundo o qual nenhuma análise de conceitos é auto-subsistente, e não há filosofia estritamente analítica de conceitos semelhantes, cuja recolha é antes de mais fenomenológica). Localizou-se, é certo, um limite da particular teoria apresentada sobre o que torna o humano humano, a ausência de tratamento do

[886] NAGEL 1986: 8.

[887] E a pergunta será então "Alguma coisa poderá alguma vez ser *contra natura*?". Na discussão deste tema é a própria noção de *antropologia* que é posta em causa (por ser paroquial, talvez). De facto, faz toda a diferença para aquilo que se entende por psicologia e antropologia a possibilidade de existir cognição artificial.

sentido *sentido* de si e de saber (a que se poderia acrescentar e sentido de querer, de desejar).

Ainda relativamente à especificidade do humano, ou, na terminologia adoptada no presente trabalho, relativamente aos sistemas cognitivos susceptíveis de pessoalidade e de Identidade Pessoal, convém recordar que foi proposto, a partir da TSI, que a pessoalidade do mental não é apenas uma questão relativa às continuidades física e/ou fenomenológica. A caracterização da Identidade Pessoal será incompleta sem aquilo a que se poderia chamar um compromisso com a vida mental própria[888], sem auto-avaliação e sem avaliação da racionalidade própria. Ainda relativamente às aplicações morais da teoria do mental, uma outra questão clássica da filosofia, a vontade livre, foi afastada do âmbito de uma discussão acerca do determinismo físico para o âmbito da interpretação intencional. Este não é evidentemente um movimento novo ou original na história da filosofia: pense-se no atrás referido monismo espinosista, e na ideia de Espinosa segundo a qual conceber a liberdade como ausência de determinação é apenas ignorância. Foi de resto a impossibilidade de explicar a causação mental o maior inimigo da divisão substancial cartesiana e do imaterialismo. Mas o problema da causação mental não desaparece com o desaparecimento do dualismo cartesiano. No caso de Dennett a estratégia adoptada relativamente à vontade livre foi idêntica à estratégia adoptada relativamente à consciência (e poder-se-ia dizer que também à racionalidade): fragmentação. Dennett defende que a vontade livre não é *um* problema, mas vários problemas que se interpenetram com os problemas da consciência e da racionalidade. A vontade livre, aquilo que somos tanto quanto formos o mais que podemos ser, é *auto-controlo meta-reflexivo* – controlo do comportamento próprio e da identidade própria por meios racionais, conseguido por alguns sistemas cognitivos físicos resultantes da evolução existentes num mundo determinista – e *avaliação forte*. Noutras palavras, a liberdade não consiste em "razões" absolutamente determinadas nem explicativas: nada no mundo descrito de forma mentalista tem essas características. Razões de um agente para a acção são selecções em sistemas cuja complexidade escapa em grande medida ao próprio agente. As questões da liberdade (e consequentemente da responsabilidade, da distribuição – nomeadamente legal – da culpa, da punição, do elogio, dos direitos e das obrigações de humanos) são questões de macro-nível, dizem respeito a um particular entendimento ou interpretação de eventos físicos como acções. Não são questões relativas ao *micro-nível* do sistema, explicado pela teoria física. Essa é a posição da TSI perante a parte do problema do mental que é o problema da agência.

[888] Dennett (DENNETT 1994f) aceita as sugestões de Carol Rovane quanto a auto-avaliação e avaliação da racionalidade própria. Como já foi sublinhado anteriormente, ele considera mesmo que Rovane aperfeiçoa as suas ideias acerca de pessoalidade.

Muito do que é caracteristicamente humano é mentalista e tudo o que é mentalista é de macro-nível e de interface, superficial relativamente àquilo que seria por exemplo um conhecimento da estrutura micro-física e insusceptível de mudar de estatuto com o avanço do conhecimento de tal estrutura. A questão da acção coloca-se precisamente tanto quanto nunca poderemos sair completamente para fora de nós próprios, da nossa natureza mentalista. Que importância moral, metafísica, pode ter então o esclarecimento das condições do agir? Se, ao contrário de tudo o que foi até aqui afirmado, existisse um eu naturalmente e essencialmente uno ou uma pessoa "absoluta", o auto-controlo meta-reflexivo através de razões seria redundante. Mas, precisamente, nenhum eu uno pode ser dado como garantia a priori, o Eu é ele próprio um efeito do funcionamento sub-pessoal do sistema e as "pessoas" pressupõem sistemas constituídos por um tal processo de centralização virtual apoiado numa auto-representação global. De acordo com Dennett, apenas o auto-controlo meta-reflexivo representa auto-consciência genuína, que fica assim indiscernivelmente ligada à agência racional. Esta caracterização de pessoa é, evidentemente, anti-essencialista e anti-imediatista. O problema é que a auto-consciência (considerada não apenas como auto-avaliação mas também como compromisso com a racionalidade própria) é assim aparentemente caracterizável como desprovida de conteúdo sentido, de sentimento de si ou de desejo (outra coisa não seria de esperar, depois da teoria da consciência apresentada no Capítulo 3 acerca da consciência fenomenal e do sentimento de si, sendo a pessoalidade de sistemas cognitivos aparentemente analisável de forma exaustiva como uma questão relativa a conhecimento e controlo). Ironicamente, também este hiper-racionalismo é uma posição clássica na teorização da liberdade da vontade, simplesmente desligado agora da pureza cartesiana do espírito.

De modo algum se exclui, pelo facto de a teoria apresentada poder fornecer, a partir de uma filosofia da ciência do mental, cri-

térios de orientação na resposta às questões clássicas enunciadas acerca do que torna o humano humano e acerca da vontade livre, questões tantas vezes chamadas antropológicas, a possibilidade, sublinhada por T. Nagel[889], de haver boas razões para se ser pessimista relativamente ao entendimento de nós mesmos que uma ciência da mente poderá vir a possibilitar. De facto, é mais do que razoável supor que existem limites à forma como podemos pensar em nós próprios. Por exemplo, cada um de nós pensa em si próprio como uma pessoa Una, Única e Singular e esse não é um traço trivial ou periférico da nossa auto-concepção mas o próprio núcleo daquilo que pensamos que somos e não pode ser posto de lado como puramente fenomenal. Assim sendo, «Pode ser que seja impossível para nós abandonarmos certas maneiras de nos concebermos a nós próprios por mais insustentáveis que elas sejam do ponto de vista científico»[890]. Uma das coisas que não podemos abandonar é a unidade da pessoa que pensamos que somos. No entanto essa unidade nada tem de absoluto. Poder-se-ia sugerir o mesmo quanto ao sentido de completude e de determinação dos conteúdos da vida mental, maugrado as afirmações de um modelo como o MEM acerca da negligência normal e da indeterminação: não podemos abandoná-los, a virtualidade e a ilusão de imanência são parte integrante da nossa natureza mental, do nosso conceito mentalista de nós próprios, mesmo se de alguma forma "desaparecem" na caracterização dos sistemas físicos que nós também somos. O desafio colocado pela dobra entre o subjectivo e o objectivo trazida para dentro de nós pelas ciências do mental é precisamente saber em que medida somos ou não somos aquilo que nos apercebemos como sendo, saber que intervalo, que grau de ilusoriedade existe entre isso que apercebemos como a nossa interioridade mental e pessoal e o que quer que na ordem objectiva subjaz às continuidades subjectivas da vida mental a que chamamos nossa e que é objecto de ciências de cognição. A desmontagem de testes da mentalidade genuína como o Quarto Chinês de J. Searle mostra que não podemos confiar na nossa intuição e muito daquilo que ficou dito no presente trabalho indica pontos de integração insatisfatória entre o subjectivo e o objectivo, colocando a reintegração moral e intelectual numa posição complicada.

[889] NAGEL 1979d, *Brain Bissection and the Unity of Consciousness.*
[890] NAGEL 1979d: 39.

Uma última palavra acerca da dimensão moral da questão da mente. Quando se trata do mental como auto-consciência e auto-agenciamento trata-se daquilo que nós somos e portanto a questão importa-nos, o que talvez justifique um certo sortilégio da dimensão psicológica[891] quando pensamos sobre o pensamento. É verdade que de acordo com muitas das ideias analisadas neste trabalho nós somos – enquanto agentes individuais centralizados e auto-conscientes, enquanto seres mentais e pessoais com uma vida mental presencial e preenchida –de uma natureza bastante superficial, muito menos fundamentais e simples do que aquilo que uma concepção imaterialista de mente como alma e pensamento fortemente individuais nos permitiria supor. No entanto o risco, referido por C. Taylor, de não ficar assim ninguém para ser responsável por nada não é o diagnóstico correcto da situação em que uma ciência do mental nos coloca. Identidade Pessoal, responsabilidade e acção são noções inauguradas num mesmo movimento: não seria possível por exemplo manter a noção de responsabilidade e perder o eu. Não é por outro lado insignificante o facto de essa natureza mental nos permitir conceber a nossa própria natureza, estruturada em torno de um eu, suposto autor de uma vida mental racional e consciente, e portanto o nível ao qual a nossa realidade pessoal se situa, bem como aquilo que provavelmente é mais fundamental do que nós.

[891] Não no sentido ainda agora referido de mecanismos sub-pessoais da cognição mas no sentido de vida sensorial, emocional, sentimental, etc, que é própria, individual.

Bibliografia

Nesta bibliografia, tal como nas notas ao texto principal, utiliza-se o sistema autor (nome maiusculado) / data. Quando existe uma divergência entre a data do primeiro aparecimento de uma obra ou de um artigo e a data da edição efectivamente utilizada, aquela data aparece entre parênteses rectos a seguir à data da edição utilizada. Sempre que se considera ser de interesse histórico tal procedimento, a referência no corpo do texto é feita à data da primeira publicação. A referência completa que aparece na presente bibliografia é no entanto a da edição utilizada.

Para além de uma listagem tanto quanto possível completa dos escritos de D. Dennett, a presente bibliografia inclui ainda um número considerável de referências a determinados autores sem os quais a intenção mais alargada do trabalho não seria concebível, com especial relevo para W. V. Quine, H. Putnam, D. Davidson e J. Fodor. Devido à intenção em parte histórica do trabalho, certas referências que estiveram na origem do interesse pelas relações entre a filosofia e a ciência cognitiva aparecem aqui incluídas, mesmo não sendo directamente citadas no corpo do texto.

I – *Obras de D. Dennett*[1]*:*

I.I *Livros:*

DENNETT, Daniel, 1986 [1969], *Content and Consciousness*, London, Routledge and Kegan Paul (2ªedição). Tradução italiana: *Contenuto e concienza*, Bologna, Il Mulino, 1992. Tradução espanhola: *Contenido y conciencia*, Barcelona, Gedisa, 1995.

[1] Uma bibliografia actualizada (até 2000) e bastante completa de D. Dennett – que inclui ainda referências a escritos dedicados ao pensamento do autor (números de revistas, obras colectivas, artigos, recensões) — pode ser encontrada em http://sun3.lib.uci.edu/indiv/scctr/philosophy/dennett, um *site* mantido por Eddie Yeghian, Bibliotecário de Filosofia da Universidade da California – Irvine (UCI). Essa bibliografia foi de grande utilidade para o presente trabalho. Optou-se ainda por referir nesta bibliografia algumas traduções dos livros do autor. Para além da versão publicada, muitos dos artigos referidos estão disponíveis *on-line* na página mantida pelo autor em http://ase.tufts.ed/cogstud/pubpage.htm.

DENNETT, Daniel, 1981 [1978], *Brainstorms – Philosophical Essays on Mind and Psychology*, Cambridge MA, MIT Press (edição original: Montgomery VT, Bradford Books). Tradução italiana: *Brainstorms. Saggi filosofici sulla mente e la psicologia*, Milano, Adelphi, 1988.

DENNETT, Daniel, 1983, *Elbow Room – Varieties of Free Will Worth Wanting*, Cambridge MA, MIT Press. Tradução alemã: *Ellenbogenfreiheit: die wunschwerten Formen von freiem Willen*, Frankfurt, Hain bei Athenäum, 1986. Tradução espanhola: *La libertad de acción: un análisis de la exigencia de libre albedrío*, Barcelona, Gedisa, 1992.

DENNETT, Daniel, 1987, *The Intentional Stance*, Cambridge MA, MIT Press. Tradução francesa: *La stratégie de l'interprète*, Paris, Gallimard, 1990. Tradução italiana: *L'atteggiamento intenzionale*, Bologna, Il Mulino, 1993. Tradução espanhola: *L'actitud intencional*, Barcelona, Gedisa, 1991.

DENNETT, Daniel, 1991, *Consciousness Explained*, Boston, Little Brown. Tradução francesa: *La conscience expliquée*, Paris, O. Jacob, 1993. Tradução italiana: *Che cosa è conscienza*, Milano, Rizzoli, 1993. Tradução espanhola: *La conciencia explicada: una teoría interdisciplinar*, Barcelona, Paidos, 1995.

DENNETT, Daniel, 1995, *Darwin's Dangerous Idea – Evolution and The Meanings of Life*, New York, Touchstone. Tradução francesa: *Darwin est-il dangereux?*, Paris, Jacob, 1999. Tradução alemã: *Darwins gefährliche Erbe. Die Evolution und der Sinne des Lebens*, Heidelberg, Hoffman, 1997. Tradução espanhola: *La peligrosa idea de Darwin: evolución y significados de la vida*, Barcelona, Galaxia Gutenberg, 2000. Tradução portuguesa: *A Ideia Perigosa de Darwin*, Lisboa, Círculo de Leitores, 2000.

DENNETT, Daniel, 1996, *Kinds of Minds, Toward an Understanding of Consciousness*, New York, Basic Books. Tradução italiana: *La mente e le menti*, Milano, Rizzoli, 1997. Tradução alemã: *Spielarten des Geistes*, M., Bertelsmann Verlag, 1999. Tradução francesa: *La diversité des esprits*, Paris, Hachette, 1999. Tradução espanhola: *Tipos de mente: hacia una comprensión de la conciencia*, Madrid, Ed. Debate. Tradução portuguesa: *Tipos de Mentes*, Lisboa, Temas e Debates.

DENNETT, Daniel, 1998, *Brainchildren – Essays on Designing Minds*, London, Penguin.

HOFSTADTER, Douglas & DENNETT, Daniel, 1981, *The Mind's I – Fantasies and Reflections on Self and Soul*, New York, Bantam Books. Tradução francesa: *Vues de l'esprit*, Paris, Interdictions, 1987. Tradução alemã: *Einsicht ins Ich. Fantasien und Reflexionen über Selbst und Seele*, Stuttgart, Klett-Cotta, 1986. Tradução italiana: *L'io della mente. Fantasie e riflessioni sul sé e sull'anima*, Milano, Adelphi, 1988.

I.II *Artigos:*

DENNETT, Daniel, 1968a, Machines Traces and Protocol Statements, *Behavioral Science*, 13 (2), 155-161.

DENNETT, Daniel, 1968b, Features of Intentional Actions, *Philosophy and Phenomenological Research*, 29 (2), 232-244.

DENNETT, Daniel, 1968c, Geach on Intentional Identity, *Journal of Philosophy*, 65, 11, 335-341.

DENNETT, Daniel, 1971, Review of C.O. Adams' *The Subject of Consciousness*, *Philosophical Quarterly*, 21 (83), 180-181.

DENNETT, Daniel, 1972, Review of J.R. Lucas' *The Freedom of the Will*, *Journal of Philosophy*, 69 (17), 527-531.

DENNETT, Daniel, 1974, Comment on Wilfrid Sellars' Meaning as Functional Classification, *Synthese* 27, 439-444.

DENNETT, Daniel, 1975, Review of G. Warnok and B. F. Skinner *Behavior Control: Freedom and Morality*, *Teaching Philosophy*, 1, 175-177.

DENNET, Daniel, 1977a, Review of Jonathan Glover (ed.) *The Philosophy of Mind*, *Teaching Philosophy*, 2, 325-326.

DENNETT, Daniel, 1977b, Review of W. Matson's *Sentience*, *Studies in Philosophy* 9, 182-183.

DENNETT, Daniel, 1978a [1971], Intentional Systems, in DENNETT 1978, originalmente publicado em *Journal of Philosophy* LXVIII, 4, 1971, 87-106.

DENNETT, Daniel, 1978b, Reply to Arbib and Gunderson, in DENNETT 1978, primeira aparição resumida em *Journal of Philosophy* LXIX, 18, 1972, 604.

DENNETT, Daniel, 1978c [1975], Why The Law of Effect Won't Go Away, in DENNETT 1978, originalmente publicado em *Journal of the Theory of Social Behavior*, V, 2, 1975, 169-187.

DENNETT, Daniel, 1978d, Skinner Skinned, in DENNETT 1978.

DENNETT, Daniel, 1978e [1975], Brain Writing and Mind Reading, in DENNETT 1978, originalmente publicado em K. Gunderson (ed.), *Language, Mind and Knowledge*, Minnesota Studies in the Philosophy of Science VII, Minneapolis, University of Minnesota Press, 1975, 403-415.

DENNETT, Daniel, 1978f [1977], A Cure For the Common Code, in DENNETT 1978, recensão de Jerry Fodor, *The Language of Thought*, originalmente publicada em *Mind* 86, 265-280.

DENNETT, Daniel, 1978g, Toward a Cognitive Theory of Consciousness, in DENNETT 1978, também em C. SAVAGE (ed.), *Perception and Cognition*, Minnesota Studies in the Philosophy of Science IX, Minneapolis, University of Minnesota Press, 1978, 201-228.

DENNETT, Daniel, 1978h, IA as Philosophy and as Psychology, in DENNETT 1978, também em M. Ringle (ed.), *Philosophical Perspectives in Artificial Intelligence*, Atlantic Highlands NJ, Humanities Press, 1979, 57-80.

DENNETT, Daniel, 1978i, Two Approaches to Mental Images, in DENNETT 1978, também em N. BLOCK 1981, 87-107.

DENNETT, Daniel, 1978j [1976], Are Dreams Experiences?, in DENNETT 1978, originalmente publicado em *Philosophical Review*, LXXIII, 1976, 151-171.

DENNETT, Daniel, 1978l, Why You Can't Make a Computer That Feels Pain, in DENNETT 1978.

DENNETT, Daniel, 1978m [1973], Mechanism and Responsability, in DENNETT 1978, originalmente publicado em T. Honderich (ed.), *Essays on Freedom of Action*, London, Routledge & Kegan Paul, 1973, 157-184.

DENNETT, Daniel, 1978n [1976], Conditions of Personhood, in DENNETT 1978, originalmente publicado em RORTY 1976, 175-196.
DENNETT, Daniel, 1978 o [1970], The Abilities of Men and Machines, in DENNETT 1978, primeira aparição resumida em *Journal of Philosophy*, LXVII, 20, 835.
DENNETT, Daniel 1978p, On Giving Libertarians What They Say They Want, in DENNETT 1978.
DENNETT, Daniel, 1978q, How to Change your Mind in DENNETT 1978.
DENNETT, Daniel, 1978r, Where Am I? in DENNETT 1978, também em HOFSTADTER & DENNETT 1981, 217-229.
DENNETT, Daniel, 1978s, Why Not The Whole Iguana (commentary on Pylyshyn), *Behavioral and Brain Sciences* 1 (1), 103-105.
DENNETT, Daniel, 1978t, Beliefs about Beliefs (commentary on Premack and Woodruff), *Behavioral and Brain Sciences*, 1 (4), 568-570.
DENNETT, Daniel, 1978u, Co-Opting Holograms, *Behavioral and Brain Sciences*, 1 (2), 232-233.
DENNETT, Daniel, 1978v, Requisition for a Pexgo, *Behavioral and Brain Sciences*, 1 (1), 56-57.
DENNETT, Daniel 1978w, Current Issues in the Philosophy of Mind, *American Philosophical Quarterly*, 15 (4), 249-261.
DENNETT, Daniel, 1978x, Review of Margaret Boden's *Artificial Intelligence and the Natural Man, Philosophy of Science*, 45 (4), 648-649.
DENNETT, Daniel, 1978y, Review of R. Aquila's *Intentionality* and E. Casey's *Imagining, Southwestern Journal of Philosophy*, 9 (3), 139-143.
DENNETT, Daniel, 1978z, What's the difference: some riddles, *Behavioral and Brain Sciences* 1 (3), 351.
DENNETT, Daniel, 1979a, Breeding Cognitive Strategies, *Behavioral and Brain Sciences* 2 (4), 599-600.
DENNETT, Daniel, 1979b, On the Absence of Phenomenology, in D. Gustafson & B. Tapscott (eds.), *Body, Mind and Method: Essays in Honor of V. Aldrich*, Dordrecht, Reidel, 93-113.
DENNETT, Daniel, 1979c, The Onus Re Experiences: A Reply to Emmett, *Philosophical Studies*, 35 (3) 315-318.
DENNETT, Daniel, 1979d, Reply to K. Emmett's 'Oneiric Experiences', *Philosophical Studies*, 34, 445-450.
DENNETT, Daniel, 1979e, Review of K. Popper & J. Eccles *The Self and Its Brain: An Argument for Interaccionism*, *Journal of Philosophy*, 76 (2), 91-97.
DENNETT, Daniel, 1980a, The Milk of Human Intentionality, *Behavioral and Brain Sciences* 3 (3), 428-430.
DENNETT, Daniel, 1980b, Passing the Buck Back to Biology, *Behavioral and Brain Sciences* 3 (1), 19.
DENNETT, Daniel, 1980c, Reply to Professor Stich, *Philosophical Books*, 21 (2), 73-76.
DENNETT, Daniel, 1981a, Making Sense of Ourselves, *Philosophical Topics*, 12 (1), 63-81.

DENNETT, Daniel, 1981b, The Nature of Images and the Introspective Trap, in BLOCK 1981, 51-61.
DENNETT, Daniel, 1981c, Perspectivas Actuales en la Filosofia de la Mente, *Teorema*, 11, 197-230.
DENNETT, Daniel (com D. Hofstadter) 1981d, Reflections on Borges' *Borges and I*, in HOFSTADTER & DENNETT 1981, 20-22.
DENNETT, Daniel, 1981e, Reflections on R. Rucker's *Software*, in HOFSTADTER & DENNETT 1981, 265-268.
DENNETT, Daniel, 1981f, Wondering Where the Yellow Went, *Monist*, 64 (1), 102-108.
DENNETT, Daniel, 1981g, Introduction, in DENNETT & HOFSTADTER 1981, 3-16.
DENNETT, Daniel, 1982a, Comments on Rorty (on R. Rorty's 'Contemporary Philosophy of Mind), *Synthese*, 53 (2), 349-356.
DENNETT, Daniel, 1982b, Grey Matter and Mind, *Radio Times*, (May 8-14), 70--72.
DENNETT, Daniel, 1982c, How to Study Human Consciousness Empirically or Nothing Comes to Mind, *Synthese*, 53 (2), 159-180.
DENNETT, Daniel, 1982d, The Imagination Extenders, *Psychology Today*, 16 (2), 32-39.
DENNETT, Daniel, 1982e, The Myth of the Computer: An Exchange, *New York Review of Books*, 29 (11), 56.
DENNETT, Daniel, 1982f, Philosophy According to Nozick, *New Boston Review*, 7, 9-11.
DENNETT, Daniel, 1982g, Recent Work in Philosophy of Interest to AI, *Artificial Intelligence*, 19 (1), 3-5.
DENNETT, Daniel, 1982h, Review of G. Ryle's *On Thinking, International Studies in Philosophy*, 14 (1), 98-99.
DENNETT, Daniel, 1982i, Review of R. Gregory's *Mind in Science, Contemporary Psychology*, 27 (8), 597-598.
DENNETT, Daniel, 1982j, Why We Think What We Do about Why We Think What We Do: Discussion on Goodman's *Thoughts Withouht Words, Cognition* 12 (2), 219-227.
DENNETT, Daniel, 1983a, Artificial Intelligence and the Strategies of Psychological Investigation, in J. Miller's, *States of Mind*, New York, Pantheon Books, 1983.
DENNETT, Daniel, 1983b, Taking the Intentional Stance Seriously, *Behavioral and Brain Sciences*, 6 (3), 379-388.
DENNETT, Daniel, 1983c, When Do Representations Explain, *Behavioral and Brain Sciences*, 6 (3), 406-407.
DENNETT, Daniel, 1984a, Carving the Mind at Its Joints: A Review of Fodor's *Modularity of Mind, Contemporary Psychology*, 29, 285-286.
DENNETT, Daniel, 1984b, Formulating Human Purposes, in R. Mason, *Information Processing* 83, Amsterdam, Elsvier.

DENNETT, Daniel, 1984c, Wishful Thinking – Commentary on B. F. Skinner 'The Operational Analysis of Psychological Terms', *Behavioral and Brain Sciences* 7(4), 556-557.
DENNETT, Daniel, 1984d, Foreword for R. Millikan's *Language Thought and Other Biological Categories*, in MILLIKAN 1984, ix-x.
DENNETT, Daniel, 1984e, Correspondent's Report: Recent Work in Philosophy II, *Artificial Intelligence*, 22, 231-234.
DENNETT, Daniel, 1984f, The Role of the Computer Metaphor in Understanding the Mind, in H. Pagels (ed.) *Computer Culture*, Annals of the NY Academy of Science, 426, 266-275.
DENNETT, Daniel, 1984g, I Could not Have Done Otherwise – so what?, *Journal of Philosophy*, 81, 553-565.
DENNETT, Daniel, 1984h, Thinking About Thinking: the Mind as Virtual Machine, in *The Human Mind, the Human Brain and Beyond*, American Broadcasting Company.
DENNETT, Daniel, 1985a, Music of the Hemispheres – Review of M. Gazzaniga's *The Social Brain*, *New York Times*, Nov. 17, 53.
DENNETT, Daniel, 1985b, When Does the Intentional Stance Work, *Behavioral and Brain Sciences*, 8 (4), 763-766.
DENNETT, Daniel, 1985c, Why Believe in Belief, Review of S. Stich *From Folk Psychology to Cognitive Science*, *Contemporary Psychology*, 30 (12), 949.
DENNETT, Daniel, 1986a, Preface to Second Edition of *Content and Consciousness*, in DENNETT [1969].
DENNETT, Daniel, 1986b, Engineering's Baby, *Behavioral and Brain Sciences*, 9 (1), 141-142.
DENNETT, Daniel, 1986c, Information, Technology and the Virtues of Ignorance, *Daedalus*, 115 (3), 135-153.
DENNETT, Daniel, 1986d, Is There and Autonomous Knowledge Level, in Z. Pylyshyn & W. Demopoulos (eds.), *Meaning and Cognitive Structure: Issues in the Computational Theory of the Mind*, Norwood NJ, Ablex, 51-54.
DENNETT, Daniel, 1986e, Philosophy as Mathematics or as Anthropology, *Mind & Language*, 1 (1), 18-19.
DENNETT, Daniel, 1986f, Review of V. Braitenberg's *Vehicles*, *Philosophical Review*, 95 (1), 137-139.
DENNETT, Daniel & AKINS, Kathleen, 1986, Who May I Say is Calling, *Behavioral and Brain Sciences*, 9 (3), 517-518.
DENNETT, Daniel 1987a, Commentary on Cam's 'Propositions about Images', *Philosophy and Phenomenological Research*, 48 (2), 339-341.
DENNETT, Daniel 1987b, Artificial Intelligence. Where Are We, *Abacus*, 4 (4), 40.
DENNETT, Daniel 1987c, «Consciousness», in R. Gregory (ed.) *The Oxford Companion to the Mind*, Oxford, Oxford University Press, 161-164.
DENNETT, Daniel 1987d, Eliminate the Middletoad!, *Behavioral and Brain Sciences*, 10 (3), 372-374.
DENNETT, Daniel 1987e, Skinner Placed, in S. Modgil & C. Modgil (eds.), B. F. *Skinner: Consensus and Controversy*, New York, Falmer Press.

DENNETT, Daniel, 1987f, Mid-Term Examination, Compare and Contrast, in DENNETT 1987.
DENNETT, Daniel, 1987g [1981], True Believers, in DENNETT 1987 (também em M. M. Carrilho (ed.), *Dicionário do Pensamento Contemporâneo*, Lisboa, Dom Quixote, 1991, 13-35, sob o título «Intencionalidade»).
DENNETT, Daniel, 1987h [1981], Three Kinds of Intentional Psychology, in DENNETT 1987.
DENNETT, Daniel, 1987i [1981], Making Sense of Ourselves, in DENNETT 1987, originalmente publicado em *Philosophical Topics*, 12 (1), 1981, 63-81.
DENNETT, Daniel, 1987j, [1982] Beyond Belief, in DENNETT 1987, originalmente publicado em A. Woodfield, *Thought and Object: Essays on Intentionality*, Oxford, Clarendon Press, 1982, 1-96.
DENNETT, Daniel, 1987l, Fast Thinking, in DENNETT 1987.
DENNETT, Daniel, 1987m [1983] Intentional Systems in Cognitive Ethology: The Panglossian Paradigm Defended, originalmente publicado em *Behavioral and Brain Sciences* 6 (3), 343-390 (também publicado como «Intentionale Systeme in der Kognitiven Verhaltensforschung», in D. Munch (ed.), *Kognitionwissenschaft*, Frankfurt, Suhrkamp, 1992).
DENNETT, Daniel, 1987n [1982-1983], Styles of Mental Representations, in DENNETT 1978, originalmente publicado em *Proceedings of the Aristotelian Society*, 83, 213-226.
DENNETT, Daniel, 1987o, Evolution, Error and Intentionality, in DENNETT 1987.
DENNETT, Daniel, 1988a, Coming to Terms with the Determined, Review of T. Honderich's *A Theory of Determinism*, *Times Literary Supplement*, November 4-10, 4466, 1219-1220.
DENNETT, Daniel, 1988b, The Intentional Stance in Theory and Practice, in WHITEN & BYRNE 1988, 180-202.
DENNETT, Daniel, 1988c, The Moral First Aid Manual, in S. McMurrin (ed.) *Tanner Lectures on Human Values*, III, Salt Lake City, University of Utah Press, 119-147.
DENNETT, Daniel, 1988d, Précis of *The Intentional Stance*, *Behavioral and Brain Sciences*, 11 (3), 495-504.
DENNETT, Daniel, 1988e, Quining Qualia, in MARCEL & BISIACH 1988, 42--77 (também em GOLDMAN 1993 e em BLOCK, FLANAGAN & GÜZELDERE 1997).
DENNETT, Daniel, 1988f, Review of J. Fodor's *Psychosemantics*, *Journal of Philosophy*, 85 (7), 384-389.
DENNETT, Daniel, 1988g, Review of W. Lyons' *The Disappearance of Introspection*, *Philosophy of Science*, 55 (4) 653-654.
DENNETT, Daniel, 1988h, Science, Philosophy and Interpretation, *Behavioral and Brain Sciences*, 11(3), 535-646.
DENNETT, Daniel, 1988i, Why Everyone is a Novelist, *Times Literary Supplement*, September 16-22, 4459: 1016, 1028-1029 (também publicado como «The Self as a Centre of Narrative Gravity» em F. Kessel, P. Cole & D. Johnson (eds.), *Self and Consciousness*, Hillsdale, NJ, Erlbaum, 1992).

DENNETT, Daniel, 1989a, Murmurs in the Cathedral, Review of R. Penrose's *The Emperor's New Mind*, *Times Literary Supplement*, September 29-October 5, 4513, 1055-1056.
DENNETT, Daniel, 1989b, The Origin of Selves, *Cogito*, 3, 163-173.
DENNETT, Daniel, 1989c, Review of R. Richard's *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior*, *Philosophy of Science*, 56 (3), 540-543.
DENNETT, Daniel, 1990a, Abstracting From Mechanism, *Behavioral and Brain Sciences*, 13 (3), 583-584.
DENNETT, Daniel, 1990b, Artificial Life: A Feast For the Imagination, Review of C. Langton's *Artificial Life*, *Biology and Philosophy*, 5 (4), 489-492.
DENNETT, Daniel, 1990c, Attitudes about ADHD: Some Analogies and Aspects, in K. Conners & M. Kinsbourne (eds.), *Attention Deficit Hyperactivity Disorder*, Munich, Medizin Verlag, 1990.
DENNETT, Daniel, 1990d, Teaching an Old Dog New Tricks, *Behavioral and Brain Sciences*, 13 (1), 76-77.
DENNETT, Daniel, 1990e, Doctor Pangloss Knows Best, *Behavioral and Brain Sciences* 13 (3), 581-582.
DENNETT, Daniel, 1990f, Betting Your Life on an Algorithm, *Behavioral and Brain Sciences* 13 (4), 660-661.
DENNETT, Daniel, 1990g, Demystifying Consciousness, *Edge*, November 1990, 5-8.
DENNETT, Daniel 1990h, Evolution of Consciousness, in J. Brockman (ed), *Speculations*, New York, Prentice Hall, 85-108.
DENNETT, Daniel, 1990i, The Interpretation of Texts, People and Other Artifacts, *Philosophy and Phenomenological Research*, 50 (Supplement), 177-194.
DENNETT, Daniel, 1990j, Memes and the Exploitation of Imagination, *Journal of Aesthetics and Art Criticism*, 48 (2), 127-135.
DENNETT, Daniel, 1990l, The Myth of Original Intentionality, in K. Mohyeldin Said, W. H. Newton-Smith, R. Viale & K. Wilkes (eds.), *Modeling the Mind*, Oxford, Clarendon Press, 43-62.
DENNETT, Daniel 1990m, Thinking With a Computer in H. Barlow, C. Blakemore & M. Weston-Smith (eds.), *Images and Understanding: Thoughts About Images, Ideas About Understanding*, Cambridge, Cambridge University Press, 1990.
DENNETT, Daniel, 1990n, Ways of Establishing Harmony: A Reply to Dretske's 'Does Meaning Matter'?, in E. Villanueva (ed.), *Information, Semantics and Epistemology*, Cambridge, Blackwell 1990, 18-27 (também em B. P. McLaughlin (ed.), *Dretske and His Critics*, Cambridge MA, Blackwell, 1991).
DENNETT, Daniel 1991a, The Brain and Its Boundaries, Review of C. McGinn's *The Problem of Consciousness*, *Times Literary Supplement*, May 10, 4597, 10.
DENNETT, Daniel, 1991b, Granny's Campaign for Safe Science in LOEWER & REY 1991, 87-94.
DENNETT, Daniel, 1991c, A Dash to Solve the Puzzle of Mind, *Boston Globe*, December 8, A21, A24.
DENNETT, Daniel, 1991d, *A Gadfly's View*, in *Teaching and Technology*, Ann Arbor, MI, Pierian Press.

DENNETT, Daniel, 1991e, Lovely and Suspect Qualities, in E. Villanueva (ed.), *Consciousness*, Atascadero, California, Ridgeview, 1991, 37-43.
DENNETT, Daniel, 1991f, Media Neutral, in W. Coleman & P. Perrin (eds.), *The Jamais Vu Papers*, New York, Harmony.
DENNETT, Daniel, 1991g, Modelli del cervello e il punto di vista dell'osservatore, in G. Giorello & P. Stratta (eds.), *L'automa spirituale: menti, cervelli e computer*, Bari, Laterza.
DENNETT, Daniel, 1991h, Mother Nature and the Walking Encyclopedia, in W. Ramsey, S. Stich & D. Rumelhart (eds.), *Philosophy and Connectionist Theory*, Hillsdale NJ, Erlbaum, 1991, 21-30.
DENNETT, Daniel, 1992a, The Architecture of Consciousness, *International Journal of Psychology*, 27 (3-4), 1.
DENNETT, Daniel, 1992b, Filling In versus Finding Out, in H. Pick, P. van den Broek & D. Knill (eds.), *Cognition: Conceptual and Methodological Issues*, Washington DC, American Psychological Association, 1992, 33-49.
DENNETT, Daniel, 1992c, Hitting the Nail in the Head, Commentary on E. Thompson, A. Palacios & F. Varela, *Ways of Coloring*, *Behavioral and Brain Sciences*, 15 (1), 35.
DENNETT, Daniel, 1992d, Privileged Access and Consciousness, in T. Nelson, *Metacognition*, Boston, Allyn & Bacon, 1992, 27-99.
DENNETT, Daniel, 1992e, Review of F. Varela, E. Thompson & E. Rosch's *The Embodied Mind* and G. Edelman's *Bright Air, Brilliant Fire*, *New Scientist*, 134 (1825), 48-49 .
DENNETT, Daniel, 1992f, Temporal Anomalies of Consciousness: Implications of the Uncentered Brain, in Y. Christen & P. S. Churchland (eds.), *Neurophilosophy and Alzheimer's Disease*, Berlin, Springer.
DENNETT, Daniel & KINSBOURNE, Marcel, 1992a, Time and The Observer: the Where and When of Consciousness in the Brain, *Behavioral and Brain Sciences*, 15 (2), 183-201.
DENNETT, Daniel & KINSBOURNE, Marcel, 1992b, Escape From the Cartesian Theater, *Behavioral and Brain Sciences*, 15 (2), 234-247.
DENNETT, Daniel, 1993a, Back From the Drawing Board, in DAHLBOM 1993, 203-235.
DENNETT, Daniel 1993b, Précis of *Consciousness Explained*, in *Philosophical and Phenomenological Research*, 53 (4), November-December, 889-892.
DENNETT, Daniel, 1993c, The Message is: There is no Medium in *Philosophical and Phenomenological Research*, 53 (4), November-December, 919-931.
DENNETT, Daniel, 1993d, Caveat Emptor, *Consciousness and Cognition*, 2, 1, 48--57.
DENNETT, Daniel, 1993e, Down With School! Up With Legoland!, Review of S. Papert's *The Children's Machine*, New Scientist, November 6, 45-46.
DENNETT, Daniel, 1993f, Evolution, Teleology, Intentionality, *Behavioral and Brain Sciences*, 16 (2), 389-391.
DENNETT, Daniel, 1993g, Learning and Labeling, *Mind and Language*, 8 (4), 540-548.

DENNETT, Daniel, 1993h, Living on the Edge, *Inquiry*, 36 (1-2), 135-159.
DENNETT, Daniel, 1993i, Review of J. Searle's *The Rediscovery of the Mind*, *Journal of Philosophy*, 90 (4), 193-205.
DENNETT, Daniel, 1994a, Counting Consciousness – None, One, Two or None of the Above, *Behavioral and Brain Sciences*, 17 (1), 178-180.
DENNETT, Daniel, 1994b, «Dennett» in GUTTENPLAN 1994, 236-244.
DENNETT, Daniel, 1994c, E Pluribus Unum, *Behavioral and Brain Sciences*, 17 (4), 617-618.
DENNETT, Daniel 1994d, The Role of Language in Intelligence, in J. Khalfa (ed.), *What is Intelligence?*, Cambridge, Cambridge University Press, 1994, 161-178.
DENNETT, Daniel, 1994e, Review of R. Jackendoff's *Patterns in the Mind* and S. Pinker's *The Language Instinct*, *London Review of Books*, 16 (12), 10-11.
DENNETT, Daniel, 1994f, Get Real, *Philosophical Topics*, 22, 1-2, 505-568.
DENNETT, Daniel, 1995a, Darwin's Dangerous Idea, *Sciences*, 35 (3), 34-40.
DENNETT, Daniel, 1995b, Dissecting Darwin – Reply, *Sciences*, 35 (5), 48.
DENNETT, Daniel, 1995c, Evolution as Algorithm – The ultimate insult?, in H. Morowitz & J. Singer (eds.), *The Mind, the Brain and Complex Adaptive Systems*, Reading MA, Addison-Wesley, 1995, 221-223.
DENNETT, Daniel, 1995d, «Homunculus», «Intentionality», «Mentalese», «Mentalism», «Topic-Neutral» in R. Audi (ed.), *The Cambridge Dictionary of Philosophy*, Cambridge, Cambridge University Press, pp. 337, 381, 482, 482--483, 807.
DENNETT, Daniel, 1995e, How to Make Mistakes, in J. Brockman & K. Matson (eds.), *How Things Are: A Science Tool-Kit For the Mind*, New York, Morrow, 1995, 137-144.
DENNETT, Daniel, 1995f, Interview With Daniel Dennett, *Journal of Cognitive Neuroscience*, 7 (3), 408-414.
DENNETT, Daniel, 1995g, Is Perception the Leading Edge of Memory?, *Iride*, 7 (4), 59-78.
DENNETT, Daniel, 1995h, The Path Not Taken, *Behavioral and Brain Sciences*, 18 (2), 252-253 (também em BLOCK, FLANAGAN & GÜZELDERE 1997).
DENNETT, Daniel, 1995i, Overworking the Hippocampus, *Behavioral and Brain Sciences*, 18 (4), 677-678.
DENNETT, Daniel, 1995j, In Defense of AI, in BAUMGARTNER & PAYR 1995, 59-69.
DENNETT, Daniel, 1995l, The Mistery of Consciousness, *New York Review of Books*, 42 (20), 83 (também em SEARLE 1997).
DENNETT, Daniel, 1995m, Review of António Damásio's *Descartes' Error*, *Times Literary Supplement*, August 25, 4821, 3-4.
DENNETT, Daniel, 1995n, Comment nous tissons notre moi, in GRIVOIS & DUPUY 1995, 147-166.
DENNETT, Daniel & KINSBOURNE, Marcel, 1995, Multiple Drafts: An Eternal Golden Braid?, *Behavioral and Brain Sciences*, 18 (4), 810-811.
DENNETT, Daniel, 1996a, Facing Backwards on the Problem of Consciousness, *Journal of Consciousness Studies*, 3 (1), 4-6 (também em SHEAR 1997).

DENNETT, Daniel, 1996b, Cow-Sharks, Magnets and Swampman, *Mind and Language*, 11 (1), 76-77.
DENNETT, Daniel, 1996c, Granny Versus Mother Nature – No Contest, *Mind and Language*, 11 (3), 263-269.
DENNETT, Daniel, 1996d, Quantum Incoherence, Review of A. Cairns-Smith's *Evolving the Mind*, *Nature*, 381 (6582), 485-486.
DENNETT, Daniel, 1996e, Review of T. Nagel's *Other Minds*, *Journal of Philosophy*, 93 (8), 425-428.
DENNETT, Daniel, 1996f, Seeing is Believing – or is it?, in K. Akins (ed.), *Perception*, Oxford, Oxford University Press, 1996, 158-172.
DENNETT, Daniel, 1996g, Consciousness: More Like Fame than Television, em http://ase.tufts.ed/as/cogstud/papers/ concfame.htm (também publicado como «Bewusstsein hat mehr mit Ruhm als mit Fernsehn zu tun» em C. Maar, E. Pöppel & T. Christaller (eds.), *Die Technik auf dem Weg Zur Seele*, Rowohlt).
DENNETT, Daniel, 1996h, The scope of natural selection (Reply to Orr), *Boston Review*, 21 (5).
DENNETT, Daniel, 1997a, Review of Andy Clark's *Being There*, *Times Literary Supplement*, May 16, 4911,5.
DENNETT, Daniel, 1997b, Review of W. Burkert's *Creation of the Sacred: Tracks of Biology in Early Religions*, *Sciences*, 37 (1), 39-45.
DENETT, Daniel, 1997c, The Cartesian Theater and Filling in the Stream of Consciousness, in BLOCK, FLANAGAN & GÜZELDERE 1997, 83-88.
DENNETT, Daniel, 1997d, Fé na Verdade, *disputatio* 3, 3-21.
DENNETT, Daniel, 1998a [1985], Can Machines Think?, in DENNETT 1998, inicialmente publicado em M. Shafto (ed.), *How We Know*, San Francisco, Harper and Row.
DENNETT, Daniel & HUMPHREY, Nicholas, 1998 [1989], Speaking for Ourselves: An Assessment of Multiple Personality Disorder, in DENNETT 1998, inicialmente publicado em *Raritan*, 9 (1), 68-98.
DENNETT, Daniel, 1998b [1992], Do It Yourself Understanding, in DENNETT 1998, inicialmente publicado como *La compréhension artisanale* em D. Fisette (ed.), *Daniel C. Dennett et les Stratégies Intentionelles*, Montreal, Université de Quebec, 1992, 27-52.
DENNETT, Daniel 1998c, [1991], Two Contrasts, Folk Craft versus Folk Science and Belief versus Opinion, inicialmente publicado em J. Greenwood (ed.), *The Future of Folk Psychology: Intentionality and Cognitive Science*, Cambridge, Cambridge University Press, 1991, 135-148.
DENNETT, Daniel, 1998d [1991], Real Patterns, in DENNETT 1998, inicialmente publicado em *Journal of Philosophy*, 88 (1), 27-51.
DENNETT, Daniel, 1998e, [1986], Julian Jaynes Software Archeology, in DENNETT 1998, inicialmente publicado em *Canadian Psychology*, 27 (2), 149-154.
DENNETT, Daniel, 1998f [1994], Real Consciousness, inicialmente publicado em A. Revonsuo & M. Kampinnen (eds.), *Consciousness in Philosophy and Cognitive Neuroscience*, Hillsdale NJ, Erlbaum, 1994, 55-63.

DENNETT, Daniel, 1998g [1994] Instead of Qualia, in DENNETT 1998, inicialmente publicado em A. Revonsuo & M. Kampinnen (eds.), *Consciousness in Philosophy and Cognitive Neuroscience*, Hillsdale NJ, Erlbaum, 1994, 129-139.
DENNETT, Daniel 1998h [1994], The Practical Requirements for Making a Conscious Robot, inicialmente publicado em *Philosophical Transactions of the Royal Society*, 349 (1689), 133-146.
DENNETT, Daniel, 1998i [1995], The Unimagined Preposterousness of Zombies, in DENNETT 1998, inicialmente publicado em *Journal of Consciousness Studies*, 2 (4), 322-326.
DENNETT Daniel, 1998j [1984], Cognitive Wheels: The Frame Problem of AI, in DENNETT 1998, inicialmente publicado em C. Hookway (ed.), *Minds, Machines and Evolution*, Cambridge, Cambridge University Press, 1984, 129-151.
DENNETT, Daniel, 1998l [1986], Producing Future by Telling Stories, in DENNETT 1998, inicialmente publicado em FORD & PYLYSHYN 1996, 1-7.
DENNETT, Daniel, 1998m [1986], The Logical Geography of Computational Approaches, inicialmente publicado em M. Brand & R. Harnish (eds.), *The Representation of Knowledge and Belief*, Tucson, University of California Press, 59-79.
DENNETT, Daniel, 1998n, [1996] Hofstadter's Quest, in DENNETT 1998, inicialmente publicado em *Complexity* 1 (6), 9-11.
DENNETT, Daniel, 1998o [1995], Foreword to R. French *The Subtlety of Sameness*, inicialmente publicado em R. French, *The Subtlety of Sameness*, Cambridge MA, MIT Press, 1995, vii-xi.
DENNETT, Daniel, 1998p [1994], Cognitive Science as Reverse Engineering: Several Meanings of 'Top-Down' and 'Bottom-Up', in DENNETT 1998, inicialmente publicado em D. Prawitz, B. Skyrms & D. Westerstahl (eds), *Logic, Methodology and Philosophy of Science* IX, Amsterdam, Elsevier Science, 1994, 679-689.
DENNETT, Daniel, 1998q [1994], Artificial Life as Philosophy, in DENNETT 1998, inicialmente publicado em *Artificial Life*, 1, 291-292.
DENNETT, Daniel, 1998r [1988], When Philosophers Encounter Artificial Intelligence, in DENNETT 1998, inicialmente publicado em *Daedalus*, 117 (1), 283-295.
DENNETT, Daniel, 1998s [1993], Review of Allen Newell's *Unified Theories of Cognition*, inicialmente publicado em *Artificial Intelligence*, 59 (1-2), 285-294.
DENNETT, Daniel, 1998t [1988], Out of the Armchair and Into the Field, in DENNETT 1998, inicialmente publicado em *Poetics Today*, 9 (1), 205-221.
DENNETT, Daniel, 1998u [1989], Cognitive Ethology: Hunting for Bargains or a Wild Goose Chase, inicialmente publicado em A. Montefiore & D. Noble (eds.), *Goals, No-Goals and Own Goals: A Debate on Goal-Directed and Intentional Behaviour*, London, Unwin Hyman, 1989, 101-116.
DENNETT, Daniel, 1998v [1995], Do Animals Have Beliefs?, in DENNETT 1998, inicialmente publicado em H. Roitblat & J.- A. Meyer (eds.), *Comparative Approaches to Cognitive Science: Complex Adaptive Systems*, Cambridge MA, MIT Press, 1995, 111-118.

DENNETT, Daniel, 1998w [1988], Why Creative Intelligence is Hard to Find: Commentary on Whiten and Byrne, in DENNETT 1998, inicialmente publicado em *Behavioral and Brain Sciences*, 11 (2), 253.
DENNETT, Daniel, 1998x [1995], Animal Consciousness, What Matters and Why, in DENNETT 1998, inicialmente publicado em *Social Research*, 62 (3), 691-710.
DENNETT, Daniel, 1998y [1994], Self-Portrait, in DENNETT 1998, originalmente publicado em GUTTENPLAN 1994, 236-244.
DENNETT, Daniel, 1998z [1986], Information, Technology and the Virtues of Ignorance, in DENNETT 1998, inicialmente publicado em *Daedalus*, 115 (3), 135-153.
DENNETT, Daniel, 1998z', Preston on Exaption: Herons, Apples and Eggs, *The Journal of Philosophy*, XCV, 11, 576-80.
DENNETT, Daniel, 1998z", Bombas de Intuição, in Brockman, John, *A Terceira Cultura*, Lisboa, Temas e Debates, 168-173.
DENNETT, Daniel, 1998z"', Things About Things, Final Draft for Lisbon Conference on Cognitive Science, em http://ase.tufts.edu/cogstud/papers/ /lisbon.htm. (também em BRANQUINHO, J., (ed.), *The Foundations of Cognitive Science*, Oxford, OUP, 2001)
DENNETT, Daniel & DENSMORE, Shannon, 1999, The Virtues of Virtual Machines, *Philosophy and Phenomenological Research*, LIX, 3, September, 747-767.
DENNETT, Daniel, 1998z"", The Myth of Double Transduction, in S. Hameroff, A. Kaszniak & A. C. Scott (eds.), *Toward a Science of Consciousness II – The Second Tucson Discussions and Debates*, Cambridge MA, MIT Press.
DENNETT, Daniel, 1999a, Review of John Haugeland's *Having Thought*, *Journal of Philosophy*, XCVI, 8, 430-435.
DENNETT, Daniel, 1999b, Is Hirsh or Wilson Confused?, em http://ase.tufts.ed/as/cogstud/ papers/hirsch.fin.html.
DENNETT, Daniel, 1999c, The Possibility of sub-isomorphic experiential differences, em http://ase.tufts.ed/as/cogstud/ papers/palmsbbs.html.
DENNETT, Daniel, 1999d, Swift and enormous, em http://ase.tufts.ed/as/ /cogstud/ papers/palmer.html.
DENNETT, Daniel, 1999e, It's not a Bug, It's a Feature, *Journal of Consciousness Studies*, 7,4, 25-27.
DENNETT, Daniel, 1999f, Sort-of symbols?, *Behavioral and Brain Sciences*, 22 (4), 613.
DENNETT, Daniel, 1999g, We Earth Neurons, em http://ase.tufts.ed/as/cogstud/ papers/earthneuron.htm.
DENNETT, Daniel, 1999h, The Zombic Hunch: Extinction of an Intuition? http://ase.tufts.ed/as/cogstud/papers/zombic.htm.
DENNETT, Daniel, 1999i, The Evolution of Culture, Charles Simonyi Lecture delivered at Oxford University, http://www.edge.org.
DENNETT, Daniel 2000a, The Case for Rorts, in Brandon, R.B. (ed.), 2000, *Rorty and His Critics*, Oxford, Blackwell (também em http://ase.tufts.ed/ /as/cogstud/ papers/rort.htm).

DENNETT, Daniel 2000b, Are we Explaining Consciousness Yet?, em http://ase.tufts.ed/as/cogstud/ papers/cognition.fin.htm.
DENNETT, Daniel 2000c, Foreword to Darwinizing Culture, em http://ase.tufts.ed/as/cogstud/ papers/aungerfore.pen.htm.
DENNETT, Daniel 2000d, From Typo to Thinko, em http://ase.tufts.ed/as/cogstud/papers/THINKO.htm.
DENNETT, Daniel 2000e, In Darwin's Wake, Where Am I? (APA Presidential Address) em http://ase.tufts.ed/as/cogstud/ papers/apapresadd.htm.
DENNETT, Daniel, 2000f, With a Little Help From My Friends, in ROSS, BROOK & THOMPSON 2000, 327-388.
DENNETT, Daniel & TAYLOR, Christopher 2000, Who's Afraid of Determinism?, em http://ase.tufts.ed/as/cogstud/ papers/kitdraft/htm.
WESTBURY, C. & DENNETT, Daniel, 2000, Mining the past for the future, in D. Schacter & E. Scarry (eds.), *Memory, Brain and Belief*, Cambridge MA, Harvard University Press.
DENNETT, Daniel, 2001, The fantasy of first person science, em http://ase.tufts.ed/as/cogstud/ papers/chalmersdeb3dft.htm.

II. *Bibliografia secundária.*

AKINS, Kathleen, 1993, What is it Like to Be Boring and Myopic? in DAHLBOM 1993, 124-160.
ALLEN, Collin & BEKOFF, Mark, 1997, *Species of Mind: The Philosophy and Biology of Cognitive Ethology*, Cambridge, MA, MIT Press.
ANSCOMBE, G.E.M.,1957, *Intention*, Oxford, Blackwell.
ARMSTRONG, David, 1968, *A Materialist Theory of the Mind*, London, Routledge and Kegan Paul

BAARS, Bernard, 1988, *A Cognitive Theory of Consciousness*, Cambridge, Cambridge University Press.
BAARS, Bernard, 1996, *In the Theater of Consciousness: The Workspace of the Mind*, New York, Oxford University Press.
BAARS, Bernard, 1997, Understanding Subjectivity: Global Workspace Theory and the Resurection of the Observing Self, in SHEAR 1997, 241-247.
BAARS, Bernard & FEHLING, Michael, Consciousness is associated with central as well as distributed processing, *Behavioral and Brain Sciences*, 15 (2), 203-204.
BARON-COHEN, Simon, 1995, *Mind Blindness: An Essay on Autism and Theory of Mind*, Cambridge MA, MIT Press.
BARRETT, Robert & GIBSON, Roger (eds.), 1990, *Perspectives on Quine*, Oxford, Blackwell.
BAUMGARTNER, Peter & PAYR, Sabine (eds.), 1995, *Speaking Minds – Interviews with Twenty Eminent Cognitive Scientists*, Princeton, New Jersey, Princeton University Press.

BECHTEL, William & GRAHAM, George (eds.), 1998, *A Companion to Cognitive Science*, Oxford, Blackwell.
BECHTEL, William, ABRAHAMSEN, Adele & GRAHAM, George, 1998, The Life of Cognitive Science, in BECHTEL & GRAHAM 1998, 2-104.
BEKOFF, Mark, 1998, Cognitive Ethology, in BECTHEL & GRAHAM 1998, 371-379.
BLOCK, Ned, 1993 [1978], Troubles with Functionalism, in GOLDMAN 1993, 231-253.
BLOCK, Ned, 1981, *Imagery*, Cambridge MA, MIT Press.
BLOCK, Ned, 1993, Review of Dennett's *Consciousness Explained*, *Journal of Philosophy* 90, 4, 181-193.
BLOCK, Ned, 1994a, «Consciousness», in GUTTENPLAN 1994, 210-219.
BLOCK, Ned, 1994b, What is Dennett's Theory a Theory of, *Philosophical Topics*, 22, 1-2, 23-57.
BLOCK Ned, 1995, *The Mind as the Software of the Brain*, in SMITH & OSHERSON 1995, 377-425.
BLOCK, Ned, 1997a [1990], Inverted Earth, in BLOCK, FLANAGAN & GÜZELDERE 1997, 677-693.
BLOCK, Ned, 1997b, On a Confusion About the Function of Consciousness, in BLOCK, FLANAGAN & GÜZELDERE 1997, 375-415.
BLOCK, Ned 1997c, Begging the Question Against Phenomenal Consciousness, in BLOCK, FLANAGAN & GÜZELDERE 1997, 175-179.
BLOCK, Ned, FLANAGAN, Owen & GÜZELDERE, Guven (eds.), 1997, *The Nature of Consciousness – Philosophical Debates*, Cambridge, MA, MIT Press.
BODEN, Margaret, 1977, *Artificial Intelligence and the Natural Man*, New York, Basic Books.
BODEN, Margaret (ed.), 1996, *The Philosophy of Artificial Intelligence*, Oxford, Oxford University Press.
BRADDON-MITCHELL, D. & JACKSON, F., 1996, *Philosophy of Mind and Cognition*, Oxford, Blackwell.
BRAITENBERG, Valentino, 1984, *Vehicles – Experiments in Synthetic Psychology*, Cambridge MA, MIT Press.
BRENTANO, Franz, 1995 [1874] *Psychology From an Empirical Point of View* (tradução inglesa de *Psychologie Vom Empyrishen Standpunkt*), London, Routledge.
BRONCANO, Fernando, 1995, El control racional da la conducta in *La Mente Humana*, Madrid, Trotta.
BROOKS, Rodney, 1997, Intelligence without Representation, in HAUGELAND 1997d, 395-420.
BURGE, Tyler, 1979, Individualism and the Mental, *Midwest Studies in Philosophy*, vol. IV, Studies in Metaphysics, Minneapolis, University of Minnesota Press, 73-121.
BURGE, Tyler, 1992, Philosophy of Language and Mind 1950-1990, *The Philosophical Review*, 101, 1, 3-51.
BYRNE, Alex, 1994, «Behaviorism», in GUTTENPLAN 1994, 132-140.

CANTWELL SMITH, Brian, 1996, *On the Origin of Objects*, Cambridge MA, MIT Press.
CARROL Lewis, 1995, *Alice in Wonderland Through the Looking Glass*, Ware, Hutfordshire, Wordsworth.
CHALMERS, David, 1996, *The Conscious Mind: In Search of a Fundamental Theory*, Oxford, Oxford University Press.
CHALMERS, David, 1997a, Facing Up to the Problem of Consciousness, in SHEAR 1997, 9-30.
CHALMERS, David, 1997b, Moving Forward on the Problem of Consciousness, in SHEAR 1997, 379-422.
CHEMERO, Anthony, 1999, *How to be an Anti-Representationalist*, Doctoral Dissertation, Department of Philosophy and Cognitive Science, Indiana University, www.psych.indiana.edu/hyplan/achemero/papers.
CHENEY, Dorothy & SEYFARTH, Robert, 1990, *How Monkeys See The World*, Chicago, University of Chicago Press.
CHERNIAK, Christopher, 1986, *Minimal Rationality*, Cambridge MA, MIT Press.
CHURCHLAND, Paul, 1981, Eliminative Materialism and the Propositional Attitudes, *Journal of Philosophy* 78, 67-90.
CHURCHLAND, Paul, 1992 [1984], *Materia y Consciencia*, Barcelona, Gedisa.
CHURCHLAND, Paul, 1994, «Folk Psychology (2)», in GUTTENPLAN 1994, 308-316.
CHURCHLAND, Paul, 1995, *The Engine of Reason, The Seat of the Soul: A Philosophical Journey into the Brain*, Cambridge MA, MIT Press.
CHURCHLAND, Patricia, 1981, On the Alleged Backwards Referral of Experiences and its Relevance to the Mind-Body Problem, *Philosophy of Science*, 48, 165-181.
CHURCHLAND, Patricia, 1981, The timing of sensations, *Philosophy of Science*, 48, 492-497.
CHURCHLAND, Patricia, 1986, *Neurophilosophy: Toward a Unified Science of the Mind-Brain*, Cambridge MA, MIT Press.
CHURCHLAND, Patricia & RAMACHANDRAN, V., 1993, Filling In: Why Dennett is Wrong, in DAHLBOM 1993, 28-52.
COELHO, Helder, 1995, *Inteligência Artificial em 25 Lições*, Lisboa, Fundação Calouste Gulbenkian.
COOPER, L. & SHEPARD, R., 1973, Chronometric Studies of the Rotation of Mental Images, in CHASE, W. (ed.), *Visual Information Processing* 1973, 75-115.
CLARK, Andy, 1997, *Being There: Putting Brain, Body and World Together Again*, Cambridge MA, MIT Press.
CRICK, Francis, 1994, *The Astonishing Hypothesis: The Scientific Search for the Soul*, New York, Scribner.
CRICK, Francis & KOCH, Christopher, 1990, Towards a neurobiological theory of consciousness, in BLOCK, FLANAGAN & GÜZELDERE 1997, 277-292.

DAHLBOM, Bo, (ed.), 1993, *Dennett and His Critics*, Oxford, Blackwell.

DAHLBOM, Bo, 1993a, Editor's Introduction, in DAHLBOM 1993, 1-12.
DAMÀSIO, António, 1992, The Selfless Consciousness, *Behavioral and Brain Sciences*, 15, 208-209.
DAMÁSIO, António, 1994, *O Erro de Descartes*, Lisboa, Europa-América.
DAMÁSIO, António, 1999, *The Feeling of What Happens: Body and Emotion on the Making of Consciousness*, New York, Harcourt Brace.
DAVIDSON, Donald, 1980, *Essays on Actions and Events*, Oxford, Oxford University Press.
DAVIDSON, Donald 1980a [1963], Actions, Reasons and Causes, in DAVIDSON 1980.
DAVIDSON, Donald, 1980b [1970], Events as Particulars, in DAVIDSON 1980.
DAVIDSON, Donald, 1980c [1976], Hempel on Explaining Action, in DAVIDSON 1980.
DAVIDSON, Donald, 1980d [1970], Mental Events, in DAVIDSON 1980.
DAVIDSON, Donald, 1980e [1974], Psychology as Philosophy, in DAVIDSON 1980.
DAVIDSON, Donald, 1980f [1973], The Material Mind, in DAVIDSON 1980.
DAVIDSON, Donald, 1982, Rational Animals, *Dialectica* 36, 318-327.
DAVIDSON, Donald, 1987, Knowing One's Own Mind, *Proceedings and Addresses of the American Philosophical Association*, 1987, 441-458.
DAVIDSON, Donald, 1984, *Inquiries into Truth and Interpretation*, Oxford, Oxford University Press.
DAVIDSON, Donald, 1984a, Radical Translation, in DAVIDSON 1984.
DAVIDSON, Donald, 1984b, Belief and the Basis of Meaning, in DAVIDSON 1984.
DAVIDSON, Donald, 1994, «Davidson», em GUTTENPLAN 1994, 231-236.
DAWKINS, Richard, 1989 [1976], *O Gene Egoísta*, Lisboa, Gradiva.
DESCARTES, René, 1985 [1641], *Meditações sobre a Filosofia Primeira*, Coimbra, Almedina.
DESCOMBES, Vincent, 1995, *La Denrée Mentale*, Paris, Minuit.
DRETSKE, Fred, 1999 [1981], *Knowledge and the Flow of Information*, Stanford California, CSLI Publications.
DRETSKE, Fred, 2000 [1983], Précis of *Knowledge and the Flow of Information*, in Dretske, F. & Bernecker, S., 2000, *Knowledge – Readings in Contemporary Epistemology*, Oxford, Oxford University Press, 103-117.
DRETSKE, Fred, 1993 [1986], *Misrepresentation*, in GOLDMAN 1993, 297-314.
DRETSKE, Fred, 1988, *Explaining Behavior – Reasons in a World of Causes*, Cambridge MA, MIT Press.
DRETSKE, Fred, 1988a, The Stance Stance, in *Behavioral and Brain Sciences* 11 (3), 511-512.
DRETSKE, Fred, 1990, Does Meaning Matter?, in E. Villanueva (ed), *Information, Semantics and Epistemology*, Oxford, Blackwell, 5-17.
DRETSKE, Fred, 1994 Differences That Make No Difference, *Philosophical Topics*, 22, 1-2, 41-57.
DRETSKE, Fred, 1994, «Dretske» in GUTTENPLAN 1994, 259-265.

DRETSKE, Fred,1995, *Naturalizing the Mind,* Cambridge MA, MIT Press.
DRETSKE, Fred, 1997, Conscious Experience, in BLOCK FLANAGAN & GÜZELDERE 1997, 773-788.
DREYFUS, Hubert, 1972, *What Computers Can't Do,* New York, Harper & Row.
DUMMETT, Michael, 1978, *Truth and Other Enigmas,* Cambridge MA, Harvard University Press.
DUMMETT, Michael, 1978b, Oxford Philosophy, in DUMMETT 1978, 431-436.
DUMMETT, Michael, 1993, *Origins of Analytic Philosophy,* Cambridge, MA, Harvard University Press.
DUMMETT, Michael, 1997, Uma Perspectiva Anti-Realista sobre a Linguagem, o Pensamento, a Lógica e a História da Filosofia Analítica (entrevista por F. Pataut), *disputatio* 3, 22-56.
DUPUY, Jean-Pierre, 1994, *Aux Origines des Sciences Cognitives,* Paris, La Découverte.

EDELMAN, Gerald, 1987, *Neural Darwinism: The Theory of Neuronal Group Selection,* New York, Basic Books.
EDELMAN, Gerald, 1989, *The Remembered Present: A Biological Theory of Consciousness,* New York, Basic Books.
ENGEL, Pascal, 1998, *Introdução à Filosofia do Espírito,* Lisboa, Instituto Piaget.
EVANS, Gareth, 1982, *The Varieties of Reference,* Oxford, Oxford University Press.

FLANAGAN, Owen, 1992, *Consciousness Reconsidered,* Cambridge MA, MIT Press, 1992.
FODOR, Jerry, 1968, *Psychological Explanation,* New York, Random House.
FODOR, Jerry, 1975, *The Language of Thought,* Cambridge MA, Harvard University Press.
FODOR, Jerry, 1983, *The Modularity of Mind, An Essay on Faculty Psychology,* Cambridge MA, MIT Press.
FODOR, Jerry, 1980 [1974], Special Sciences or the Disunity of Science as a Working Hypothesis, in Ned Block (ed.), 1980, *Readings in the Philosophy of Psychology,* vol. I., Cambridge MA, Harvard University Press, 120-133.
FODOR, Jerry, 1981, *Representations – Essays on Cognitive Science,* Cambridge MA, MIT Press.
FODOR, Jerry, 1981a, Methodological Solipsism as a Research Strategy, in FODOR 1981.
FODOR, Jerry, 1987, *Psychosemantics,* Cambridge MA, MIT Press.
FODOR, Jerry, 1990, *A Theory of Content and Other Essays,* Cambridge MA, MIT Press.
FODOR, Jerry, 1990a, Fodor's Guide to Mental Representations, in FODOR 1990.
FODOR Jerry & LEPORE, Ernest, 1992, *Holism: a Shopper's Guide,* London, Blackwell.
FODOR, Jerry & LEPORE, Ernest, 1992, D.C. Dennett and the Normativity of Intentional Ascription (and a little more about Davidson) in FODOR & LEPORE 1992.

FODOR, Jerry, 1994a, *The Elm and the Expert – Mentalese and Its Semantics,* Cambridge MA, MIT Press.
FODOR, Jerry, 1994b, «Fodor», in GUTTENPLAN 1994, 292-300.
FODOR, Jerry, 1998a, *Concepts – Where Cognitive Science Went Wrong,* Oxford, Clarendon Press .
FODOR, Jerry, 1998b, *In Critical Condition: Polemical Essays on Cognitive Science and the Philosophy of Mind,* Cambridge MA, MIT Press.
FODOR, Jerry, 2000, *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology,* Cambridge MA, MIT Press.
FORD, K. & PYLYSHYN, Z., 1996, *The Robot's Dilemma Revisited: The Frame Problem in Artificial Intelligence,* Norwood NJ, Ablex.
FOX, Ivan, 1994, Our Knowledge of the Internal World, *Philosophical Topics,* 22, 1-2, 59-106.
FRANKFURT, Harry, 1971, Freedom of the Will and the Concept of a Person, *Journal of Philosophy,* 68, 5-20.

GARDNER, Howard, 1985, *The Mind's New Science – A History of the Cognitive Revolution,* New York, Basic Books.
GIBSON, James J., 1966, *The Senses Considered as Perceptual Systems,* Boston, Houghton-Mifflin.
GIBSON, James J., 1979, *The Ecological Approach to Visual Perception,* Boston, Houghton Mifflin.
GIL, Fernando, 1998a, Eu, in GIL 1998, 31-49.
GIL, Fernando, 2000, *La Conviction,* Paris, Flammarion.
GLEITMAN, Lila & LIBERMAN, Mark, 1995, *An Invitation to Cognitive Science,* vol. I, Language, Cambridge MA, MIT Press.
GOLDMAN, Alvin (ed.), 1993, *Readings in Philosophy and Cognitive Science,* Cambridge MA, MIT Press.
GOODMAN, Nelson, 1978, *Ways of World Making,* Indianopolis, Indiana, Hackett.
GOULD, S.J. & LEWONTIN, R., 1978 The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptacionist Program, in SOBER 1997, 73-90.
GRICE, H.P., 1957, Meaning, *Philosophical Review* 66, 377-388.
GRIFFIN, Donald, 1976, *The Question of Animal Awareness: Evolutionary Continuity of Mental Experience,* New York, Rockefeller University Press.
GRIFFIN, Donald, 1983, Thinking About Animal Thoughts, *Behavioral and Brain Sciences,* 6 (3), 364.
GRIFFIN, Donald, 1991, Progress Toward a Cognitive Ethology, in RISTAU 1991, 3-17.
GRIFFIN, Donald, 1992, *Animal Minds,* Chicago, University of Chicago Press.
GUTTENPLAN, Samuel (ed), 1994, *A Companion to the Philosophy of Mind,* Oxford, Blackwell.
GÜZELDERE, Güven, 1997, Preface and Acknowledgments, in BLOCK, FLANAGAN & GÜZELDERE 1997, xi-xxix.

HAHN, Edwin e SCHILPP, Paul A. (eds), 1998, *The Philosophy of W.O Quine,* (expanded edition), The Library of Living Philosophers, vol. XVIII, Chicago Illinois, La Salle, Open Court.
HARDCASTLE, Valerie, 1998, Binding Problem, in BECHTEL, Wiilliam & GRAHAM, George 1998, 555-565.
HARDIN, C.L., 1988, *Color for Philosophers – Unweaving the Rainbow,* Indianopolis, Hackett.
HARDIN, C.L., 1993, *Color Subjectivism,* in GOLDMANN 1993, 493-507.
HARMAN, Gilbert, 1995, Rationality, in SMITH & OSHERSON 1995, 175-211.
HAUGELAND, John (ed.), 1981, *Mind Design I,* Cambridge MA, MIT Press.
HAUGELAND, John, 1982, The Mother of Intention, *Noûs,* November, 16 (4), 613-619.
HAUGELAND, John, 1985, *Artificial Intelligence,* Cambridge MA, MIT Press.
HAUGELAND, John, 1993, Patterns and Being, in DAHLBOM 1993, 53-69.
HAUGELAND, John, 1997, *Having Thought: Essays in the Metaphysics of Mind,* Cambridge MA, Harvard University Press.
HAUGELAND, John, 1997a, *Understanding: Dennett and Searle,* in HAUGELAND 1997.
HAUGELAND, John, 1997b, *Mind embodied and embedded,* in HAUGELAND 1997.
HAUGELAND, John, 1997c, *Truth and Rule-Following,* in HAUGELAND 1997.
HAUGELAND, John, (ed.), 1997d, *Mind Design II – Philosophy, Psychology, Artificial Inteligence,* Cambridge MA, MIT Press.
HOBSON, J.A., 1999a, Dreaming, in WILSON & KEIL 1999, 242-244.
HOBSON, J. A., 1999b, Sleep, in WILSON & KEIL 1999, 772-775.
HOFSTADTER, Douglas, 1980, *Gödel, Escher, Bach: An Eternal Golden Braid. A metaphorical fugue on minds and machines in the spirit of Lewis Carroll,* London, Penguin.
HOFSTADTER, Douglas 1981a, Reflections (on J. R. Searle's *Minds, Brains and Programas),* in HOFSTADTER & DENNETT 1981, 373-372.
HOFSTADTER, Douglas, 1981b, Reflections on D. Hofstadter's *Prelude. Ant Fugue,* in HOFSTADTER & DENNETT 1981, 191-201.
HUME, David, 1985 [1739], *A Treatise of Human Nature, Being An Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects,* London, Penguin.
HUME, David 1988 [1779], *Dialogues Concerning Natural Religion and the Posthumous Essays,* Indianopolis, Indiana, Hackett.
HUMPHREY, Nicholas, 1988 [1976], *The social function of intellect,* in WHITEN & BYRNE 1988, 13-26.
HUMPHREY, Nicholas, 1995, *Una Historia de la Mente,* Barcelona, Gedisa.
HUMPHREY, Nicholas, 1997, Review of *Kinds of Minds, Journal of Philosophy,* February 1997, 94 (2), 97-103.

JACKSON, Frank, 1993, Appendix A (for philosophers) in *Philosophical and Phenomenological Research,* 53 (4), November-December, 899-903.
JACKSON, Frank, 1997a [1986], *What Mary Didn't Know,* in BLOCK, FLANAGAN & GÜZELDERE 1997, 567-570.
JACKSON, Frank, 1997b, *Finding the Mind in the Natural World* in BLOCK, FLANAGAN & GÜZELDERE 1997, 483-491.
JAYNES, Julian, 1976, *The Origins of Consciousness in the Breakdown of the Bicameral Mind,* Boston, Houghton Mifflin.
JOHNSON-LAIRD, P., 1983, *Mental Models: Toward a Cognitive Science of Language,* Cambridge, Cambridge University Press.
JOHNSON-LAIRD, P., 1988, A Computational Analysis of Consciousness, in MARCEL & BISIACH 1988, 357-368.
JOHNSON-LAIRD, P., 1988, *The Computer and the Mind – An Introduction to Cognitive Science,* London, Fontana Press.

KIM, Jaegwon, 1996, *Philosophy of Mind,* Boulder, Colorado, Westview Press.
KIM, Jaegwon, 1998, *Mind in a Physical World,* Cambridge MA, MIT Press.
KOSSLYN, Stephen & OSHERSON, Daniel, 1995, *An Invitation to Cognitive Science,* vol. II, Visual Cognition, Cambridge MA, MIT Press.
KOSSLYN, S.M., 1980, *Image and Mind,* Cambridge MA, Harvard University Press.
KOSSLYN, S.M, 1994, *Imagery and Brain: The Resolution of the Imagery Debate,* Cambridge MA, MIT Press.
KOSSLYN, S.M., 1995, Mental Imagery, in KOSSLYN & OSHERSON 1995, 267-296.
KRIPKE, Saul, 1979, A Puzzle about Belief, in A. Margolis (ed.), *Meaning and Use,* Dordrecht, Reidel, 239-283.

LETTVIN, J., MATURANA, H., McCULLOCH, W. & PITTS, W., 1959, What the Frog's Eye Tells the Frog's Brain, in *Proceedings of the Institute of Radio Engineers,* 1940-1951.
LEVELT, W., 1989, *Speaking,* Cambridge MA, MIT Press.
LEVINE, Joseph, 1983, Materialism and the Qualia, *Pacific Philosophical Quarterly,* 64, 354-361.
LEVINE, Joseph, 1997, On leaving out what's like, in BLOCK, FLANAGAN & GÜZELDERE 1997, 543-555.
LEWIS, David, 1966, An Argument for the Identity Theory, *Journal of Philosophy,* 63, 17-25.
LIBET, B., WRIGHT, E., FEINSTEIN, B. & PEARL, D. K., 1979, Subjective Referral of the Timing for a Conscious Sensory Experience, *Brain,* 102, 193-224.
LIBET, B., 1981, The Experimental Evidence for Subjective Referral of a Sensory Experience Backwards in Time: Reply to P.S. Churchland, *Philosophy of Science,* 48, 182-197.
LIBET, B., 1985, Unconscious Cerebral Iniciative and the Role of Conscious Will in Voluntary Action, *Behavioral and Brain Sciences,* 8, 529-566.

LIBET, B., 1989, The Timing of a Subjective Experience, *Behavioral and Brain Sciences* 12, 183-185.
LIBET, B., 1998, Solutions to the Hard Problem of Consciousness, in SHEAR 1997, 301-301.
LOCKE, John, 1999 [1690], *Ensaio Sobre o Entendimento Humano,* Lisboa, Fundação Calouste Gulbenkian (trad. E. Soveral).
LORMAND, Eric 1994, Qualia! Now Showing at a Theater Near You!, in *Philosophical Topics*, 22, 1-2, 127-156.
LOUI, Ronald, 1996, Back to the Scene of the Crime: or, Who Survived Yale Shooting?, in FORD & PYLYSHYN 1996, 89-98.
LUCAS, J.R, 1961, Minds, Machines and Gödel, *Philosophy* 36, 112--127.
LYCAN, William, 1992, UnCartesian Materialism and Lockean Introspection, in *Behavioral and Brain Sciences*, 15 (2), 216-217.

MACHUCO ROSA, António, 2002, *Dos Sistemas Centrados aos Sistemas Acentrados – Modelos da Razão em Ciências Cognitivas, Teoria Social e Novas Tecnologias da Informação*, Lisboa, Vega.
MALCOLM, Norman, 1956, Dreaming and Skepticism, *Philosophical Review*, LXV, 14-37.
MALCOLM, Norman, 1959, *Dreaming*, London, Routledge and Kegan Paul.
MARCEL, Anthony & BISIACH, Edoardo, (eds.) 1988, *Consciousness in Contemporary Science,* Oxford, Oxford University Press.
MARR, David, 1982, *Vision*, San Francisco, Freeman and Co.
McCARTHY, J. & HAYES, P., 1969, Some Philosophical Problems from the Standpoint of Artificial Intelligence, in D. Michie (ed.), *Machine Intelligence*, vol. 4, Edinburgh, Edinburgh University Press, 463-502.
McCULLOCH, W. & PITTS, W. H., 1943, A Logical Calculus of the Ideas Immanent in Nervous Activity, *Bulletin of Mathematical Biophysics*, 5, 115-133.
McGINN, Colin, 1991, *The Problem of Consciousness*, Oxford, Blackwell.
MELO, Paulo Jorge, 1994, *Imanência e Transcendência na Teoria Filosófica de W.V. Quine*, Dissertação de Doutoramento, Lisboa, FSCH-Universidade Nova.
MILLER, George, 1956, The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information, *Psychological Review* 63, 81-97.
MILLIKAN, Ruth, 1984, *Language, Thought and Other Biological Categories*, Cambridge, MA, MIT Press.
MILLIKAN, Ruth, 1993, *White Queen Psychology and Other Essays for Alice,* Cambridge, MA, MIT Press.
MILLIKAN, Ruth, 1993a, Fodor, Dretske and Millikan on Teleosemantics, in MILLIKAN 1993.
MILLIKAN, Ruth, 1993b, Biosemantics, in MILLIKAN 1993.
MILLIKAN, Ruth, 1993c, What is behavior?, in MILLIKAN 1993.
MILLIKAN, Ruth, 1993d, Explanation in Biopsychology, in MILLIKAN 1993.
MILLIKAN, Ruth, 1993e, On Mentalese Ortography, in MILLIKAN 1993 e também em DAHLBOM 1993, 97-123.

MILLIKAN, Ruth, 1993f, In Defense of Proper Function, in MILLIKAN 1993.
MILLIKAN, Ruth, 1993g, The Green Grass Growing All Around: A Philosophical Essay on Ethology and Individualism in Psychology, in MILLIKAN 1993.
MILLIKAN, Ruth, 1996, Pushmi-pullyu Representations, in J. Tomberlin (ed.), *Philosophical Perspectives*, vol. IX, Ridgview Publishing, 185-200.
MILLIKAN, Ruth, 1998, A Common Structure for Concepts of Individuals, Stuffs and Real Kinds: more Mama, more Milk, more Mouse, *Behavioral and Brain Sciences*, 21 (1), 55-65.
MINSKY, Marvin, 1985, *The Society of Mind,* New York, Simon & Schuster.
MOONEY, T. Brian, 2000, Dennett on Ethics: Fitting the Facts Against Greed for the Good, in ROSS, BROOK & THOMPSON 2000, 309-326.

NAGEL, Thomas, 1979, *Mortal Questions,* Cambridge, Cambridge University Press, 1979.
NAGEL, Thomas, 1979a [1974], What is it like to be a bat, in NAGEL 1979.
NAGEL, Thomas, 1979 b, Moral Luck, in NAGEL 1979.
NAGEL, Thomas, 1979c, Panpsiquism, in NAGEL 1979.
NAGEL, Thomas, 1979d, Brain Bissection and the Unity of Consciousness, in NAGEL 1979.
NAGEL, Thomas, 1979e, Subjective and Objective, in NAGEL 1979.
NAGEL, Thomas, 1986, *The View From Nowhere,* Oxford, Oxford University Press.
NAGEL, Thomas, 1995, *Other Minds*, Oxford, Oxford University Press .
NAGEL, Thomas, 1995a, Review of *Content and Consciousness,* in NAGEL 1995.
NAGEL, Thomas, 1995b, Review of *Consciousness Explained,* in NAGEL 1995.
NAGEL, Thomas, 1997, *The Last Word,* Oxford, Oxford University Press.
NEWELL, Allen, 1980, Physical Symbol Systems, *Cognitive Science*, 4, 135-183.
NEWELLL, Allen, 1990, *Unified Theories of Cognition*, Cambridge MA, Harvard University Press.
NIETZSCHE, Friedrich, 1987 [1882], *A Gaia Ciência*, Lisboa, Guimarães Ed.
NOZICK, Robert, 1993, *The Nature of Rationality,* Princeton New Jersey, Princeton University Press.

PAIVIO, Allan, 1971, *Imagery and Verbal Information Processing*, New York, Holt Rinehart and Winston.
PARFIT, Derek, 1984, *Reasons and Persons,* Oxford, Oxford University Press.
PENCE, Greg, 1991, Virtue Theory in SINGER 1991, 249-258.
PENROSE, Roger, 1989, *The Emperor's New Mind – Concerning Computers, Minds and Laws of Physics,* Oxford, Oxford University Press.
PENROSE, Roger, 1994, *Shadows of the Mind: A Search for the Missing Science of Consciousness,* Oxford, Oxford University Press.
PINKER, Steven, 1994,*The Language Instinct,* London, Penguin.
PINTO, João Alberto, 1999, *Superveniência, Materialismo e Experiência,* Dissertação de Mestrado, Faculdade de Letras da Universidade do Porto (no prelo).
PLATÃO, s/d, *Teeteto,* Lisboa, Inquérito.

POLGER, Thomas, 2000, Zombies Explained, in ROSS, BROOK &THOMSON 2000, 259-286.
POPPER, Karl, & ECCLES, John, 1977, *The Self and Its Brain*, Berlin, Springer Verlag.
PREMACK, D., & WOODRUFF, G., Does the Chimpanzee Have a Theory of Mind, *Behavioral and Brain Sciences* 3, 515-526.
PREMACK, David, Does the Chimpanzee Have a Theory of Mind Revisited, in WHITEN & BYRNE 1988, 160-179.
PUTNAM, Hilary, 1975, *Philosophical Papers – Mathematics, Matter and Method*, vol. 1, Cambridge, Cambridge University Press.
PUTNAM, Hilary, 1975, *Philosophical Papers – Mind, Language and Reality*, vol. 2, Cambridge, Cambridge University Press.
PUTNAM, Hilary, 1983, *Philosophical Papers – Realism and Reason*, vol. 3, Cambridge, Cambridge University Press.
PUTNAM, Hilary, [1960]1975a, Minds and machines, in PUTNAM 1975.
PUTNAM, Hilary, [1962]1975b, Dreaming and 'depth grammar', in PUTNAM 1975.
PUTNAM, Hilary, [1963]1975c, Brains and behavior, in PUTNAM 1975.
PUTNAM, Hilary, [1964]1975d, Robots: machines or artificially created life?, in PUTNAM 1975.
PUTNAM, Hilary, [1967]1975e, The nature of mental states, in PUTNAM 1975.
PUTNAM, Hilary, [1967] 1975f, The mental life of some machines, in PUTNAM 1975.
PUTNAM, Hilary, [1975]1975g, The meaning of meaning, in PUTNAM 1975.
PUTNAM, Hilary, 1983a, Why reason can't be naturalized, in PUTNAM 1983.
PUTNAM, Hilary, 1983b, Why there isn't a ready-made world, in PUTNAM 1983.
PUTNAM, Hilary, 1983c, Computational psychology and interpretation theory, in PUTNAM 1983.
PUTNAM, Hilary, 1988, *Representation and Reality*, Cambridge, MA, The MIT Press.
PUTNAM, Hilary, 1992, *Renewing Philosophy*, Cambridge, MA, Harvard University Press.
PUTNAM, Hilary, 1994, «Putnam», in GUTTENPLAN 1994, 507--513.
PYLYSHYN, Zenon, 1979, Do Mental Events Have Durations?, *Behavioral and Brain Sciences*, 2, 277-278.
PYLYSHYN, Zenon, 1984, *Computation and Cognition*, Cambridge MA, MIT Press.

QUINE, W.V., 1953, *From a Logical Point of View*, Cambridge MA, Harvard University Press.
QUINE, W.V., 1953 a, Two Dogmas of Empiricism, in QUINE 1953.
QUINE, W.V., 1953b, On What There Is, in QUINE 1953 .
QUINE, W.V., 1960, *Word and Object*, Cambridge MA, MIT Press.
QUINE, W.V., 1966, Quantifiers and Propositional Objects, in Quine, W.O., 1966, *The Ways of Paradox*, New York, Random House.

QUINE, W.V., 1969, *Ontological Relativity and Other Essays*, New York, Columbia University Press.
QUINE, W.V., 1969 a, Propositional Objects in QUINE 1969.
QUINE, W.V., 1969 b, Epistemology Naturalized in QUINE 1969.
QUINE, W.V., 1985, States of Mind, *Journal of Philosophy* LXXXII, 5-8.
QUINE, W.V., 1987, Indeterminacy of translation, *Journal of Philosophy*, LXXXIV, 5-10.
QUINE, W.V., 1990, *Pursuit of Truth*, Cambridge MA, Harvard University Press.
QUINE, W.V., 1990 a, Three Indeterminacies in BARRETT & GIBSON 1990, 1-16.
QUINE, W.V., 1990 b, Comment on Katz in BARRETT & GIBSON 1990, 198--199.
QUINE, W.V., 1990 c, Comment on Stroud in BARRETT & GIBSON 1990, 334-335.
QUINE, W.V., 1995, *From Stimulus to Science*, Cambridge MA, Harvard University Press.

RAWLS, John, 1971, *A Theory of Justice*, Cambridge MA, MIT Press.
RISTAU, Carolyn, 1991, *Cognitive Ethology*, Hillsdale, Erlbaum.
RORTY, Amélie A., (ed) 1976, *The Identities of Persons*, Berkeley, University of California Press.
RORTY, Richard, 1965, Mind-Body Identity, Privacy and Categories, *Review of Metaphysics*, 19, 24-54.
RORTY, Richard, 1970, Incorrigibility as the Mark of the Mental, *Journal of Philosophy* 69, 203-220.
RORTY, Richard, 1972a, Functionalism, Machines and Incorrigibility, *Journal of Philosophy*, 69, 203-220.
RORTY Richard 1972b, Dennett on Awareness, *Philosophical Studies* 23, 153-162.
RORTY, Richard, 1988, *A Filosofia e o Espelho da Natureza*, Lisboa, Dom Quixote.
RORTY, Richard, 1991a, *Objectivity, Relativism and Truth – Philosophical Papers*, vol. 1, Cambridge, Cambridge University Press.
RORTY, Richard, 1991b, Non Reductive Physicalism, in RORTY 1991.
RORTY, Richard, 1993, Holism, Intrinsicality and the Ambition of Transcendence, in DAHLBOM 1993, 184-202.
RORTY, Richard, 1994 *Consciousness, Intentionality and the Philosophy of Mind*, in WARNER & SUBKT 1994, 121-127.
ROSCH, Eleanor, VARELA, Francisco & THOMPSON, Evan, 1991, *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge MA, MIT Press.
ROSENTHAL, David, 1993, Multiple Drafts and Higher-Order Thoughts, in *Philosophical and Phenomenological Research*, 53 (4), November-December, 911-918.
ROSENTHAL, David, 1994, First-Person Operationalism and Mental Taxonomy, in *Philosophical Topics* 22, 1-2, 319-349.
ROSENTHAL, David, 1997, A Theory of Consciousness, in BLOCK, FLANAGAN & GÜZELDERE 1997, 729-753.

ROSENTHAL, David, 2000, Content, Interpretation and Consciousness, in ROSS, BROOK & THOMSON 2000, 287-308.
ROSS, Don, BROOK, Andrew & THOMPSON, David (eds.), 2000, *Dennett's Philosophy: A Comprehensive Assessment*, Cambridge MA, MIT Press.
ROSS, Don & BROOK, Andrew, 2002, *Daniel Dennett*, Cambridge, Cambridge University Press.
ROVANE, Carol, 1994, The Personal Stance, in *Philosophical Topics*, 22, 1-2, 351-396.
ROWLANDS, Mark, 1997, Teleological Semantics, in *Mind*, 106, 422, 279-303.
RYLE, Gilbert, 1966 [1949], *The Concept of Mind*, London, Penguin.
RYLE, Gilbert, *On Thinking*, Totowa, NJ, Rowman and Littlefield.
RUDDER BAKER, Lynn, 1994a, Content Meets Consciousness, *Philosophical Topics* 22, 1-2, 1-22.
RUDDER BAKER, Lynn, 1994b, Instrumental Intentionality, in S. Stich & T. Warfield (eds.), *Mental Representation*, Cambridge MA, Blackwell.

SCHICK, Frederic, 1997, *Making Choices – A Recasting of Decision Theory*, Cambridge, Cambridge University Press.
SEAGER, William, 1991, *The Metaphysics of Consciousness*, London, Routledge.
SEARLE, John, 1980, Minds, Brains and Programs, *Behavioral and Brain Sciences*, 3 (3), 417-424.
SEARLE, John, 1998 [1992], *A Redescoberta da Mente*, Lisboa, Instituto Piaget.
SEARLE, John, 1997, *The Mistery of Consciousness*, New York, New York Review Book.
SEARLE, John, 1997a, An Exchange With D.C. Dennett, in SEARLE 1997, 115-131.
SELLARS Wilfrid, 1963, Empiricism and the Philosophy of Mind in Sellars, W., *Science, Perception and Reality*, London, Routledge and Kegan Paul.
SEYFARTH, R., CHENEY, D. & MARLER, P., 1980, Monkey Responses to Three Different Alarm Calls: Evidence of Predator Classification and Semantic Communication, *Science*, 210, 801-803.
SHAFIR, E. & TVERSKY, A., 1995, Decision Making in SMITH & OSHERSON 1995, 77-100.
SHEAR, Jonathan (ed.), 1997, *Explaining Consciousness – The Hard Problem*, Cambridge MA, MIT Press .
SHEPARD, R.N. & METZLER, J, 1971, Mental Rotation of Three-Dimensional Objects, *Science*, 171, 701-703.
SHEPARD, R.N. & COOPER, L., 1982, *Mental Images and Their Transformations*, Cambridge MA, MIT Press.
SHOEMAKER, Sidney, 1997, The Inverted Spectrum, in BLOCK, FLANAGAN & GÜZELDERE 1997, 643-662.
SIEWERT, Charles, 1993, What Dennett Can't Imagine and Why, in *Inquiry* 36, 1-2, 93-112.
SIMON, Herbert, 1995, Technology is not the problem, in BAUMGARTNER & PAYR 1995, 231-248.
SIMON, Herbert, 1996 [1969], *The Sciences of the Artificial*, Cambridge MA, MIT Press.

SKINNER, B.F., 1953, *Science and Human Behavior*, New York, Free Press.
SKINNER, B.F., 1971, *Beyond Freedom and Dignity*, New York, Knopf.
SKINNER, B.F., 1974, Behaviorism at Fifty, in T. Wann (ed.), *Behaviorism and Phenomenology: Contrasting Bases for Modern Psychology*, Chicago, University of Chicago Press.
SKINNER, B.F. 1983, A Better Way to Deal With Selection, *Behavioral and Brain Sciences*, 6 (3), 377-378.
SLOTE, Michael, 1994,The Problem of Moral Luck, in *Philosophical Topics*, 22, 1--2, 397-409.
SLUGA, Hans & STERN, 1996, David (eds.), *The Cambridge Companion to Wittgenstein*, Cambridge, Cambridge University Press.
SLUGA, Hans, 1996, Whose House is that? Wittgenstein on the self, in SLUGA & STERN 1996, 320-353.
SMART, J.J.C., 1963, *Philosophy and Scientific Realism*, London, Routledge and Kegan Paul.
SMITH, Edward & OSHERSON, Daniel, 1995, *An Invitation to Cognitive Science*, vol. III (Thinking), Cambridge MA, MIT Press.
SOBER, Elliot, 1993, *Philosophy of Biology*, Oxford, Oxford University Press.
SOBER, Elliot, 1994, *From a Biological Point of View*, Cambridge, Cambridge University Press.
SOBER, Elliot (ed.), 1997, *Conceptual Issues in Evolutionary Biology*, Cambrige, MA, MIT Press.
SOUSA, Ronald de, 1976, *Rational Homunculi*, in RORTY 1976, 217-238.
SOUSA, Ronald de, 1971, How to Give a Piece of Your Mind, or the Logic of Belief and Assent, *Review of Metaphysics*, September 1971.
SPERBER, D. & WILSON D, 1986, *Relevance: A Theory of Communication*, Cambridge MA, Harvard University Press.
STERBERG, Saul & SCARBOROUGH, Donald, 1995, *An Invitation to Cognitive Science*, vol. IV (Conceptual Foundations), Cambridge MA, MIT Press.
STICH, Stephen, 1993, *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation*, Cambridge MA, MIT Press.
STILLIGS, N., WEISLER, S., CHASE, C., FEINSTEIN, M., GARFIELD, J. & RISSLAND, E., 1995, *Cognitive Science – An Introduction*, Cambridge MA, MIT Press.
STROUD, Barry, 1990, Quine's Physicalism, in BARRETT & GIBSON 1990, 321-333.
SUTHERLAND, Stuart, 1992, *Irrationality: Why We Don't Think Straight*, New Brunswick, Rutgers University Press.

TAYLOR, Charles, 1964, *The Explanation of Behavior*, London, Routledge and Kegan Paul.
TAYLOR, Charles, 1976, *Responsability for Self* in RORTY 1976, 281-299.
TAYLOR, Charles, 1992, *The Sources of Self*, Cambridge MA, Harvard University Press.
TAYLOR, Charles, 1992, *The Ethics of Authenticity*, Cambridge MA, Harvard University Press.

TAYLOR, Charles, 1996, *Human Agency and Language – Philosophical Papers vol. I*, Cambridge, Cambridge University Press.
TAYLOR, Charles, 1996, *Philosophy and the Human Sciences – Philosophical Papers*, vol. II, Cambridge, Cambridge University Press.
THOMPSON, David, 2000, Phenomenology and Heterophenomenology, in ROSS, BROOK & THOMPSON 2000, 201-218.
TOLLIVER, Joseph Thomas, 1994, Interior Colors, in *Philosophical Topics* 22, 1-2, 411-441.
TURING, Alan, 1950, Computing Machinery and Intelligence, *Mind*, 59, 433-460.
TVERSKY, Amos & KAHNEMAN, Daniel, 1981, The Framing of Decisions and the Psychology of Choice, *Science* 211, 453-458.
TVERSKY, Amos & KAHNEMAN, Daniel, 1982,Judgment under Uncertainty: Heuristics and Biases, in Kahneman, D., Slovic, P. & Tversky, A. (eds.), *Judgment under Uncertainty: Heuristics and Biases*, Cambridge, Cambridge University Press.
TVERSKY, Amos & KAHNEMAN, Daniel, 1993, *Probabilistic Reasoning*, in GOLDMAN 1993, 43-68.
TYE, Michael, 1993, Reflections on Dennett and Consciousness, in *Philosophical and Phenomenological Research*, 53 (4), November-December, 893-898.

VAN GULLICK, Robert, 1997, Time for More Alternatives, in BLOCK, FLANAGAN & GÜZELDERE, 181-183.

WARNER, R., & SUBKT, T. (eds.), 1994, *The Mind-Body Problem*, Oxford, Blackwell.
WEISKRANZ, Larry, 1986, *Blindsight: A Case Study and Implications*, Oxford, Oxford University Press.
WEISKRANZ, Larry, 1997, *Consciousness Lost and Found – A Neuropsychological Exploration*, Oxford, Oxford University Press.
WHITE, Stephen, 1994, Color and Notional Content, in *Philosophical Topics* 22, 1-2, 471-503.
WHITEN, A. & BYRNE, R, (eds.) 1988, *Machiavellian Intelligence – Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans*, Oxford, Clarendon Press.
WILLIAMS, Bernard, 1981, *Moral Luck*, Cambridge, Cambridge University Press.
WILSON, R. & KEIL, F. (eds.) 1999, *The MIT Encyclopedia of Cognitive Science*, Cambridge, MA, MIT Press.
WITTGENSTEIN, Ludwig, 1987 [1953], *Investigações Filosóficas*, Lisboa, Fundação Calouste Gulbenkian (trad. M.S.Lourenço).

YATES, J.F. & ESTIN, P.A., 1998, Decision Making in BECTHEL & GRAHAM 1998, 186-196.

ZILHÃO, António, 1993, *Linguagem da Filosofia e Filosofia da Linguagem – Estudos sobre Wittgenstein*, Lisboa, Colibri.
ZILHÃO, António, 1998/99, Folk Psychology, Rationality and Human Action, in *Grazer Philosophische Studien*, 56, 1-28.

Índice Onomástico

Anscombe, G. E. M., 11, 35, 71, 72, 97, 98, 118-127, 574
Armstrong, D., 46, 51, 574
Ayer, A.J., 52

Baars, B., 38, 253, 254, 279, 389, 574
Babbage, C., 34
Baron Cohen, S., 457, 574
Berkeley, G., 45, 135
Block, N., 154, 171, 250, 262, 263, 271, 272, 297, 316, 317, 336, 337, 575
Boole, G., 34
Brentano, F., 61, 65, 512, 575
Brooks, R., 443, 550, 575
Burge, T., 189, 192, 575, 576

Cantwell Smith, B., 17, 480, 482, 500, 501, 502, 503, 504, 505, 506, 507, 508, 576
Chalmers, D., 15, 27, 38, 50, 51, 159, 251, 260, 269, 270, 271, 273, 274, 275, 276, 277, 278, 288, 382, 384, 386, 387, 388, 481, 528, 576
Changeux, J.P., 80, 341
Cheney, D., 450, 452, 455, 456, 460, 462, 576
Chomsky, N., 33, 99, 138, 140, 247, 440
Church, A., 34
Churchland, Patricia S., 34, 48, 55, 111, 137, 151, 180, 200, 239, 332, 450, 488, 532, 576
Churchland, Paul M., 27, 34, 137, 151, 180, 200, 211, 239, 488, 532, 576
Clark, A., 154, 488, 576
Crick, F., 38, 252, 253, 254, 576

Dahlbom, B., 48, 49
Damásio, A., 256-259, 473, 534, 577
Davidson, D., 26, 119, 121, 172, 173, 184, 198, 218, 231, 240, 419, 457, 510, 512, 515, 516, 517, 542, 548, 549, 577
Dawkins, R., 192, 193, 341, 450, 452, 453, 577
Descartes, R., 28, 34, 97, 257, 405, 537, 538, 577
Descombes, V., 438, 577
Dretske, F., 14, 27, 28, 51, 108, 135, 139, 189, 190, 192, 201, 204, 205, 206, 207, 218, 225-230, 233, 234, 238, 239, 240, 500, 521, 528, 577, 578
Dreyfus, H., 115, 447, 490, 491, 578
Dummett, M., 31, 32, 52, 55, 187, 208, 449, 578
Dupuy, J.P., 33, 553, 578
Eccles, J., 327, 584
Edelman, G., 38, 80, 251, 254-256, 338, 341, 578
Espinosa, B., 28, 550, 556
Evans, G., 209, 356

Flanagan, O., 270, 282, 578
Fodor, J., 13, 14, 25, 27, 28, 30, 31, 34, 36, 64, 66, 93, 130, 136, 137, 139, 140-146, 147, 148, 149, 151, 154, 168, 171, 180, 189, 192, 197, 200, 201-206, 210-213, 218, 224, 225, 226, 229, 230-238, 239, 243, 247, 271, 277, 298, 299, 347, 419, 440, 484-485, 496, 500, 502, 503, 509, 510, 521, 547, 549, 554, 578, 579
Frankfurt, H., 420, 423, 579
Frege, G., 31, 208, 232, 234, 235, 236

Gibson, J., 154, 488, 579
Gil, F., 9, 554, 579
Goodman, N., 325, 329, 330, 579
Gould, S.J., 189, 247, 450-453, 579
Grice, P., 417, 418, 419
Griffin, D., 443, 445, 446, 450, 451, 454, 579

Hardin, C., 374, 375, 376
Haugeland, J., 17, 195-200, 490-500, 508, 580
Hayes, P., 446, 464, 582
Hebb, D., 72
Hobbes, T., 45, 395
Hume, D., 28, 34, 46, 47, 145, 246, 284, 397, 465-468, 537
Humphrey, N., 369, 374, 378, 413-414, 450, 451, 453, 473-475

Jackson, F., 15, 27, 38, 267, 268, 269, 270, 271, 272, 275, 278, 381, 382, 383, 431, 487, 581
Jaynes, J.,316, 344, 409, 581
Johnson-Laird, P., 279

Kant, I., 28, 34, 283, 400, 405, 409, 411, 425, 496
Kaplan, D., 210
Kim, J., 27, 274, 481
Kinsbourne, M., 329, 337
Koch, C., 38, 252, 576
Kosslyn, S., 107, 108, 297, 299, 300, 301, 302, 303, 304, 307, 361-363

Leibniz, G., 28, 326, 538
Levine, J., 251, 278
Lewis, D., 46, 51, 581
Libet, B., 325-327, 330-337
Locke, J., 236, 392, 415, 582

Machuco, A., 9, 550, 582
Marcel, A., 103, 355
McCarthy, J., 33, 446, 464
McGinn, C., 270, 271, 582
Miller, G., 33
Millikan, R., 13, 14, 27, 28, 118, 154, 192, 201, 204, 205, 206, 207, 209, 216-225, 227, 229, 236, 238, 239, 240, 449, 450, 488, 506
Minsky, M., 33, 311, 315, 338, 351, 355, 356, 372, 422, 440, 531, 548
Moore, G.E., 52

Nagel, T., 7, 15, 31, 46, 48, 49, 74, 75, 131, 159, 269-271, 278, 295, 319, 360, 384, 387, 391, 425, 426, 443, 445, 547, 555
Newell, A., 33, 72, 73, 87, 115, 338, 548
Nietzsche, F., 32, 249, 352, 395
Nozick, R., 158, 157, 175, 394, 404, 512, 520

Parfit, D., 289, 394, 395, 583
Penrose, R., 38, 251, 259, 260, 583
Pinker, S., 418, 583
Pinto, J.A., 9, 37, 50, 262, 263, 273, 274, 295, 540, 545
Platão, 350, 583
Premack, D., 450, 451
Putnam, H., 11, 12, 30, 34, 41, 46, 51, 66, 67-70, 82, 96, 99, 130, 135, 140, 143, 146, 173, 179, 189, 190, 192, 210, 213, 233, 234, 290, 291, 537, 544-548
Pylyshyn, Z., 108, 195, 283, 297, 298, 299, 302-307, 362, 472, 584

Quine, W. O., 11, 13, 27, 35, 41, 48, 51, 55, 61-66, 70, 76-79, 137, 142, 164, 167, 173, 198, 209, 213, 216, 446, 511, 512, 513, 514, 537, 540, 543-545, 548, 549, 584-585

Rawls, J., 186, 419
Ristau, C., 450, 585
Rorty, R., 101, 106, 131, 151, 243, 265, 487, 488, 490, 492, 493, 499, 500, 528, 532, 585
Rosenthal, D., 51, 103, 272, 364-366
Rovane, C., 416, 419, 420, 421, 435
Rudder Baker, L., 211, 214, 215, 244, 319

Ryle, G., 11, 28, 35, 48, 49, 51-60, 70, 77, 94, 95, 97, 98, 112, 113, 114, 121, 131, 141, 142, 145, 158, 318, 350, 487, 543, 551, 586

Searle, J., 15, 17, 27, 34, 38, 48, 73, 139, 179, 189, 194, 195, 196, 197, 251, 253, 260, 269-271, 276, 278, 295, 345, 383, 384, 387, 437, 490-498, 504, 533, 538, 558, 589
Seyfarth, R., 450, 452, 455, 456, 460-462
Shepard, R., 108, 284, 298, 362
Simon, H., 33, 72, 73, 87, 115, 152, 163, 185, 437-439, 441, 473, 517, 586
Skinner, B. F., 138, 163, 164, 165, 407, 450, 452, 454
Smart, J. J. C., 46, 51, 587
Sober, E., 457, 587
Sousa, R., 392, 587
Stich, S., 181, 182, 183, 187, 210, 405, 514, 518, 554, 587

Taylor, C., 11, 35, 70, 71, 72, 164, 396, 422-425, 543, 559
Thorndike, E., 49, 138
Turing, A., 34, 47, 66-69, 118, 143, 234, 338, 393, 438
Tversky, A., 153, 515, 516, 588

Weiskranz, L., 38, 253, 588
Williams, B., 425, 426, 588
Wittgenstein, L., 11, 28, 31, 35, 48, 49, 51-60, 70, 95, 97-99, 111, 131, 141, 143, 145, 216, 289, 291, 292, 446, 450, 543, 551, 588
Woodruff, G., 450, 451, 584

Zilhão, A., 516, 588

Índice Temático

abstractum, 359
acção intencional, 118-127
actos de fala, 15, 281, 282, 313, 318, 351-356, 365-366
adaptacionismo, 194, 242, 243, 452-453
agentes, 23, 40, 93, 114, 121, 126, 137, 152, 153, 160, 182-186, 195, 198, 284, 311, 312, 316, 352, 354, 356, 358, 360, 391, 393, 396, 401-402, 405, 407-411, 420, 426-428, 431-433, 446, 463
analógico, 297, 298, 299, 300, 303, 304
Argumento do Conhecimento, 38, 275, 383, 387
artificial, 16, 39, 73, 115, 163, 245, 309, 310, 440-442, 555
autismo, 457
auto-consciência, 255, 256, 415, 417-421, 435, 492, 519, 538, 557
auto-controlo meta-reflexivo, 16, 39, 406, 421, 422-426, 520, 556-557
auto-engano, 170, 360, 392, 511
auto-monitorização, 67, 193, 365, 430-526
auto-representação, 358, 412, 413, 445, 492, 517, 557
avaliação forte, 16, 396, 422-424, 556

barreira da fusão, 74
behaviorismo, 13, 48, 49, 59, 60, 71, 72, 86, 137-139, 159, 162-165, 171, 272-273, 444, 458
behaviorismo lógico, 49, 142, 152, 154, 305

cérebro dividido, 360
cibernética, 33, 72
coelho cutâneo, 325, 328, 331
compatibilismo, 15, 391, 399
comportamento dirigido por fins, 86-89
conexionismo, 444, 550
conexões aferentes-eferentes, 82-83
consciência de intenção, 325, 333-334
continuismo, 385, 522
controlo, 12, 16, 101-107, 351-357, 403-405, 406-410, 422, 434
cor, 111-112, 373-380

deliberador, 16, 426-433
descrições conceptuais e explicações causais, 35
desordem de personalidade múltipla, 277, 414
determinismo, 16, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 420, 427
disposições comportamentais, 54, 154
dor, 12, 15, 57, 68-69, 93-94, 97, 98, 99, 308-310, 318

Efeito Baldwin, 338, 342, 344
eliminativismo, 14, 180, 200-201, 238-241, 262, 273, 389
eliminativista, 24, 64, 110, 151, 239, 261, 390
embebimento, 21, 179, 534
empirismo, 142, 538, 548, 550
engenharia invertida, 16, 34, 437, 440
entendimento genuíno, 40, 241, 383, 480, 497-498, 521, 522
epifenomenismo, 28, 380, 381, 382
epistemologia naturalizada, 27, 62, 540

erro categorial, 52, 53, 74, 79, 121
esboços múltiplos, 15, 19, 25, 37, 213, 250, 252, 312, 313, 315, 316, 322-324, 330, 356, 357, 360, 389, 475, 524
ética da virtude, 16, 429
etologia cognitiva, 17, 30, 39, 445, 448-462, 535
eu, 16, 16, 29, 34, 38, 40, 49, 256-258, 279-280, 357-360, 393-395, 414, 422, 492, 526, 529-531, 553, 557
evolução no cérebro, 80, 88, 278, 340-343
evolucionismo, 13, 14, 48, 90, 135, 163, 245, 451

fenómeno phi, 325, 328, 329, 333
fenomenologia, 31, 37, 94, 131, 142, 159, 212, 256, 255-266, 293, 305, 307, 311, 317-322, 337, 388, 447, 471, 523-526, 529, 551, 554
Fido, 90, 91, 510, 513
frases de observação, 63, 64
frases permanentes, 64
free floating rationales, 87, 411, 452, 520
funcionalismo, 12, 17, 25, 30, 35, 36, 39, 46, 48, 67-69, 96-101, 135, 140, 146, 147, 154, 201, 250, 274, 275, 348, 370, 393, 437, 473-475, 484, 534, 543, 545, 546, 547

geografia lógica, 52, 61
gradualismo, 392, 429, 434, 503

harmonia pré-estabelecida, 28, 468
heterofenomenológico, 36, 214, 278, 306, 317-322, 482, 528-531
hipóteses analíticas, 64
holismo, 11, 13, 75, 96, 131, 165, 168, 221, 231, 238, 413, 497, 548

IA forte, 73
Ícones Intencionais, 192, 204, 208, 209, 216-225, 238, 239, 241, 488
idealismo, 45, 548, 549, 551
identidade pessoal, 23, 39, 257, 258, 357-360, 392-395, 411-414, 422-426, 480, 497, 542
ilusão do utilizador, 15, 17, 311-315, 337, 347-351, 445, 520, 523, 526, 527, 530, 535
imagens mentais, 12, 15, 107-112, 296-308, 361-363
imaterialismo, 537, 556
incorrigibilidade, 12, 25, 36, 98-103, 159, 165-166, 262, 265, 266, 285, 291, 294, 493, 513, 524, 527, 532
indeterminação da tradução, 64-65, 164
infalibilidade, 159, 261, 262, 265, 285, 286, 532
informação, 225-227
instrumentalismo, 13, 25, 146, 154, 171, 201, 205, 230, 244
intelectualismo, 54, 370, 372, 389, 524, 527, 528
Intencionador Central, 313, 314, 352-354, 363, 373
interface, 17, 346-349, 438-439, 482, 542-543, 557
interpretação, 13, 14, 24, 26, 39, 40, 61-65, 159-170, 180-194, 238-244, 445, 479-520, 522
interpretação radical, 26, 198, 512
introspecção, 57-60, 96-107, 116, 190, 214-215, 264, 282, 285, 361, 372, 469

justiça, 419, 426

lei do efeito, 49, 138
leis *ceteris paribus*, 148, 484
leis estritas, 121, 484
ligações virtuais, 343
linguagem do pensamento, 13, 26, 67, 94, 136, 137, 140-146, 180, 208-215, 354, 485

manual de tradução, 459
Máquina de Turing, 68-70, 147, 345-346
máquina semântica, 25, 154, 405
Máquina Virtual, 15, 17, 276, 311, 337, 345-356, 363, 384, 520, 523, 526, 530
materialismo, 45, 51, 52, 67, 268, 269, 273, 274
materialismo eliminativo, 137

maximização racional, 395-396, 405, 410, 429-433
medium imagético, 300, 361-363
memes, 341, 343, 345
mentes animais, 17, 39, 448-462, 535
meta-contraste, 325, 328, 330
metafísica, 14, 27, 31, 38, 51, 200, 251, 262, 268, 273, 316, 360, 382, 384-389, 480-490, 501, 505, 538-540, 542, 545, 548, 550, 553, 557
Modelo dos Esboços Múltiplos, 322-324
Modos de Apresentação, 205, 234
módulo, 279
monismo, 28, 52, 68, 71, 506, 508, 544-550
morcego, 15, 38, 269, 275, 380, 384
mundo nocional, 207-216, 319, 386, 459, 482, 506, 528, 531

naturalização da epistemologia, 30, 540, 548
neurocriptógrafo, 168-170
niilismo do significado, 24, 25, 36, 64
novo misteriosismo, 270

ontologia, 17, 32, 39, 66, 75, 78, 172, 197, 200, 205, 269, 308, 479, 480, 486, 490, 500, 501, 548
origem dos objectos, 500-508
Otto, 313, 317, 352, 368, 369, 372, 379, 385

padrões, 13, 25, 36, 176-177, 180, 197-201, 247, 363, 377, 400, 409-412, 486, 494, 498, 503-508
pandemónio, 15, 279, 316, 343, 356
pensamentos de ordem superior, 364
pessoalidade, 50, 120, 392, 414-425, 433-435, 479, 517, 556-559
possibilidade epistémica, 16, 407, 427, 429
preenchimento, 107-112, 303, 329, 370-372
problema da indução, 466-468
problema de Hume, 46, 47, 145, 284
problema de Libet, 333
problema do enquadramento, 17, 21, 39, 446-447, 462-472, 514
problema dos pássaros de Platão, 350
produção da fala, 294, 353-356
produção de futuro, 339-340, 464, 471
programa, 74, 88, 116, 155-157, 196, 276, 308, 346-348, 404, 437, 442
protocolos humanos, 115-118
prudência naturalizada, 39, 396, 429-433
psicologia cognitiva sub-pessoal, 153, 184, 357, 485-486
psicologia de senso comum, 122, 151-154, 184, 312, 484-485
psicologismo, 31, 553
psicologista, 214, 482, 543

qualia, 12, 14, 15, 25, 35, 37, 38, 46, 107-112, 251, 261-267, 269, 272, 324, 361, 366, 368, 373-389, 480, 522, 527
Quarto Chinês, 194-197, 380-384, 496, 558
querer-dizer, 15, 280, 294, 317, 351-355, 364, 417-418

racionalidade, 13, 17, 23, 24, 26, 40, 93, 140, 145, 152, 157, 159-170, 181-186, 201, 222-225, 242-243, 257, 276, 409, 410-414, 419, 421, 431, 437-440, 509-520, 554
rapaz que vende limonada, 181
rastros da máquina, 115
referência para trás no tempo, 325, 330, 331
responsabilidade, 16, 123-125, 419-421, 422-428, 559
revisões orwellianas e estalinistas, 327
robótica de Brooks, 550
satisficing, 152, 185, 429-430, 439, 489, 515
self, 16, 21, 36, 256-258, 358, 393-394, 410-414, 422-423, 542
sentimento de si, 25, 29, 256-259, 414, 435, 527, 554, 556-557
serialidade, 276, 314, 345, 349, 351, 357, 526

sistemas dinâmicos, 550
sonhos, 14, 38, 286-296, 305, 313, 525
sorte moral, 424-426
superveniência, 51, 268, 272-275, 384, 385, 389, 524, 540, 545, 546
swampman, 231

Teatro Cartesiano, 14, 37, 249, 361, 266-267, 311, 312, 314, 351, 361, 363, 367, 388, 445, 479, 480, 492, 526-528, 533, 535
teleofuncionalismo, 13, 24, 36, 70, 201-207, 215, 225, 229, 230, 240, 534
teoria da identidade, 35, 46, 69, 74, 130, 146, 147
teoria deflacionária, 12, 24, 95, 105, 276, 314, 461, 480
terapia conceptual, 55, 61
termostato, 172, 175, 178, 179, 288, 404, 494
Terra Gémea, 143, 179, 190, 191, 210, 212, 213, 233
testemunha, 15, 46-48, 109, 296, 367, 383
transcendental, 24, 29, 544, 548

utilitarismo, 395, 409, 430, 433

verificacionismo, 48, 97, 252, 274, 275, 287, 289-292, 324, 522, 533
virtual, 15, 16, 17, 39, 276, 311, 314, 337, 343-351, 370-372, 393, 410, 444, 445, 448, 479, 492, 525-526, 529-530
visão cega, 15, 38, 253, 277, 355, 365, 367-370, 527

zombie, 38, 49, 50, 253, 271-275, 295, 352, 361, 365, 366, 367, 379, 380, 382, 383