

<b>Sumário</b>	<b>Noema e percepção na fenomenologia de E. Husserl</b>	<b>3</b>
	Pedro M. S. Alves	
	<b>Do princípio da indiscernibilidade dos idênticos ao princípio da identidade de essência</b>	<b>69</b>
	Antônio Pedro Mesquita	
	<b>Pessoas e Acções – A filosofia da mente e os fundamentos cognitivos da filosofia moral</b>	<b>81</b>
	Sofia Miguens	
	<b>O Eu como humanidade singular em Montaigne</b>	<b>135</b>
	Adelino Cardoso	
	<b>L'amphibologie des concepts de la réflexion: appendice ou pièce maîtresse de la <i>Critique de la raison pure</i>?</b>	<b>149</b>
	Michèle Cohen-Halimi	
	<b>NOTAS DE LEITURA</b>	
	<b>Afecto, formas, evidência. O retorno a Hipócrates em filosofia da medicina. Sobre <i>O Espelho Declinado</i>, de Manuel Silvério Marques</b>	<b>161</b>
	Paulo Tunhas	
	<b>John Searle, <i>The Mystery of Consciousness</i>, The New York Review of Books, New York, 1997, 224 pp.</b>	<b>167</b>
	Maria Cerezo	
	<b>ABSTRACTS   RESUMÉS</b>	<b>171</b>

**análise** 23 | 2002  
Revista Quadrimestral de Filosofia

**Directora** Maria Luísa Couto Soares

**Investigador Responsável** Fernando Gil

**Administrador** Manuel Villaverde Cabral

**Secretariado da redacção** Adelino Cardoso e Paulo Tunhas

**Redacção** Adelino Cardoso, António Braz Teixeira, Diogo Pires Aurélio, Fernando Gil, Manuel Barroso Silvério Marques, Manuel Villaverde Cabral, Maria José Cantista, Maria Luísa Couto Soares, Marta de Mendonça, Paulo Tunhas.

**Propriedade do título** Gabinete de Filosofia do Conhecimento

**Registo de título e propriedade** 110517

**ISSN** 0870-3124

© Gabinete de Filosofia do Conhecimento  
Biblioteca Nacional 1751 Lisboa

**Editor** Campo das Letras, Editores, SA

Rua D. Manuel II, 33 - 5.º

4050-345 Porto

Telef.: 22 600 77 28 Fax: 22 600 40 19

Site: [www.campo-letras.pt](http://www.campo-letras.pt)

E-mail: [campo.letras@mail.telepac.pt](mailto:campo.letras@mail.telepac.pt)

Todos os assuntos relativos a assinaturas devem ser tratados com o editor

**Preço de capa** € 10

**Assinatura anual** € 23,94

**Professores de Ensino Secundário e Estudantes** € 20,95

**Grafismo** Loja das Ideias

**Impressão** Papelmunde SMG, Lda. - V.N. Famalicão

**Depósito Legal** n.º 36973

**Distribuidora** ECL - Empresa de Comércio Livreiro

**Tiragem** 1 000 exemplares

**Data da edição** Março de 2002

Pede-se permuta

Pidese canje

On demande l'échange

Exchange requested

Man bittet um Austausch

Publicação patrocinada por:

**FCT**

Fundação para a Ciência  
e a Tecnologia  
CENTRO NACIONAL DE INVESTIGAÇÃO EM TECNOLOGIA



FUNDAÇÃO ENG. ANTÓNIO DE ALMEIDA

# Pessoas e Acções – A filosofia da mente e os fundamentos cognitivos da filosofia moral<sup>1</sup>

Sofia Miguens\*

"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, *Moral Luck*<sup>2</sup>

## 1. As noções de Pessoa e de Acção a partir das teorias do conteúdo e da consciência de D. Dennett.

As noções que sustentam a filosofia moral – noções como a unidade e a continuidade do agente, a voluntariedade e a racionalidade da acção – estão suspensas de esclarecimentos que cabem à teoria da mente. O objectivo do presente artigo é identificar as consequências na filosofia moral,<sup>3</sup> nomeadamente nas noções de pessoa

---

\* Faculdade de Letras da Universidade do Porto

<sup>1</sup> O presente artigo corresponde quase totalmente ao Capítulo 4 da dissertação de doutoramento intitulada *Uma teoria fiscalista do conteúdo e da consciência – D. Dennett e os debates da filosofia da mente* (Faculdade de Letras da Universidade do Porto, 2001), orientada pelo Professor Doutor Fernando Gil. A 'extracção' do capítulo envolve, como é óbvio, alguma artificialidade. Nomeadamente, as questões tratadas não são antecedidas pela exposição das teorias dennettianas do conteúdo e da consciência (respectivamente uma teoria quiniana da interpretação e um conjunto de argumentos deflacionários em torno de um modelo psicológico funcionalista e dos conceitos de *qualia* e de *zombie*). Existe no entanto uma certa autonomia dos temas tratados e é nela que se aposta.

<sup>2</sup> In T. NAGEL 1979, *Mortal Questions*, p. 38.

<sup>3</sup> A filosofia moral é entendida como consistindo em tentativas concretas de resposta à questão 'Como se deve viver?' bem como em análises dos factores que possibilitam as respostas, como é o caso, precisamente, da noções de pessoa e de acção. Não se discutirá portanto as distinções frequentemente feitas entre a ética (considerada como reflexão anterior à moral e moralmente neutra, cf. B. WILLIAMS 1985, *Ethics and the Limits of Philosophy*) e a moral, que estabelecerá princípios universalmente vinculativos. As análises apresentadas terão a certo ponto uma dimensão (quase) prescritiva e não apenas descritiva, não sendo assim de qualquer modo desadequado o termo 'moral'. De acordo com uma outra definição do âmbito da reflexão ética ou moral esta diz respeito às entidades abrangidas pelo uso da palavra 'nós'. O problema é que o 'nós' não está definido *a priori*, nada garantindo aliás que ele seja claramente delimitado: aquilo que define o 'nós' está precisamente dependente de teorias da intencionalidade, da racionalidade e da consciência e portanto dos problemas discutidos na filosofia da mente. Procurar saber que entidades além dos humanos pertencem ao círculo do 'nós' é no entanto uma extensão a que o ético naturalista é obrigado, dado os seus compromissos teóricos (supõe-se que de um ponto de vista naturalista nada distingue em exclusivo e desde logo os humanos como únicas entidades éticas).

e de acção, da teoria da mente desenvolvida pelo filósofo americano D. Dennett, examinando o prolongamento das descrições cognitivas em posições propriamente filosóficas. Existe de resto toda uma corrente na obra de Dennett que se ocupa de questões de razão prática<sup>4</sup> e na qual se esboça – a partir das teorias do conteúdo e da consciência – uma posição acerca da acção, das condições de pessoalidade e da identidade pessoal a que se chamará naturalista e gradualista<sup>5</sup> e que se procura aqui expor.

Na obra de Dennett o mental é abordado através de uma dupla via: de um lado ficam as questões do conteúdo, tratadas através da Teoria dos Sistemas Intencionais (TSI), uma teoria quiniana da interpretação supondo a racionalidade<sup>6</sup>; do outro lado ficam as questões da consciência, abordadas através de modelos psicológicos funcionalistas cuja interpretação filosófica é, na sua forma última, o Modelo dos Esboços Múltiplos<sup>7</sup>. Embora as noções de pessoa e de acção decorram sobretudo da teoria do conteúdo mental, elas apoiam-se numa teoria cognitiva do eu, do controlo e da consciência elaborada à luz do Modelo dos Esboços Múltiplos.

A primeira consequência geral que decorre da teoria do conteúdo é a normatividade dos conceitos de pessoa e de acção. O aspecto normativo do conceito de pessoa diz respeito às condições de pessoalidade (*conditions of personhood*)<sup>8</sup> e à identidade pessoal. Entende-se por condições de pessoalidade determinadas características em

---

<sup>4</sup> Essa corrente vai desde o tratamento de pessoas e de intenções em *Content and Consciousness* (DENNETT 1986, 1.ª ed. 1969), à Parte IV de *Brainstorms* (DENNETT 1981) intitulada *Free Will and Personhood*, na qual aparecem dois artigos que são usualmente apontados por Dennett como sendo fulcrais para a sua posição geral, os artigos *Mechanism and Responsibility* e *Conditions of Personhood*, a *Elbow Room* (DENNETT 1983) e *Darwin's Dangerous Idea* (DENNETT 1995).

<sup>5</sup> Cf. C. ROVANE 1994, *The Personal Stance*.

<sup>6</sup> Cf. por exemplo D. DENNETT 1987, *The Intentional Stance*, e DENNETT 1991, *Real Patterns*.

<sup>7</sup> Cf. por exemplo D. DENNETT 1991, *Consciousness Explained*.

<sup>8</sup> O termo *personhood* é especificado sobretudo no artigo *Conditions of Personhood* (DENNETT 1981b). No entanto já desde *Content and Consciousness* Dennett defende que o problema da mente não pode ser separado do problema das pessoas (p.189) e que a eficácia de razões no controlo da acção de uma entidade é fundamental para a própria existência de pessoas, mesmo se a voluntariedade envolvida deve ser concebida como consistindo (apenas) em apercebimento sem observação, ao nível pessoal, do conteúdo dos controlos eferentes e não em alguma característica especial a mais. Em *Brainstorms* surge 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 1981c, *How to Change Your Mind*). A intenção do artigo *How to Change Your Mind* no qual essa distinção é proposta é marcar a diferença entre o estado cognitivo atribuído no âmbito da TSI aos Sistemas Intencionais (SI) em geral, a crença, e a opinião, exclusivamente presente na 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. Ronald de SOUSA 1971, *How to Give a Piece of Your Mind*) obtém-se que é o assentimento (*assent*) que caracteriza aquilo que é

função das quais uma entidade será ou não uma pessoa. A noção de condições de pessoalidade, mais à frente analisada, opõe-se desde logo a uma concepção absoluta de pessoa (segundo à 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<sup>9</sup>. Uma concepção absoluta de pessoa ou de acção traduz-se em distinções rigorosas entre pessoa e não-pessoa e entre acção e não-acção. Dennett não aceita concepções absolutas de pessoa e de acção e é precisamente o facto de não as aceitar que o conduz à sua posição gradualista e a afirmar que é impossível separar um conceito a que se poderia 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 uma entidade responsável pelas suas acções). Afirmará assim que a pessoalidade se constitui moralmente, i. e., que uma entidade no mundo se constitui enquanto pessoa através de acontecimentos que têm determinadas características intencionais.

O conceito de acção é, por seu lado, 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 agente – por contraste com aquilo que simplesmente acontece. Sendo conceitos mentais, os conceitos de pessoa e de acção são, como se disse, conceitos normativos. São-no em dois sentidos. Pessoa e acção são conceitos normativos no sentido geral da teoria do conteúdo, i. e., devido ao facto de suporem o apercebimento intencional e a racionalidade, e também devido ao facto de terem uma dimensão moral, prática. É devido

---

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, um acto que inicia um compromisso. R. de Sousa chama a esse acto *assenting* e considera que ele é qualquer coisa como 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, mas a existência de opiniões sim. Esta diferença é muito importante para a possibilidade de pessoas. Além do mais, é 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 conhecer qual seria o melhor curso de acção, opta por outro curso de acção. O auto-engano supõe logicamente uma divisão do agente, em agente que sabe algo e o oculta e o agente que não sabe, dentro de um mesmo indivíduo. A fraqueza da vontade e o auto-engano são os exemplos mais usuais na filosofia da mente de irracionalidade na acção e no pensamento devido à fragmentação da (suposta) unidade do agente. A diferença entre crença e opinião permite finalmente a Dennett isolar uma característica definitiva 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).

<sup>9</sup> Cf. LOCKE: 1999, *Ensaio sobre o Entendimento Humano*. J. Locke acentua aí o facto de 'pessoa' ser uma categoria moral (Livro II, XXVII).

a este duplo aspecto normativo que não é possível, por exemplo, identificar pessoas com corpos humanos num instante do tempo ou acções com movimentos de tais corpos.

Para além de ter consequências relativas ao estatuto dos conceitos de pessoa e acção a teoria do conteúdo tem ainda consequências relativas à questão da identidade pessoal. Esta não é estritamente identificável nem com a questão das condições de pessoalidade, sendo de certo modo menos abstracta, ou melhor, sendo o seu ponto de aplicação individual, nem com a noção de eu. O Eu é uma questão cognitiva prévia à questão da identidade pessoal e susceptível de uma teoria funcionalista subpessoal. De acordo com o Modelo dos Esboços Múltiplos (MEM), o modelo dennetiano da consciência, o Eu é uma auto-representação linguística e explícita tomada como centro de controlo virtual de uma (ou numa) sociedade de agentes em competição. Esta descrição é por sua vez uma interpretação psicológica funcionalista daquilo que ao nível do *hardware* é processamento paralelo distribuído. Isto significa que de um ponto de vista cognitivo a unificação e a centralidade dos sistemas cognitivos físicos conscientes são, de acordo com o MEM, *virtuais* (o que não significa que a unificação e a centralidade sejam *irreais*). O Eu é ainda, de acordo com o MEM, o centro de gravidade de uma narrativa autobiográfica automaticamente produzida<sup>10</sup> pelos seres humanos, um objecto abstracto surgido na/da interpretação linguística de si de um sistema cognitivo físico, interpretação pela qual aquilo que esta é e faz é unificado e continuado no tempo. Este Eu é uma extensão, ligada à linguagem, do *self* biológico, o princípio de organização em função do qual uma entidade distingue a si própria do resto<sup>11</sup>, sendo este *self* anterior à unificação dos agentes (vários) num só Agente racional e consciente que o Eu proporciona virtualmente.

O desenvolvimento da teoria dennetiana do Eu inicia-se em *The Mind's I – Fantasies and Reflections on Self and Soul*<sup>12</sup> e continua em textos como *The Self as a Center of Narrative Gravity*<sup>13</sup>, *The Origin of Selves*<sup>14</sup>, *The Reality of Selves*<sup>15</sup>, *Speaking for Ourselves*<sup>16</sup> e

---

<sup>10</sup> Por 'automaticamente produzida' entenda-se: a narrativa surge como se proviesse de um centro, no entanto esse centro é um produto da própria narrativa.

<sup>11</sup> O processo da constituição desta diferença tem que iniciar-se com (passar sempre por) um reconhecimento daquilo que é controlado como controlado. Nem todos os autocontroladores são Eus, embora todo o organismo seja um autocontrolador. Note-se por exemplo que qualquer bebé humano tem que *aprender* a controlar-se, descobrindo que controla e o que controla nomeadamente o seu corpo. De resto, a verificação do controlo de si não deixa nunca de estar subjacente aos Eus humanos e à identidade pessoal.

<sup>12</sup> DENNETT & HOFSTADTER 1981, *The Mind's I*.

<sup>13</sup> DENNETT 1992, inicialmente publicado sob o título *Why Everyone is a Novelist* (*Times Literary Supplement*, September 16-22, 1988).

<sup>14</sup> DENNETT 1989.

<sup>15</sup> O capítulo 13 de DENNETT 1991.

<sup>16</sup> DENNETT, D. & HUMPHREY, N., 1989.

em geral no âmbito do modelo de *Consciousness Explained*. *The Mind's I* é uma obra sobre as questões do eu<sup>17</sup> escrita – ou melhor, composta e arranjada, visto que se trata de uma antologia de textos de natureza díspar – em parceria com o cientista de computadores Douglas Hofstadter, na qual as questões do Eu são tratadas como reportando-se sobretudo às condições da auto-referência num sistema cognitivo,

---

<sup>17</sup> *The Mind's I* (HOFSTADTER & DENNETT 1981), cujo subtítulo é *Fantasies and reflections on self and soul* (1983), é uma antologia de textos literários e filosóficos reunidos por D. Dennett e Douglas Hofstadter e ao mesmo tempo uma viagem em busca do eu que culmina 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? (D. HOFSTADTER, in HOFSTADTER & DENNETT 1981: 408). Para os autores, o eu é o eu da mente. A resposta reconhece a atracção da posição dualista e a inultrapassável cisão entre a visão subjectiva e a visão objectiva de si. Dennett e Hofstadter pretendem 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 de si como também na física (cf. o texto 3, *Rediscovering the Mind*, de H. Morowitz), na lógica matemática, com Gödel (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 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*). Esta noção de símbolo-do-Eu é a noção central da teoria do eu proposta em *The Mind's I*. Em *The Mind's I* a teoria é defendida de acordo com o 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 substrato físico é paradigmaticamente defendida por Alan Turing, no seu célebre artigo de 1950, *Computing Machinery and Intelligence* (texto 4), comentado em *The Turing Test a Coffee 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 Hofstadter procuram explorar as consequências da aplicação da ideia funcionalista de mente como programa à noção de eu. Uma primeira consequência é (1) a ideia de *mind-tape*: a mente de uma pessoa, aquilo que a caracteriza como exactamente 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 se ser um eu ser-se um entidade orgânica, nascida de outra entidade orgânica. Outra consequência possível é (3) o espalhamento da implementação do programa (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 ser 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, da (5) Indistinção entre cópias e originais. Esta está em causa em situações como o teletransporte de pessoas, por reprodução exacta, partícula a partícula (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, são extensões ficcionais da teoria cognitiva proposta: Além disso, 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, entre natural

nomeadamente à constituição de uma (auto) representação de alto nível<sup>18</sup> da unidade global de um sistema cognitivo constituído por partes.

A identidade pessoal, aquilo pelo qual uma pessoa se (re)identifica como sendo uma e a mesma pessoa em cada instante e através do tempo, é um problema de outro nível. Em primeiro lugar, a reidentificação envolve provavelmente – mas não apenas – a narrativa autobiográfica acima referida, a qual tem como ‘centro de gravidade’ o Eu. O facto de existir uma vida mental – uma reunião das experiências a cada instante como constituindo as experiências de uma só pessoa, e alguém que apercebe a vida mental como sua, de forma continuada – sugere que alguma forma de continuidade (a continuidade psicológica ou a continuidade de algum sistema físico<sup>19</sup>) é a pedra de toque da identidade pessoal. A continuidade que aparentemente constitui o critério da sobrevivência de uma pessoa como a mesma pessoa esbarra no entanto com diversos problemas, levantados nomeadamente pela possível variação do suporte físico da continuidade<sup>20</sup>, pelas inúmeras interrupções (normais) da continuidade<sup>21</sup> de si conscientemente apercebida, pela possibilidade de fissão do suporte físico da continuidade<sup>22</sup>, pela indefinição das margens temporalmente mais distantes da suposta continuidade psicológica centrada no momento

---

e artificial, entre 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 reimplementação). 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. Hofstadter), *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 in DENNETT & HOFSTADTER 1981: 459). É nos comentários dos autores às selecções que aparece uma teoria do Eu. Propõe-se que o eu é a autoconsciência (*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 autoconsciente tem que haver um símbolo-do-Eu ‘mostrado’ a um *self*.

<sup>18</sup> I. e., não simplesmente biológica, existente em seres sem linguagem, como se supõe ser o caso daquilo a que se chama *self*.

<sup>19</sup> 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.

<sup>20</sup> Nesta situação (pelo menos até agora apenas imaginada) a mesmidade qualitativa de uma vida mental experienciada deixa de corresponder à mesmidade numérica de um suporte físico (por exemplo um mesmo cérebro). É para cobrir estas situações que D. Parfit (PARFIT 1984, *Reasons and Persons*, Oxford, Clarendon Press) fala de causas normais (como por exemplo a persistência de um mesmo cérebro) e de causas não normais da continuidade psicológica.

<sup>21</sup> Estas interrupções são uma das razões para pôr dúvida a existência de uma resposta definida à questão ‘será que as experiências foram minhas?’ (ligada a um determinado corpo) em cada instante do tempo.

<sup>22</sup> Cf. por exemplo PARFIT 1984, capítulo 12, §89.

presente (a qual torna as pessoas aparentemente menos apegadas aos estádios de si temporalmente mais distantes<sup>23</sup>), e até pelo facto de aparentemente não constituir uma característica fenomenológica específica das experiências o facto de elas serem *minhas*<sup>24</sup>. Não negando a importância das questões da continuidade, a teoria dennetiana do conteúdo conduz àquilo a que se chamará um critério normativo da identidade pessoal, para além dos dois mencionados critérios.

De acordo com D. Dennett certos agentes físicos capazes de unificação, de controlo e de sofisticação do controlo do comportamento próprio devido a uma representação de unificação, constituem-se como pessoas tanto quanto há lugar para acções guiadas por razões apercebidas (intencionalmente) como tal. A constituição da personalidade e a criação da identidade pessoal surgem neste lugar (mas não necessariamente a explicação da acção por razões que seriam as suas causas<sup>25</sup>).

No que respeita à teoria da acção o que importa aqui notar é que a posição de Dennett constitui uma alternativa quer a teorias reducionistas<sup>26</sup> quer a teorias da maximização racional como o utilitarismo e a deontologia<sup>27</sup>. 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 e com apoio em métodos heurísticos, conduzem Dennett a descrever – e até certo ponto a propor – aquilo a que chamará uma prudência naturalizada<sup>28</sup>. A ideia de

---

<sup>23</sup> D. Parfit, que utiliza a noção de conexão psicológica (PARFIT 1984:206) para referir a ligação entre os instantes/estádios unidos na continuidade psicológica, considera que pode existir conexão forte ou conexão fraca entre os estádios. Só no primeiro caso (a conexão forte) se poderia falar de *continuidade* psicológica propriamente dita.

<sup>24</sup> Cf. DAMASIO 1999, *The Feeling of What Happens* (New York, Harcourt Brace), para uma sugestão em contrário.

<sup>25</sup> Este problema é parte do problema do estatuto da explicação psicológica, um problema que cruza transversalmente, de forma inevitável, o presente problema das pessoas e das acções. Uma das razões por que a referência a Davidson não é aqui maior é o facto de este admitir, ao contrário de Dennett, não apenas que as razões podem ser causas de acções, mas também uma concepção absoluta de pessoa.

<sup>26</sup> Alude-se aqui às posições às quais Dennett chama *greedy reductionism* (reduccionismo ganancioso). Um exemplo é a tentativa skinneriana de uma ética behaviorista 'para além da liberdade e da dignidade', outro é a sociobiologia (cf. DENNETT 1995: 481-493). Aquilo que Dennett recusa não é, note-se, a ideia de uma análise evolucionista das normas éticas, tal como se encontra já em autores como T. Hobbes e F. Nietzsche.

<sup>27</sup> Cf. DENNETT 1995, capítulo 17 (*Can Ethics Be Naturalized*). Utilizar-se-á como é hábito o termo deontologia para referir posições em filosofia moral que tomam a filosofia moral kantiana como referência e que consideram que a resposta às questões 'Como se deve viver?', 'O que se deve fazer?' têm a forma de regras. Teorias morais deontológicas são caracterizadas por serem centradas-em-regras (*rule-centered*). Por vezes falar-se-á também em kantismo.

<sup>28</sup> Cf. DENNETT 1988, *Moral First Aid Manual* e DENNETT 1995, Capítulo 17, "Redesigning morality". Em *Darwin's Dangerous Idea*, uma obra cujo objectivo é a interpretação das implicações filosóficas do darwinismo, encontram-se aliás várias contribuições para a teoria da acção. Isto acontece por um lado porque a exploração filosófica do darwinismo que aí se empreende passa inevitavelmente por uma teoria da possibilidade e da

prudência naturalizada, i.e., a ideia de que certas disposições devem ser cultivadas pelos agentes de modo a subjazerem em geral e ao longo de várias situações à deliberação prática, criando-se assim uma ponderação devida a hábitos de pensamento e acção, opõe-se à ideia de maximização racional, uma ideia muito frequente em ética, é certo, mas cujo maior defeito é ignorar os problemas práticos da decisão em tempo real. A necessidade de considerar estes problemas é reflectida pela ideia de um 'manual de primeiros socorros morais' (*moral first-aid manual*). Para Dennett a teoria moral não deve ser ideal ou idealizante como o são inevitavelmente as éticas da maximização racional mas sim reflectir internamente os constrangimentos práticos da situação de deliberação<sup>29</sup>.

Poder-se-ia considerar, de uma forma aliás tradicional na filosofia, que o problema geral que unifica das questões da razão prática ou da filosofia moral é o problema da vontade livre. No entanto, a estratégia adoptada por Dennett relativamente à vontade livre é idêntica à estratégia adoptada relativamente à consciência. Dennett defende que a vontade livre não é um problema, mas vários problemas (problemas relativos ao controlo, ao Eu, à decisão, às razões na acção, ao 'poder ter feito de outra maneira', às atribuições de culpa), que se interpenetram aliás com os problemas da consciência e da racionalidade. Na corrente dedicada às questões da razão prática que atravessa toda a obra de Dennett o livro *Elbow Room – The Varieties of Free Will Worth Wanting* (ER)<sup>30</sup> ocupa uma posição central no que respeita à distinção entre esses problemas. Em ER Dennett propõe-se explicitar a possibilidade do agente livre, e portanto do agente moral, a partir de uma teoria cognitiva da agência (*agency*), do controlo, do eu

---

actualidade. Ora, também a teoria da acção supõe uma teoria da possibilidade. Por outro lado, uma abordagem evolucionista pode pretender dar conta da continuidade e da diferença das razões ou fins na acção em entidades muito diferentes, desde os organismos mais simples até aos humanos conscientes. De acordo com Dennett uma perspectiva evolucionista sobre a acção mostra um panorama que se estende desde os *free floating rationales* até às razões que não apenas existem no sistema como são razões para um sistema capaz de racionalidade explícita. Em *Kinds of Minds* (DENNETT 1996), uma outra obra em que o ponto de vista naturalista da TSI é explorado, a investigação dos fundamentos da ética explícita um tema que já vem, também ele, de longe e que pode perturbar a TSI: esse tema é a ligação da atribuição de mente à consideração a ter pelo 'sentimento' de outras entidades, o que traz nomeadamente o problema da dor e do sofrimento como parte do problema das outras mentes. Este tema não se restringe obviamente às pessoas, e estende-se no mínimo a alguns animais.

<sup>29</sup> De resto, Dennett utiliza a expressão 'manual de primeiros socorros morais' para sublinhar que a pressão-para-a-decisão em tempo real, usualmente considerada como teoricamente desprezável não é teoricamente desprezável nem insignificante. Pelo contrário, o que não se justifica é uma oposição entre o que seria realmente ética, cujo objecto seriam os princípios para um agente ideal em todas as circunstâncias e um 'manual de primeiros socorros morais', que prescreveria o que fazer de modo a agir de forma pronta em situações concretas de pressão temporal (cf. DENNETT 1988:123). Assim "The Moral First Aid Manual should (...) be considered not merely as a grubby compromise with practicality, but itself just as pure an ideal vision as any other in ethics: if you like, it is the book the ideally rational agent would write as his own vade mecum, written in the light of his perfect self-knowledge about his many limitations (DENNETT 1988: 138).

<sup>30</sup> DENNETT 1983.

e da deliberação. A análise cognitiva da agência estabelece a possibilidade de deliberadores deterministas. A questão da vontade livre é assim afastada do âmbito de uma discussão acerca da relação entre a vontade e o determinismo físico para o âmbito da teoria cognitiva e da interpretação intencional. Como se passa com a consciência, existem ilusões na própria delimitação do problema da vontade livre que importa começar por afastar e uma dessas ilusões é precisamente o suposto contraste entre a vontade livre e o determinismo. Se, como Dennett sugere em *ER*, aquilo que queremos quando queremos vontade livre é ter controlo e autocontrolo, não tem que existir qualquer incompatibilidade ou contraste entre o determinismo físico e a liberdade. O que está em causa no chamado problema da vontade livre são problemas relativos ao controlo, à decisão, à racionalidade na acção e não a prova da existência de um ingrediente a mais na realidade que seria a vontade livre.

Em *ER* Dennett acaba por caracterizar a chamada vontade livre como autocontrolo meta-reflexivo<sup>31</sup> – o controlo do comportamento próprio e da identidade própria por meios racionais conseguido por alguns sistemas cognitivos físicos resultantes de evolução e existentes num mundo determinista – e avaliação forte<sup>32</sup>. Deliberadores deterministas são capazes de liberdade e de responsabilidade tanto quanto são capazes de autocontrolo meta-reflexivo e de avaliação forte. O naturalismo gradualista sustentado pela TSI e pelo MEM é assim uma forma de compatibilismo (de acordo com o compatibilismo a liberdade da vontade – seja ela o que for, neste caso sugere-se que é autocontrolo meta-reflexivo e avaliação forte – é compatível com o determinismo físico).

Um terceiro ponto de referência para o percurso que aqui se segue é o artigo *The Moral First-Aid Manual*<sup>33</sup>. Dennett critica nesse artigo, a partir da análise da estrutura cognitiva dos agentes humanos, a falta de pertinência psicológica e cognitiva das éticas da maximização racional, propondo como alternativa uma ética centrada na ideia de prudência naturalizada. Curiosamente, esta proposta aproxima Dennett das temáticas desenvolvidas nas éticas da virtude.

A ideia subjacente ao presente artigo é que é possível destacar a partir das teorias dennettianas do conteúdo e da consciência propostas relevantes para a filosofia moral. Daqui em diante serão considerados três pontos focais da teoria da razão prática: a

---

<sup>31</sup> I.e., o controlo de si próprio através de razões, a que se poderia chamar também controlo racional.

<sup>32</sup> A ideia de avaliação forte é reportada às concepções de C. Taylor sobre a responsabilidade pela identidade própria (cf. TAYLOR 1976, *Responsibility for self*). Um avaliador forte é um agente racional que se coloca as questões 'O que é que eu quero ser?'; 'Quero ser/ continuar a ser aquilo que agora sou?'. Ele avalia assim não apenas aquilo que fez/faz mas aquilo que é. A resposta a tais questões e portanto a avaliação forte supõe a possibilidade de controlo – no sentido de avaliação e determinação – daquilo que se é. Segundo C. Taylor o vocabulário da avaliação é uma parte importante do vocabulário moral pós-nietzscheano e é esquecida pelas acima chamadas éticas da maximização racional.

<sup>33</sup> DENNETT 1988.

agência e a liberdade<sup>34</sup>, a unidade e a identidade pessoal<sup>35</sup> e a deliberação<sup>36</sup>. O problema maior, que não será aqui totalmente tratado na medida em que se pretende antes de mais descrever aportações, é avaliar a consistência global da teoria da razão prática desenvolvida a partir das teorias do conteúdo e da consciência.

## 2. A fragmentação do problema da vontade livre e as formas que o 'problema da vontade livre' não tem: vontade numérica, indeterminismo físico, capacidades mentais não mecânicas<sup>37</sup>.

"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* II, III, 2

Antes de seguir o trajecto de exposição proposto – que parte de uma teoria do controlo e da agência e vai até às noções de pessoa, avaliação forte e prudência – explicitar-se-á a forma que o problema da vontade livre, de acordo com Dennett, não tem, justificando assim a exclusão de determinadas abordagens do chamado problema da vontade livre.

Dennett pensa que não é uma boa estratégia perguntar directamente o que é a vontade livre já que esta não corresponde a algo de único e bem definido. Deve-se por isso perguntar não o que é a vontade livre mas 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 a *certain looseness* de que fala Hume, uma indefinição ou indeterminação ou plasticidade na maneira como as coisas serão. As coisas futuras parecem poder vir a ser de várias maneiras, dependendo das nossas acções. Trata-se então de elucidar teoricamente essa

---

<sup>34</sup> Sob este título incluir-se-á desde o controlo em sistemas físicos não racionais e não conscientes até ao tipo de controlo característico dos seres racionais e conscientes a que se chamará liberdade de acção.

<sup>35</sup> Sob este título incluir-se-á a teoria do *self*, do eu, da personalidade, da identidade pessoal.

<sup>36</sup> Sob este título serão consideradas as condições cognitivas da deliberação e da escolha, a possibilidade, a sorte moral e o 'poder ter feito de outro modo' (*could have done otherwise*), que é usualmente tomado como sinal seguro da acção responsável de um agente moral.

<sup>37</sup> Acerca do argumento geral em defesa da existência de 'elbow room' e portanto daquilo que se deve querer quando se quer vontade livre que Dennett apresenta em DENNETT 1983 poder-se-ia sempre dizer que um tal argumento visa um alvo erróneo. Pode-se negar a vontade livre sem se ser reducionista e determinista, partindo de premissas de ordem completamente diferente das premissas de que Dennett parte nos argumentos contra o reducionismo e o determinismo que a seguir se discute. Como afirma Arthur Danto (DANTO 1986): "a negação tradicional de margem de manobra (*elbow room*) é a ausência lógica de espaço entre mim próprio (*myself*) e eu (*me*)".

*certain looseness*. De acordo com Dennett, não há razão para reportar essa indefinição ou indeterminação à questão do determinismo na física – nomeadamente ao indeterminismo quântico – ao contrário do que vários autores sugerem. A questão física do determinismo não tem qualquer relação directa com o chamado problema da vontade livre. Não têm também qualquer relação com ele as capacidades mentais humanas que estariam para além daquilo que pode ser conseguido por computações mecânicas<sup>38</sup>, ao contrário do que sugerem os autores que abordam a vontade livre por meio de argumentos que envolvem o Teorema de Gödel. Para Dennett as questões da liberdade e da responsabilidade de humanos são questões de macronível, questões que dizem respeito a um particular entendimento ou interpretação de eventos físicos como acções<sup>39</sup> e não questões relativas ao *micronível* 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. Colocada deste modo, a posição de Dennett parece ser uma rejeição grosseira de sofisticadas discussões acerca de vontade livre. Ela é no entanto a única posição coerente possível a alguém que crê que tudo o que é mental é 'mentalista', i. e., concomitante de uma descrição interpretativa global de sistemas físicos.

Em *Mechanism and Responsibility*<sup>40</sup> Dennett pretende mostrar que o determinismo<sup>41</sup> suposto na explicação científica nem pode ser seriamente contestado nem é incompatível com a responsabilidade. O problema da responsabilidade num mundo determinista diz respeito à relação das descrições intencionais<sup>42</sup> com as explicações físicas. Ora, estas descrições e explicações apenas aparentemente são conflituais.

Num sentido elementar, responsabilidade significa poder dar razões das acções próprias, o que envolve a apresentação pelo agente de descrições intencionais de

---

<sup>38</sup> Cf. DENNETT 1981d. Para Dennett esses argumentos que supõe, que Jones o humano, ou pelo menos o humano-matemático, transcende os limites do mecanicismo porque pode pensar aquilo que não é mecanicamente computável, falham porque 1. negam implicitamente 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 de 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, *The Emperor's New Mind* e PENROSE 1994, Oxford, OUP, *Shadows of the Mind*. O argumento foi apresentado por J. R. Lucas em 1961 no artigo *Minds, Machines and Gödel*.

<sup>39</sup> É uma posição semelhante por exemplo à de Peter Strawson. Segundo Strawson chamamos livres às coisas que estamos preparados para considerar como responsáveis mas não existe nenhuma caracterização não normativa, i. e., que não envolva avaliação, do que essas coisas livres e responsáveis possam ser.

<sup>40</sup> DENNETT 1981a, *Mechanism and Responsibility*.

<sup>41</sup> 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, se segue E. Se de acordo com o determinismo existe uma explicação causal para tudo aquilo que acontece, o comportamento humano supostamente livre, as acções guiadas por razões, está também incluído.

<sup>42</sup> De acordo com a TSI não existem explicações intencionais.

eventos, e portanto o apercebimento de movimentos e intenções próprios como algo diferente de meros acontecimentos. De acordo com a TSI a Estratégia Intencional, que envolve a atribuição holista de racionalidade ao sistema cognitivo, é uma condição da atribuição de responsabilidade a agentes. Isto aparentemente torna a 'Estratégia Moral' (*moral stance*) perigosamente arbitrária, resultante de uma decisão em última análise instrumental do intérprete. Dennett nunca abdica da ideia segundo a qual a Estratégia Intencional é uma pré-condição do estatuto moral de entidades no mundo. No entanto, em lugar de essa dependência mostrar a arbitrariedade da estratégia moral, o que ela mostra é a pertinência da transposição para o campo moral da relação das descrições intencionais com as explicações físicas de sistemas explorada na teoria do conteúdo. De acordo com a TSI, a Estratégia Intencional e a Estratégia Física não são alternativas mutuamente exclusivas de abordagem da realidade. Uma explicação física não exclui uma descrição intencional e evidentemente uma descrição intencional (de uma acção) também não exclui a explicação física. É certo que Dennett sempre defendeu que não existiria nunca, por princípio, uma distinção absoluta entre as coisas pelas quais somos responsáveis e aquelas pelas quais não o somos: é impossível defender no seio da TSI uma concepção absoluta de acção. O que interessa mostrar, no entanto, é que o determinismo da explicação física apenas se traduziria numa irrelevância da descrição intencional se se partisse de inaceitáveis premissas incompatibilistas. De acordo com o incompatibilismo, a liberdade da vontade é incompatível com o determinismo. Existem, evidentemente, pelo menos dois tipos de incompatibilismo: alguns incompatibilistas defendem que a realidade da vontade livre prova que o determinismo não é verdadeiro, outros defendem que sendo o mundo determinista não existe liberdade da vontade. Os incompatibilistas do primeiro tipo serão daqui em diante referidos como libertários (*libertarians*). Dentro da TSI, que considera a explicação física como 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*<sup>43</sup> Dennett defende que é possível conceder aos libertários – que são incompatibilistas – o que eles afirmam querer. O que os libertários afirmam querer é indeterminação anterior à decisão ou à formação da intenção. Sem essa indeterminação nunca poderia existir liberdade da vontade. É usualmente objectado aos libertários que se se considera que as acções livres dependem da aleatoriedade, uma vez que 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 os libertários se inclinam para fora da natureza na sua tentativa de conceber a vontade livre. De acordo nomeadamente com a doutrina da causação do agente (*agent-causation*), agir seria provocar determinados eventos sem que o agente fosse causado de modo a causar esses eventos. A acção livre seria assim a iniciação não

---

<sup>43</sup> DENNETT 1981e, *On Giving Libertarians What They Say They Want*.

causada de uma cadeia causal. Chamar-se-á a esta ideia a irrupção do agente, ou do actor. Dada a irrupção do agente, a sua insubmissão à causalidade, o agente seria capaz de agir exclusivamente por razões. Ora, Dennett considera que a ideia de um agente puramente movido a razões é 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"<sup>44</sup>. A indeterminação desejada pelos libertários não tem que supor uma irrupção do agente nem um agente puramente movido a razões. A indeterminação que os libertários consideram ser necessária para a existência de liberdade da vontade não é o indeterminismo microfísico, que fundaria a indeterminação macroscópica da acção, nem a indeterminação devida a uma vontade livre numérica. A indeterminação que segundo Dennett importa para a acção é apenas aleatoriedade, ligada à inexistência de padrões. Ora, aleatoriedade nesse sentido é possível num mundo determinista. Para reportar a indeterminação que importa na vontade livre à aleatoriedade resta retirar a esta a conotação de arbitrariedade.

Recapitulando, aquilo que interessa quando se trata da chamada vontade livre não são movimentos de partículas descritos pela teoria física e nem sequer movimentos esqueléticos de humanos, mas sim acontecimentos considerados como acções através de descrições mentalistas e normativas. Se o determinismo ou indeterminismo físicos não têm relação com a acção livre ou com o conceito de liberdade é porque estes são conceitos normativos. As acções podem ser realizadas de diversas formas, e quando quisermos saber perante determinados eventos de que acção se trata o que queremos não é saber exactamente por meio de que movimentos físicos ela foi levada a cabo. Só é possível fazer sentido de uma acção como constituindo uma particular acção através de uma descrição intencional. Nestas condições, o comportamento humano pode ser determinado ou indeterminado do ponto de vista físico e previsível ou imprevisível intencionalmente, de forma independente. Uma acção pode não ser arbitrária (i. e., pode ser intencionalmente previsível) mesmo que consista numa descrição de eventos resultantes de indeterminismo físico. Por outro lado, um agente pode ser intencionalmente imprevisível mesmo que o determinismo físico seja verdadeiro. Não são portanto o determinismo ou o indeterminismo físicos que estão em causa quando se trata da inteligibilidade intencional da acção. No artigo *On Giving Libertarians What They Say they Want* Dennett dá como exemplo deste último ponto o comportamento de uma 'caixa de respostas'<sup>45</sup>. Trata-se de um dispositivo electrónico imaginado que regista respostas de pessoas a questões simples. Tem dois botões (sim e não), dois pedais (sim e não) e um ecrã dividido ao meio, onde aparece de um lado a instrução 'use os botões' e do outro lado a instrução 'use os pedais'. Na situação experimental apenas metade do ecrã é iluminada de cada vez. O sujeito é exposto a questões simples, às quais deve dar

---

<sup>44</sup> DENNETT 1983: 49.

<sup>45</sup> DENNETT 1981e: 290.

como resposta sim ou não, seguindo as instruções do ecrã. A 'escolha' da metade do ecrã que está em causa de cada vez resulta de um processo aleatório no dispositivo.

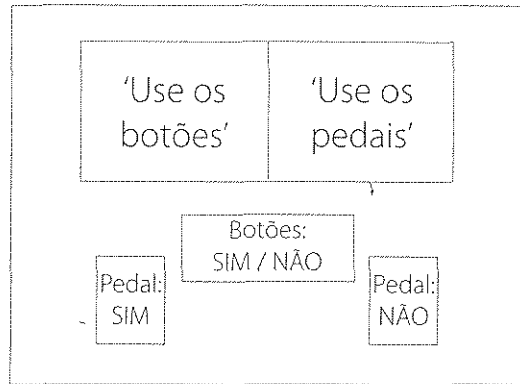


Figura 1: Ecrã, botões e pedais.

É prometido ao sujeito um bônus pelas respostas correctas. O problema é: em que situação estarão o físico e o teórico intencional para a previsão do comportamento do sujeito, supondo que este é um sistema fisicamente determinista mas que está a ser afectado pela aleatoriedade do processo de escolha? A resposta é que o teórico intencional (cada um de nós) prevê melhor, não porque a sua teoria seja melhor – ele nem sequer pode prever melhor do que o físico se o sujeito vai fazer determinado movimento – mas porque aquilo que interessa não são movimentos físicos do sujeito e sim as acções deste.

Este é um exemplo de previsão intencional de acontecimentos baseados em processos físicos cujas características microfísicas não se conhece e que são aleatórios num determinado sentido. Dennett sugere que processos semelhantes aos da caixa de respostas subjazem de qualquer modo a qualquer decisão humana. Assim, num certo sentido existem processos indeterminados (ou mesmo indeterministas) que são intencionalmente previsíveis. Uma acção é praticável de diversas maneiras e são factores aleatórios que determinam a maneira particular de realizar a acção intencionada, i. e., existem 'pontos de escolha' possivelmente ubíquos a nível inconsciente e que não aparecem fenomenologicamente. Ao escolher uma das mil maçãs que estão na minha frente quando eu penso 'quero uma maçã', não fico dias a ponderar e calcular qual será a melhor maçã particular que satisfaz o meu querer 'uma maçã'. No entanto, de modo algum o facto de a implementação de qualquer acção supor vários destes pontos de escolha indeterminados torna no entanto o agente intencionalmente imprevisível.

A proposta de Dennett, desenvolvida sobretudo em *Elbow Room* e *Darwin's Dangerous Idea* é que é precisamente a exploração inteligente da aleatoriedade num

mundo determinista que é responsável pelo espaço de manobra (*elbow room*) para a acção. Este espaço de manobra seria aquilo queremos quando queremos vontade livre, não se identificando assim nem com o indeterminismo físico nem com uma vontade livre não física, consistindo antes na exploração inteligente da aleatoriedade que subjaz à deliberação e à decisão num mundo determinista. O modelo da deliberação e decisão que Dennett propõe – e recorde-se que Dennett considera que as decisões são os fenómenos focais esclarecedores da vida mental do agente livre como se viu através da distinção entre crença e opinião a que se aludiu na nota 9 – é então esquematicamente o seguinte:

1. Produção aleatória. Para haver comportamento de decisão deve haver geração de uma variedade de alternativas. Se é esta a indeterminação que o libertário deseja, o teórico naturalista pode conceder-lha sem ter que admitir qualquer causação do agente, indeterminismo físico ou vontade livre numérica.

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) não podem considerar exaustivamente as alternativas geradas. É necessário o uso 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 nomeadamente pelo cálculo de probabilidades, pela teoria dos jogos e pela lógica temporal. Trata-se de um problema de ciência cognitiva, aliás mais tratado até hoje por economistas e estatísticos do que por psicólogos e filósofos<sup>46</sup>. Evidentemente, uma vez que se trata aqui de humanos e do seu sentimento de agir não do comportamento do mercado ou de jogos de azar, fica o problema fenomenológico de saber o que constitui o nosso sentimento de sermos agentes livres e o que contribui para ele. Esta será, nos termos de Dennett, a questão de saber como é que um processo subpessoal nos aparece ao nível pessoal e consciente quando deliberamos e pensamos que deliberamos e decidimos. Está fora de questão afirmar que as tomadas de decisão aparecem ao agente sob a forma do processo subpessoal no qual de certo modo consistem (cf. os pontos acima enumerados). Retrospectivamente o agente considerará que existiu uma razão para a sua decisão. No entanto, tendo sido gerada por um processo como o descrito, essa razão não será a Razão *sub specie aeternitatis* ou, numa perspectiva de maximização

---

<sup>46</sup> YATES & ESTIN 1998, *Decision Making*.

racional, a razão única. Assim, Dennett surge que o sentido de liberdade do agente a nível pessoal (a referida *certain looseness*, mais o sentimento de a ter determinado) depende menos da decisão tomada ela própria do que de decisões subsidiárias, tais como a decisão de terminar a deliberação, de ignorar outras considerações e de passar ao acto.

### 3. A liberdade num mundo determinista: aleatoriedade, controlo, espaço de manobra (*elbow room*) e descrição intencional.

#### 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.

Mesmo que venha a ser possível considerar certos eventos como acções por meio da descrição intencional e como escolhas livres devido à aleatoriedade a constatação que se impõe antes de mais a partir do ponto de vista naturalista adoptado 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 são portanto o exacto oposto de livres. Para nomear esta última condição Dennett fala da *sphexishness*<sup>47</sup> do comportamento, em referência à fixidez revelada pelo comportamento animal, mesmo pelo comportamento animal aparentemente inteligente. O exemplo dessa escolhido em *Elbow Room* é um comportamento de uma vespa. A revelação da limitação da inteligência da vespa escavadora *Sphexichneumonius* dá-se quando se interfere com os passos rotineiros do seu comportamento<sup>48</sup>. Quando está na altura de pôr os ovos, a vespa 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 acontecerá quantas vezes for necessário, sem que a vespa alguma vez 'pense' e resolva trazer o animal imediatamente para o interior já verificado do buraco. Ora, 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 aliás esta uma objecção constantemente levantada em relação à simulação de inteligência em máquinas. A fixidez e repetição

---

<sup>47</sup> O termo foi proposto por Douglas Hofstadter, cf. DENNETT 1983: 11.

<sup>48</sup> DENNETT 1983:10-11.

torna-se mais óbvia nas mentes humanas em situações patológicas, como por exemplo situações de mania e de obsessão.

O filósofo moral tem que lidar com o facto de haver uma grande rigidez naquilo que somos pelo facto de sermos seres físicos e organismos. Ser um sistema físico e um organismo determinado envolve uma limitação *a priori* do leque de possibilidades de acção (desde logo, para começar, do leque de possibilidades de movimento e de 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 dez toneladas). 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 ainda respeito à existência de mais ou menos alternativas possíveis de acção (margens de erro, opções de comportamento). O espaço de manobra é por vezes identificado como subjazendo à *certain looseness* humeana<sup>49</sup> atrás nomeada. No entanto, se um humano tem muito mais *elbow room* para o seu comportamento do que uma vespa *Sphex*, isso não significa 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<sup>50</sup>. Nem um humano é totalmente livre no sentido de não ser para um mundo determinado e 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 extradeterminado 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<sup>51</sup>.

No entanto, "nunca ninguém afirmou que porque o determinismo é verdadeiro, o termostato não controla a temperatura"<sup>52</sup>. Esta observação de Robert Nozick<sup>53</sup> introduz

---

<sup>49</sup>"The necessity of any action, whether of matter or of the mind, is not properly a quality in the agent, but in any thinking or intelligent being, who may consider the action, and consists in the determination of his thought to infer its existence from some preceding objects: As liberty or chance is nothing but the want of that determination, a certain looseness, which we feel in passing or not passing from the idea of one to that of the other"; HUME 1985: 455-456.

<sup>50</sup>"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 1981: 244-245).

<sup>51</sup>"in this sense any behavior controlled by a finite mechanism must be tropistic" DENNETT 1981: 245.

<sup>52</sup> Citado em DENNETT 1983: 50, *Control and self-control*.

<sup>53</sup> R. NOZICK 1981: 315.

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 dispor de controlo sobre a nossa situação nem sobre nós próprios nem estar sob o controlo de factores externos (seja o factor externo um outro agente seja uma situação na qual não existem alternativas de acção). É em parte o temor de se ser controlado, manipulado, um juguete de forças exteriores, um 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 contemporâneos. Saber o que é o controlo e explorar as suas condições é por isso uma tarefa básica da teoria da acção. Na medida em que apenas é possível controlar os estados ou as actividades que estão dentro do leque de liberdade de alguma entidade, pode ser que não saibamos o que estamos a querer (i. e., pode ser que estejamos a querer aquilo que não podemos querer) quando queremos liberdade *tout court*, liberdade radical, iniciação *ex nihilo* de não importa que acção, como se as nossas acções pudessem ser quaisquer umas.

Esta abordagem da teoria da acção através das ideias de controlo e de autocontrolo num mundo determinista coloca-se evidentemente contra a ideia libertária de acção livre concebida como irrupção do agente livre no mundo determinista. Dennett subscreve a objecção de B. Williams à vocação extramundana da filosofia moral: "nenhuma característica humana relevante para graus de estima moral pode deixar de ser uma característica empírica, sujeita a questões empíricas, história psicológica e variação humana"<sup>54</sup>. Assim, a filosofia moral sustentar-se-á sobre uma teoria do controlo que se inicia com as determinações empíricas, mundanas, dos humanos enquanto sistemas cognitivos físicos. Para Dennett, estas determinações empíricas são limitações objectivas de dois tipos. Constitui limitação objectiva não apenas a já referida determinação física e de *design* das entidades que poderão ser agentes, como também a sua limitação cognitiva. É nesse sentido que, em oposição ao kantismo e em geral às éticas da maximização racional, o teórico naturalista da acção nota que não é possível uma 'Vontade Pura', se por tal se entende uma vontade infinitamente sensível à voz da razão. Uma vontade pura decidiria que acção empreender após a consideração, na deliberação, de todos, exaustivamente todos, os factores relevantes. Ora, não é possível a entidades cognitivamente limitadas considerar exaustivamente todos os factores relevantes numa deliberação antes de passarem à acção<sup>55</sup>. Na terminologia da TSI,

---

<sup>54</sup> B. WILLIAMS 1973: 228.

<sup>55</sup> Considere-se o seguinte exemplo que S. Stich (STICH 1993: 151-152) vai buscar ao trabalho de C. Cherniak sobre as implicações epistemológicas da avaliação, no âmbito da ciência de computadores, da exequibilidade (*feasibility*) de diversas classes de algoritmos: "Among the surprising results in this area is the demonstration

reafirma-se, agora num novo sentido, que não existem máquinas semânticas. A máquina semântica seria aqui entendida como o Agente Racional, que agiria por significações e razões por oposição a causas, e que seria uma vontade pura, cognitivamente ilimitada. Em primeiro lugar as máquinas semânticas são fisicamente impossíveis porque todos sistemas físicos estão submetidos à causalidade e 'Razões' são descrições intencionais da causação. 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. A existência de uma capacidade de cálculo perfeita, implícita nessa imagem, não é plausível em sistemas finitos deliberando em tempo real. Noutras palavras, uma das consequências da TSI é que o agente racional de Descartes, de Espinoza ou de Kant não pode existir neste mundo, i. e., nenhum sistema físico com *design* pode ser um agente racional puro. A teoria da acção deve partir da apreciação dos constrangimentos que o tempo real coloca à decisão e à acção e de uma apreciação da maneira que sistemas finitos têm de lidar com esses constrangimentos. Curiosamente, esta abordagem aproxima Dennett, como já se fez notar, do reconhecimento, frequente nas éticas da virtude, antigas e medievais, ao contrário do que acontece nas éticas modernas, da importância da prudência e da sageza dadas as limitações da acção racional humana. Simplesmente, a prudência é agora 'naturalizada', i. e., considerada como uma característica do *design* de sistemas deliberadores resultantes de evolução por selecção natural.

#### 4. Da teoria do controlo ao controlo meta-reflexivo e à avaliação forte.

##### 4.1. Deliberação, decisão, oportunidade. Previsibilidade e imprevisibilidade.

A deliberação e a decisão: vantagens da insensibilidade e da arbitrariedade. Deliberação e possibilidade epistémica.

Tendo reportado as ideias de acção e de responsabilidade ao apercebimento das razões para agir e ao envolvimento destas no controlo do comportamento, importa saber em que consiste o controlo e qual é a relação deste com a causação e o determinismo. A questão que Dennett coloca perante a descrição de sistemas que se autocontrolam, é, tal como acontecia com o modelo da consciência, o que pensaríamos se vissemos desenvolver-se um sistema capaz de alterar e comandar o seu comportamento em virtude

---

that many quite familiar algorithms – including some that would likely be of considerable pragmatic utility to any cognitive system that could compute them easily – require immensely more computational power than could be packed into a human skull. 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 unfeasible for a human brain, it is unfeasible even for what we might suppose to be the ideal, physically buildable computer.<sup>7</sup>

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.

A ideia básica de controlo com que se inicia a teoria do agente racional, aquela que na teoria dos autómatos dá lugar a conceitos técnicos, é a seguinte:

**Definição:** A controla B se 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 queira. 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, assim, idêntico a afirmar que A pode fazer B fazer o que quer que seja. Só se pode controlar os estados de uma entidade que pertencem ao leque dos seus 'graus de liberdade'. Considerando sistemas físicos variados, a ideia de graus de liberdade de entidades mostra ser pertinente muito antes da existência de agentes morais. Ela pode ser usada em relação a artefactos, nomeadamente, e não supõe agência ou liberdade das entidades em consideração. Um dos pré-requisitos do agente racional é portanto a existência de graus de liberdade e de controlo. Um subproblema concreto aparece também: como é que se sabe que se controla? O problema coloca-se não apenas porque a existência (ou não existência) de controlo tem que ser descoberta, não sendo à partida conhecida, como também porque pode haver ilusão de controlo. No caso do agente individual, só por experiência e variação ele descobrirá se controla e o que é que controla. Relativamente à definição geral de controlo, e tomando como exemplo precisamente artefactos controlados, é também conveniente relevar o facto de que, quando se controla alguma coisa (quando por exemplo um piloto controla um avião) não se controla todas as causas que a influenciam (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<sup>56</sup>. O que interessa notar aqui é que é possível ao controlador tomar metadecisões de modo a aumentar o controlo (por exemplo, o piloto não conduz o avião em direcção à tempestade, pois aí o seu espaço de controlo seria muito menor). Estas decisões são estratégias que podem, elas sim, ser chamadas estratégias de maximização, mas de maximização do *elbow room*, e não da utilidade esperada.

A definição de controlo acima apresentada configura o controlo como uma relação entre entidades individuais. No entanto, o controlo pode remontar a um ponto anterior

---

<sup>56</sup> Longe de estarmos submetidos ao controlo pelo nosso passado biológico, foi precisamente a evolução biológica que nos preparou para sermos dispositivos autocontrolados.

à existência de agentes (humanos, animais ou máquinas). Os psicólogos behavioristas, e nomeadamente B. F. Skinner, falam por exemplo de 'controlo do comportamento de um organismo pelo ambiente'. Ora, o ambiente não é um agente, nem tem desejos e na definição de controlador apresentada ter desejos era criterial. Quando um behaviorista fala de comportamentos que estão 'sob o controlo de estímulos', fala de um controlo 'sem agente' (*agentless*) e refere o facto de mudanças em A (características do ambiente que funcionam como 'estímulos') serem fiavelmente reflectidas por mudanças em B (organismos, comportamentos). A situação difere da simples causação física. Para Dennett, salvaguardados os desvios ideológicos do uso behaviorista da noção de controlo, a utilidade da noção behaviorista de controlo do comportamento do organismo pelo ambiente é chamar a atenção para a transição conceptual (e real) entre a 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 e nesse sentido controla-as, fazendo com que os seus comportamentos sejam apropriados. Como diz Dennett "evidentemente que o mundo como um todo não é contra nós nem a nosso favor"<sup>57</sup> (ou de qualquer outro organismo), logo as noções de controlo apropriado e controlo inapropriado aqui em causa têm que ser tomadas *cum grano salis*. A suposta tendência benigna do ambiente controlador de que fala B. F. Skinner<sup>58</sup> – sabendo-se que o ambiente é indiferente, não é um agente, não tem desejos – sublinha o facto de criaturas com o *design* certo para fazerem as discriminações apropriadas terem evoluído.

O uso behaviorista da noção de controlo propicia uma fusão do controlo com a causação, produzindo uma ideia ameaçadora e negativa do controlo. Esta fusão, nota Dennett<sup>59</sup>, estende-se nomeadamente ao entendimento de situações experimentais relativas ao comportamento de organismos, e das variáveis dependentes (que o experimentador *controla*) e variáveis independentes (que o experimentador *não controla*) nestas situações. O isolamento de variáveis dependentes permite que experiências semelhantes sejam interpretadas como revelando a causa do comportamento (a variável independente) aparentemente eliminando assim qualquer espaço de manobra. Este entendimento das causas de um comportamento, sublinhando 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, praticamente incalculável. É fundamentalmente por isso que o processo é incontrolável, por oposição ao controlo que o experimentador exerce sobre as variáveis dependentes.

---

<sup>57</sup> DENNETT 1983:58.

<sup>58</sup> Cf. comentários em DENNETT 1983:57-61.

<sup>59</sup> DENNETT 1983:59.

As condições da previsibilidade ou imprevisibilidade de processos causais na sua relação com o controlo constituem o problema seguinte na 'construção' de um agente racional. As estratégias de alto nível de maximização do controlo envolvem não apenas o autoconhecimento mas também a posse de conhecimento acerca de outros agentes no ambiente. A importância da teoria dos jogos<sup>60</sup> para a teoria da decisão racional consistiu em ter mostrado a diferença que faz a presença de outros agentes no cálculo das opções racionais de acção. Se existem outros agentes racionais agindo e recolhendo informação no ambiente, a situação a controlar, i. e., aquilo que é previsível ou imprevisível, muda totalmente. Agentes podem utilizar informação para prever e controlar outros agentes. Uma teoria da racionalidade na acção não pode assim deixar de considerar os princípios que regem a racionalidade de situações nas quais agentes competem pela informação, com desejo e necessidade de ocultar a informação que possuem. Nestas situações, o princípio geral será que quanto mais informação se possui mais 'livre' se é. Embora seja indesejável ser controlado pelos outros agentes (no limite, ser controlado pelo Neurologista Maligno, que induz em nós crenças desejos e pensamentos... i. e., toda a vida mental e racional) não é possível não ser controlado por nada. A 'liberdade radical' não poderia consistir numa ignorância total da informação que é 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 e competição de agentes racionais, o espaço de manobra, gerado através do metacontrolo que permite maiores margens de liberdade e de erro é conseguido, segundo Dennett, através dos usos da desordem. É através do uso da desordem que agentes racionais iludem possíveis controladores (os outros agentes que compreenderiam bem demais 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 seguir padrões visíveis nas suas actividades, elas terão que ser aleatórias. De novo, esta aleatoriedade não é indeterminismo físico mas sim aleatoriedade no sentido da teoria da informação, segundo o qual uma série aleatória é informacionalmente incompressível. A ausência de padrão, ou a imperscrutabilidade do padrão tem um papel importante nas técnicas de controlo e autocontrolo e um valor estratégico em situações de jogo (contra adversários, ou contra a natureza) e as estratégias de maximização do espaço de manobra através da exploração de processos aleatórios dependem em grande parte do autoconhecimento do sistema que a pratica.

Por outro lado, Dennett acentua que sistemas que precisam de decidir para agir e que são recolhedores imperfeitos de informação, têm necessidade de incorporar despoletadores arbitrários de decisões (tais como a técnica da moeda ao ar). A imperscrutabilidade do padrão estará portanto em parte ligada a decisões arbitrárias que é racional substituir à completa e racional ponderação quando existe interesse numa rápida decisão. Dennett desde sempre encontrou interesse na descrição da origem da

---

<sup>60</sup> Cf. YATES & ESTIN 1998 e F. SCHICK 1997:88-105.

consciência pelo psicólogo Julian Jaynes<sup>61</sup>, e em *Elbow Room* recorda as análises do valor cognitivo de técnicas exo-psíquicas de decisão, como o lançamento de uma moeda ao ar ou a leitura de entranhas de pássaros, feitas por J. Jaynes<sup>62</sup>. A importância dessas técnicas é mostrar que alguma coisa, não importa o quê, é necessária para interromper a geração de considerações e despoletar uma decisão quanto ao que deve ser feito ou pensado.

De novo considerando as decisões, os pontos focais da vida mental do agente livre, é importante notar que este tipo de considerações acerca dos mecanismos cognitivos subpessoais da deliberação e da decisão se encontram se encontram 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 de liberdade (a imagem de Kant, Espinosa, Descartes ou Locke<sup>63</sup>) apenas uma decisão perfeitamente racional é considerada livre. A racionalidade perfeita de uma decisão pressupõe que essa decisão é tomada na plena posse pelo agente de tudo aquilo que é pensável na situação. Nestas circunstâncias, após a deliberação, revelar-se-ia um único curso racional e livre para a acção. Daí a coincidência entre liberdade, necessidade e racionalidade. No entanto esta imagem é completamente abalada por considerações empiricamente inspiradas. Uma racionalidade pressionada pelo tempo como a racionalidade humana, no contexto da qual ser racional é ser capaz de tomar as decisões necessárias em tempo real e não ponderar inteligentemente *ad infinitum*, requer, ao contrário do que a imagem racionalista prevê, insensibilidade à informação e um certo grau de arbitrariedade. Para deliberadores físicos finitos não é pensável um curso de acção que seria o melhor e o único, uma vez tudo considerado. De um ponto de vista naturalista a cláusula 'uma vez considerados todos os factores' (*all things considered*) na descrição de processos de decisão é totalmente implausível. Se não o fosse, nunca deliberações complexas como por exemplo as 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 são comuns, embora situações patológicas possam representar aproximações dessa situação<sup>64</sup>. Aparentemente, entidades racionais naturais têm que estar desenhadas de modo a ignorar mais informação a partir de certo ponto e a passar ao acto<sup>65</sup> (a dificuldade de simular esta (in)capacidade é uma das origens (suge-

---

<sup>61</sup> J. JAYNES, 1976.

<sup>62</sup> Cf. DENNETT 1983: 69 e JAYNES 1976: 245.

<sup>63</sup> Obviamente caricaturada e exagerada como hiper-racionalista.

<sup>64</sup> Cf. DAMÁSIO 1994, *O Erro de Descartes*. Lisboa, Europa América para a 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.

<sup>65</sup> A dificuldade de simular esta (in) capacidade está, de acordo com Dennett, na raiz do *frame problem* na IA.

ridas) do problema do enquadramento na Inteligência Artificial). A insensibilidade informacional, essencial à racionalidade humana, pode parecer rigidez, *sphexishness*. No entanto, sem ela a acção humana decidida não poderia existir.

Por seu lado, o ingrediente de arbitrariedade que se revela necessário à prática da racionalidade em entidades naturais demonstra uma outra diferença relativamente ao modelo clássico da racionalidade na acção, o qual incorpora uma exclusão entre a racionalidade e a espontaneidade ou criatividade. A oposição, quando se trata de conceber a motivação para a acção, entre espontaneidade e deliberação racional em seres com capacidade informacional reduzida é totalmente ilusória. De acordo com o modelo clássico da maximização racional, a racionalidade indicaria um único caminho de acção, o qual seria tomado pelo 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. Mesmo o mais racional dos humanos, sendo humano e por isso objectivamente limitado enquanto processador da informação relevante, não pode deixar de seguir por vezes o meta-imperativo racional que ordena 'deixar-se ir espontaneamente'; i. e., apostar.

#### 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<sup>66</sup>

A identificação daquilo que queremos quando queremos vontade livre com o autocontrolo conduz o teórico da acção a entidades que têm algum tipo de relação consigo próprias. Não pode existir uma teoria do agente e do controlo sem uma teoria do *self* ou proto-eu e sem uma teoria do Eu<sup>67</sup>. O *self* enquanto princípio de organização pelo qual uma entidade se distingue do resto do mundo é anterior ao Eu propriamente dito. Por outro lado, afirmou-se no início deste artigo que se considera este Eu, enquanto centro de controlo, uma questão mais básica do que a identidade pessoal, uma questão propriamente cognitiva, relativa à organização funcional subpessoal. A pessoalidade e a identidade pessoal tal como importam do ponto de vista moral e não apenas cognitivo

---

<sup>66</sup> HOFSTADTER in HOFSTADTER & DENNETT 1981: 278.

<sup>67</sup> Quer em termos de evolução de entidades, quer em termos de indivíduos humanos, o que significa que ninguém nasce sendo já um eu pleno. Um eu humano é posterior e não prévio (à acção, à experiência).

são em grande medida possibilitadas no e pelo autocontrolo de sistemas nos quais e para os quais funciona um símbolo-do-Eu e nos quais é assim criado um 'centro virtual de controlo' e um centro de gravidade narrativo'. Ao supor que o Eu resulta de um funcionamento subpessoal afasta-se a ideia de um Eu constituído e prévio que seria ele próprio a condição de possibilidade da experiência e o centro do controlo da acção. De acordo com o MEM, a personalidade e a identidade pessoal são constituídas sobre um Eu que não é originário mas posterior a um *self* e resultado da instituição de um centro virtual de controlo num sistema de agentes especialistas.

A teoria da consciência de Dennett envolvia já uma teoria da origem das razões. As razões surgem conjuntamente com as fronteiras de entidades e com os interesses destas, condições que marcam o início da divisão do mundo em bom e mau. Quando uma criatura tem interesses, passam a existir razões, *free floating rationales*, que relacionam a criatura com o ambiente. Uma criatura complexa como já o é a vespa SpheX, tomando o exemplo de Dennett para a fixidez característica de todo o comportamento determinado pelo *design*, tem muitos interesses, e age em direcção a fins hierarquizados, a partir dos fins básicos que são perseverar no seu ser e replicar-se. Mas a história da origem das razões não é apenas o início da teoria da consciência. Ela é também o início da teoria da racionalidade de agentes. Uma história natural do agente livre mostra a continuidade entre as razões (*free floating rationales*) e a racionalidade (num ser autoconsciente e capaz de se auto-agenciar). A ideia kantiana da liberdade de acordo com qual a liberdade consistiria em dar-se a si próprio as razões da acção é uma ideia indefensável do ponto de vista naturalista da TSI na medida em que 'razões' não são condições físicas no mundo mas maneiras de descrever processos causais. Uma teoria da razão na acção só pode edificar-se sobre razões que já são razões nos sistemas (*free floating rationales*) sem serem ainda razões para os sistemas (*fully appreciated rationales*). A razão na acção começa inevitavelmente por actos que são inteligentes em função de razões alheias ao actor. Basta o estudo do comportamento animal ou de sistemas artificiais verificar que estes actos ocorrem constantemente.

Agentes racionais reais não são máquinas semânticas. É preciso portanto encontrar um substituto natural da vontade numérica puramente racional. A sugestão de Dennett consiste em procurá-lo na relação dos Eus com as razões, razões que guiam o comportamento das entidades nas quais Eus funcionam como centros virtuais de controlo. Como se afirmou anteriormente, no núcleo das questões do *self* e do Eu estão as questões da auto-referência ou auto-representação num sistema cognitivo. A auto-referência é muito importante na criação de uma interioridade mental em sistemas físicos. Podem ser consideradas como situações de auto-referência progressivamente mais conscientes situações tais como 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 (i. e., eu sei que sou *eu* que penso este pensamento p). Até aqui tem-se falado da exploração da previsibilidade e da

imprevisibilidade do ambiente por agentes de modo a conseguir o controlo e o autocontrolo. Não se explicou no entanto o que está em jogo na *auto-referência* de sistemas suposta no *autocontrolo*, nomeadamente um *Eu* e/ou um *self*. O primeiro princípio de uma abordagem funcionalista é que o *Eu* não é uma coisa ou lugar no sistema mas um funcionamento, a instituição de um centro virtual, que tornará possíveis formas sofisticadas de *auto-referência* e *auto-representação*<sup>68</sup>.

Procurando compreender a constituição de uma *auto-referência* estável num sistema cognitivo apesar da instabilidade última que é a não identificação do ponto de vista objectivo com o ponto de vista subjectivo do sistema, Dennett chega ao comportamento colectivo de um sistema de indivíduos ou partes e às propriedades surpreendentes do funcionamento global. É o comportamento colectivo de um sistema de indivíduos ou partes que pode nomeadamente ter a propriedade de ser algo como um indivíduo de nível mais elevado (como a Aunt Hilary de D. Hofstadter, composta pelas formigas da colónia, em *Prelude... Ant Fugue*<sup>69</sup>, como países ou partidos políticos, etc., 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, partes que têm as suas próprias finalidades. O *Eu* envolve, no entanto, para além do comportamento colectivo das partes de um sistema, a existência nesse comportamento colectivo daquilo a que se poderia chamar um nível simbólico. Já em *The Mind's I* se propõe que a noção de *Eu* seja concebida neste quadro.

A origem dos paradoxos da *auto-referência* que frequentemente se encontram na exploração teórica das relações de um objecto consigo próprio, como é o caso com a questão do eu, está aí, e é deste modo caracterizada por D. 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'<sup>70</sup>.

O problema que se coloca em seguida diz respeito evidentemente ao poder explicativo de uma tal noção. A explicação do *Eu* (ou a explicação pelo *Eu*) é uma

---

<sup>68</sup> A história paradigmática de Dennett acerca da má maneira de abordar o problema do eu em *Elbow Room* é um relato de uma busca por Dennett do 'subway', em Londres, emergindo do outro lado da rua sem ter encontrado qualquer metropolitano (DENNETT 1983: 74).

<sup>69</sup> Cf. HOFSTADTER, in DENNETT & HOFSTADTER 1981 e também HOFSTADTER 1979, *Gödel, Escher, Bach* (um tema central é precisamente a relação entre o mental e os fenómenos da *auto-referência*).

<sup>70</sup> HOFSTADTER in DENNETT & HOFSTADTER 1981: 278.

explicação de alto nível num sistema cognitivo. Hofstadter defende que a maneira correcta de a conceber é uma maneira holista e finalista<sup>71</sup> a que chama '*soulism*'<sup>72</sup>. Do ponto de vista cognitivo, a origem do Eu radica na automonitorização, auto-representação e cruzamento de níveis num sistema cognitivo físico e um tal resultado não pode ser capturados 'a partir de baixo' e 'a partir de antes'. O holismo é considerado como causalidade dirigida para baixo, (*downward*) e opõe-se ao reducionismo (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 (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 deste é uma existência de nível elevado na organização do sistema, dependente das respostas (inseparáveis) a estímulos externos e internos. De acordo com esta teoria, é impossível separar a autopercepção da percepção e da categorização desta por meio de modelos internos. Um sistema não responde apenas a estímulos externos: 'responde' também a si próprio a vários níveis.

Note-se que não existe observação interna neste funcionamento. O que mais próximo está dessa observação interna é o já aludido símbolo-do-Eu (*self symbol*), um subsistema que é 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<sup>73</sup>.

Sistemas Representacionais como o descrito perceberão o seu próprio estado interno em termos do seu próprio repertório de conceitos. Auto-'observações' conceptuais do sistema (como 'estou confuso' pensado por um humano) são leituras de alto nível do que se passa no cérebro, que reentram no sistema para serem por sua vez pensadas.

A ideia básica de uma tal teoria é a seguinte: mostrar a um *self* um símbolo-do-Eu não constitui um redundância uma vez 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. Assim, algumas das representações para o sistema serão representações do sistema como-se-fosse-a-partir-de-fora. Estas não têm porque se identificar com 'o sistema em si' ou o sistema tal como se sente ser: apesar de serem representações do próprio sistema no próprio sistema são representações como quaisquer outras. Repita-se: não é inútil mostrar um símbolo-do-Eu a si próprio devido à cisão entre subjectivo e objectivo, à divisão em muitos sistemas cognitivos do ponto de vista de primeira pessoa e do ponto de vista de terceira pessoa acerca de si próprio, divisão esta que é inultrapassável nos sistemas cognitivos que são Eus.

---

<sup>71</sup> HOFSTADTER in DENNETT & HOFSTADTER 1981.

<sup>72</sup> Que se poderia traduzir por animista...

<sup>73</sup> HOFSTADTER in DENNETT & HOFSTADTER 1981: 200.

Admitindo a razoabilidade de uma teoria do símbolo-do-Eu (*self-symbol*), resta saber se este tem que ser consciente e se se identifica com o sentimento de consciência pessoal. De acordo com a teoria apresentada um tal símbolo do Eu não tem que ser ele próprio consciente. Embora não fosse possível pensar numa remoção do símbolo-do-Eu deixando intactos o apercebimento e a consciência da entidade<sup>74</sup>, o apercebimento e consciência da entidade não se identificam directamente com o símbolo-do-Eu. O sentimento de consciência pessoal corresponde ao ponto de vista da organização funcional na qual existe um tal símbolo, uma vez fisicamente realizada. A organização funcional na qual se encontra o símbolo-do-Eu é evidentemente subpessoal.

A partir do momento em que o Eu assim 'explicado' é também considerado 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. D. Dennett e N. Humphrey<sup>75</sup> defendem no artigo *Speaking for Ourselves* que a primeira observação que se impõe quando se analisa perturbações da identidade é que mesmo nos casos não patológicos um Eu resulta de eleições mentais, não pré-existindo. 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 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 D. Dennet e N. Humphrey fazem relativamente à Desordem de Personalidade Múltipla (ou *dissociative identity disorder*)<sup>76</sup>. 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 subpessoal) uma amálgama de muitos agentes, cada um com 'vontade' própria.

Uma crítica importante e generalizada ao tratamento da consciência por Dennett é que ele rebate a questão da consciência fenomenal, do sentir-se ser das entidades, sobre o auto-acesso, os pensamentos sobre pensamentos, e o Eu. No entanto, e é apenas isso que se pretende aqui propor, o que é contestável na teoria do Eu apresentada no âmbito do MEM não é a pretensão de que o Eu, enquanto representação

---

<sup>74</sup> Cf. DENNETT 1981: 268.

<sup>75</sup> DENNETT & HUMPHREY 1989. Para mais acerca de psicopatologia filosófica cf. GRAHAM & STEPHENS 1994, *Philosophical Psychopathology*.

<sup>76</sup> DENNETT & HUMPHREY 1989.

de unidade, não esteja presente desde o início mas apenas a pretensão de que o auto-apercebimento – para o qual, de forma reveladora, D. Dennett e D. Hofstadter em *The Mind's I* tanto utilizam a palavra ‘mente’, como a palavra ‘eu’ como a palavra ‘consciência’ – 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 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<sup>77</sup>.

#### 4.4.3. Do eu à personalidade.

Das teorias funcionalistas subpessoais do Eu resulta clara, pelo menos, a existência de um fundo irredutivelmente não mental em qualquer entidade que venha a aperceber-se de si ao nível pessoal como una e unificada e a ser moralmente uma pessoa: qualquer pessoa tem que ser primeiro um Eu, e essa é uma questão de organização cognitiva subpessoal. As ideias de acção e (correlativamente) de personalidade (correspondente a um agente autoconsciente capaz de acções voluntárias) devem ser extraídas gradualmente desse fundo por uma sequência de aproximações normativas. É o que D. Dennett faz no artigo *Conditions of Personhood*<sup>78</sup>.

Se as questões até agora consideradas – o controlo e a representação de unidade – são abordáveis de um ponto de vista de terceira pessoa, na ciência cognitiva, como questões subpessoais, com a questão da personalidade está-se já fora do nível subpessoal e da caracterização em terceira pessoa.

A teoria da personalidade é uma teoria normativa e de acordo com qualquer teoria normativa há dois aspectos a considerar. Por um lado deve existir algum cumprimento de critérios pela entidade em causa para ela ser o tipo de entidade que é, por outro lado não existe adequação integral da entidade a esses critérios. Dennett propõe-se saber se existem condições necessárias e suficientes para a consideração de entidades físicas como pessoas, partindo dos dois conceitos de pessoa a que atrás se chamou o conceito metafísico e o conceito moral de pessoa. Isto significa que ele se recusa a afirmar imediatamente que o auto-agenciamento ligado à moralidade (a situação em que alguém dá normas a si próprio) é possível devido à autoconsciência. Isso seria conceder à partida a autoconsciência, sem qualquer explicação e o problema que se coloca consiste precisamente em saber como exactamente se relacionam o *auto-agenciamento* (a base do conceito moral de pessoa) e a *autoconsciência* (a base do conceito metafísico de pessoa). Já J. Locke notara que eles se sustentam mutuamente, sustentando ambos

---

<sup>77</sup> DAMÁSIO 1994.

<sup>78</sup> DENNETT 1981b. A exposição que se segue apoia-se muito em ROVANE 1994. As principais sugestões de correcção feitas por C. Rovane são aliás aceites por Dennett (cf. DENNETT 1994, *Get Real*.)

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 reflexão 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 que, pelo que me parece, é essencial para este”<sup>79</sup>. Ao que 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”<sup>80</sup>. Para J. Locke a definição de pessoa conseguida a partir destas duas contribuições era perfeitamente geral e não se atinha necessariamente aos humanos, até agora as únicas pessoas conhecidas. Os conceitos de pessoa e de ser humano não são, assim, identificáveis já que as pessoas se distinguem por algo de mental, ou melhor, por algo que apenas pode ser descrito em termos mentais.

É ao procurar esclarecer a relação entre os dois componentes da definição lockeana de pessoa, que Dennett propõe seis condições de pessoalidade. Todas elas são condições *necessárias* da pessoalidade. No entanto, não constituem, em conjunto, condições *suficientes* para a pessoalidade<sup>81</sup>. O problema é compreender a relação de dependência entre as seis. Como se esperaria, sabendo que a distinção entre pessoa e não-pessoa é considerada uma distinção de grau, praticamente todas as condições podem ser satisfeitas por não-pessoas. Está-se longe das concepções absolutas de pessoa para as quais existe uma distinção nítida entre as entidade que possuem autoconsciência e racionalidade reflexiva e as outras.

As seis condições de pessoalidade são as seguintes:

1. Racionalidade
2. Ser objecto de atribuições intencionais
3. Dependere de uma estratégia adoptada (a Estratégia Intencional)
4. Capacidade de reciprocidade na adopção da Estratégia Intencional
5. Capacidade de comunicação verbal
6. Autoconsciência

De acordo com Dennett a ordem de apresentação das seis condições é uma ordem de dependência. O ponto central do artigo *Conditions of Personhood* é precisamente a

---

<sup>79</sup> Cf. LOCKE 1999: 443 (Livro II, XXVII, 11)

<sup>80</sup> Cf. LOCKE 1999: 459 (Livro II, XXVII, 28).

<sup>81</sup> Segundo Rovane (ROVANE 1994: 357) isto acontece porque em *Conditions of Personhood* Dennett faz uma caracterização kantiano-rawlsiana de pessoalidade moral, que faz com que Dennett oculte a variedade de imagens morais incompatíveis disponíveis para as pessoas. Sobre esta crítica repousa a proposta de Rovane de um critério ético de pessoa, que não incorre nos defeitos do (insuficiente, de acordo com Rovane) critério moral de Dennett.

análise da relação da sexta condição com as outras cinco, e portanto a análise da forma como a autoconsciência de uma entidade depende dos factores enumerados. A pessoalidade seria alcançada através de uma edificação progressiva das condições 4, 5 e 6 sobre as três primeiras condições. As seis condições constituiriam assim naturalmente dois grupos de três condições. As três primeiras são interdependentes, e aliás derivam directamente da definição de Sistema Intencional (por definição e de acordo com a TSI “ser racional é ser intencional, é ser objecto de uma certa estratégia”<sup>82</sup>). As três últimas condições põem em jogo a importância da reciprocidade e da comunicação linguística na autoconsciência e na pessoalidade. Talvez não seja possível considerá-las tão independentes como Dennett propõe em *Conditions of Personhood*<sup>83</sup>. De acordo com a TSI existem inúmeros Sistemas Intencionais (SI) sem consciência, embora a consciência de tipo humano só possa advir à intencionalidade de um SI e não a qualquer outra parte não estruturada e não desenhada do mundo físico. Noutras palavras, na ausência de *design* não poderá existir a consciência, embora possa perfeitamente existir *design* sem consciência.

Não basta no entanto, para capturar a natureza das entidades que são pessoas, afirmar que as pessoas são a subclasse dos SI constituída pelos seres que *realmente* têm crenças e desejos porque são conscientes e não apenas funcionam como se os tivessem, quando abordados através da EI. Esse movimento não apenas pressuporia o que se procura esclarecer: ele é pura e simplesmente impedido pelos princípios da TSI. 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 entidade, ‘tendo’ portanto não apenas crenças mas crenças acerca de crenças. Vários animais são SI de segunda ordem: vários animais são por exemplo capazes 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 o desejo de induzir crenças falsas no outro. Retomando um exemplo favorito de Dennett, existem pássaros que fingem ter uma asa partida para enganar o predador, que assim os considera presa fácil e os persegue, afastando-se do ninho onde estão as crias do pássaro, que são assim salvas, 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

---

<sup>82</sup> DENNETT 1981b:271.

<sup>83</sup> É a opinião de Rovane (ROVANE 1994).

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 autoconsciência (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' – na expressão de Elizabeth Anscombe que Dennett procura desde sempre obsessivamente compreender – mas que não corresponde a intenções conscientes dos intervenientes na situação nem a eventos, por exemplo neuronais, actuais. Existe inteligência nestas situações e comportamentos, mas essa inteligência não é inteligência do agente, é inteligência no agente. O facto de os comportamentos que envolvem 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. No entanto, a reciprocidade da atribuição de crenças acerca de crenças é um grande passo em direcção ao tipo humano de personalidade (poder-se-ia afirmar que se trata de ser 'sujeito' e não apenas objecto da EI, embora esta afirmação possa criar problemas).

Em *Conditions of Personhood* Dennett considera que a análise griceana da significação<sup>64</sup> é esclarecedora relativamente à diferença que uma linguagem natural (condição 5) faz num SI. O que é particularmente esclarecedor é a célebre ideia de significação não natural (*non natural meaning*), que P. 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 um sistema B reconheça que A (in)intenciona que B produza uma dada resposta. Para começar, 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 usam 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. Ele 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. Em suma, Grice pretende mostrar que se a comunicação linguística humana envolve a possibilidade de manipulação é porque ela depende da confiança. É a existência de

---

<sup>64</sup> GRICE 1957.

expectativas determinadas que permite a subsunção de intenções, a introdução de um outro discurso nas entrelinhas daquilo que é dito<sup>85</sup>. 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 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 que Grice analisa. 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 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 autoconsciência. Isto significa que todas estas condições da proto-agência ou protovoluntariedade (considerando que é a voluntariedade a característica propriamente pessoal) são obtidas sem qualquer apelo à consciência. Apenas a partir destas cinco condições seria (estranhamente) 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, a partir do momento em que esta é considerada como parte da teoria da escolha racional. Como é sabido, de acordo com a teoria da justiça de Rawls<sup>86</sup> o acordo quanto aos princípios da justiça não supõe qualquer bondade ou moralidade prévias das partes contratantes. A solução óptima (o acordo quanto aos princípios) é apenas a solução racional pela qual calculadores racionais interessados no seu próprio bem, e devendo levar em conta as crenças de outrem, devem optar. Esse dever não é desde logo 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 a seguinte situação: assim como acreditar o verdadeiro é a norma da crença<sup>87</sup> e

---

<sup>85</sup> 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-nos, segundo Pinker como "scheming, second-guessing, social animals" (PINKER 1994: 230), característica esta que constitui a base do humor, da metáfora, dos eufemismos de etiqueta, etc.

<sup>86</sup> RAWLS 1971, *A Theory of Justice*.

<sup>87</sup> A ideia é defendida não apenas por Dennett como também por Davidson, Fodor, etc.

dizer o verdadeiro é a norma da enunciação<sup>88</sup>, tratar o outro como pessoa é a norma da interacção pessoal<sup>89</sup>. Que esta norma é uma norma em relação à qual os desvios são constantes é demasiado claro: a interacção de humanos não é de modo algum necessariamente pessoal e justa, no sentido descrito.

Nada na exploração das primeiras cinco condições explica porque é que podemos (se é que podemos) considerar por vezes pessoas individuais concretas como responsáveis pelas suas acções. Tudo o que foi afirmado serviria até para fornecer argumentos no sentido de defender que nunca o podemos fazer. Até justiça como regulação da interpessoalidade parece poder existir sem qualquer autoconsciência 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<sup>90</sup>, por exemplo, comenta que as três últimas condições de pessoalidade se põem conjuntamente e que as condições 4 e 5 (a capacidade de reciprocidade na adopção da Estratégia Intencional e a capacidade de comunicação verbal) só podem referir-se a relações que são já interpessoais, no sentido em que envolvem a captação dos princípios normativos que constituem a racionalidade reflexiva e que Dennett liga exclusivamente à condição 6. A racionalidade reflexiva constitui, evidentemente, o topo da construção da pessoalidade levada a cabo em *Conditions of Personhood*. A importância da sugestão é a seguinte: a autoconsciência (Condição 6) é mais do que auto-apercebimento, auto-referência e auto-representação, é mais do que ser considerado como racional e intencional por outras entidades que por sua vez são consideradas como racionais e intencionais. A autoconsciência que importa para a constituição de uma pessoa é a *auto-avaliação* e um comunicador griceano pode não ser capaz desta auto-avaliação.

É através da tese de Harry Frankfurt<sup>91</sup> acerca da ligação entre a liberdade da acção e o conceito de pessoa que fica esclarecida a última condição de pessoalidade, a autoconsciência como racionalidade reflexiva e auto-avaliação. Ser livre aparentemente consiste em fazer aquilo que se deseja. No entanto, H. Frankfurt chama a atenção para o facto de ser possível ser-se livre de se fazer o que se deseja e não 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 implica ser-se livre para se querer o que se quer querer. Assim, a liberdade da vontade 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).

---

<sup>88</sup> GRICE 1957.

<sup>89</sup> Carol Rovane faz o seguinte reparo a este princípio kantiano: seria conveniente um recuo do kantismo que abrisse a possibilidade de desacordo ético entre as pessoas para não haver consequências indesejáveis na teoria.

<sup>90</sup> ROVANE 1994.

<sup>91</sup> FRANKFURT 1971, *Freedom of the Will and the Concept of a Person*.

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 controlo resultante da avaliação reflexiva que entra assim em jogo existe a autoconsciência que estabelece definitivamente a personalidade. Note-se que a sugestão de H. Frankfurt sobre a relação da liberdade da vontade com o conceito de pessoa se traduz numa teoria sobre a natureza da liberdade de acordo com a qual é a *incapacidade de deliberar racionalmente sobre o querer* (e não o determinismo) que impede 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<sup>92</sup>. Aparentemente ela envolveria uma capacidade ilimitada e recursiva de auto-referência, uma capacidade de pensamentos sobre pensamentos sobre pensamentos sobre pensamentos (...) e de desejos acerca de desejos acerca de desejos (...). A estrutura não é no entanto ilimitadamente iterada, como se verá.

O controlo resultante de avaliação reflexiva (i. e., o controlo racional de si e das acções que constituem personalidade e identidade) supõe a adopção em relação a si próprio, por um sistema físico, causal, do papel de exigidor de razões<sup>93</sup>. Se existisse um eu naturalmente e essencialmente uno ou uma pessoa 'absoluta' este autocontrolo meta-reflexivo através de razões pareceria redundante. Mas esse eu uno não pode ser dado como garantia *a priori*, pois é ele próprio um efeito do funcionamento subpessoal do sistema e as 'pessoas' pressupõem sistemas constituídos por um tal processo de centralização virtual pela auto-representação global. De acordo com Dennett, apenas o controlo meta-reflexivo representa a autoconsciência genuína. A autoconsciência encontra-se assim indiscernivelmente ligada com a agência racional. Uma pessoa é autoconsciente porque é um agente racional, um auto-avaliador e um autocontrolador meta-reflexivo, que pode ser considerado livre na medida em que se auto-agencia racionalmente. J. Locke, como nota Dennett, sabia que uma parte muito importante do pensamento humano era o pensamento acerca do pensamento, o pensamento acerca das crenças e desejos próprios, suposto no poder de os interromper e controlar. De acordo com Locke, tendo a mente "um poder de suspender a execução e a satisfação de qualquer um dos seus desejos (...) está em *liberdade* para considerar as finalidades dos mesmos,

---

<sup>92</sup> Dennett abandona a referência às ordens dos SI quando chega à sexta condição. C. Rovane nota de forma perspicaz que entretanto Dennett começou a usar dois critérios diferentes para avaliar a sofisticação da racionalidade: a ordem dos SI mas também o envolvimento em actividades de avaliação. Estas não requerem necessariamente ordens muito mais elevadas do que as já mencionadas. Ora este segundo critério não é de modo nenhum continuista (i. e., contradiz a ideia de que capacidades racionais e sociais são uma questão de grau e não existe nenhuma corte qualitativo); há um único tipo de seres envolvidos em actividades de auto-avaliação. Rovane duvida que as capacidades racionais que envolvem pensamento de ordem elevada sejam assim separáveis da racionalidade reflexiva (i. e., ela pensa que os pensamentos de ordem elevada pressupõem já racionalidade reflexiva).

<sup>93</sup> Observe-se o quanto esta ideia é paralela à defesa da importância das práticas de estimulação linguística para a consciência no âmbito do MEM.

para os examinar em todos os ângulos, e para os comparar (...) Aqui se baseia a liberdade que o homem possui (...) temos um poder de suspender a realização deste desejo, como todos podem experimentar em si próprios no dia-a-dia<sup>94</sup>.

Dennett virá a admitir uma crítica de Carol Rovane<sup>95</sup> segundo a qual a agência racional de que se fala em *Conditions of Personhood* envolve 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 seriam constitutivos da racionalidade que interessa Dennett em *Conditions of Personhood*, a racionalidade considerada enquanto unidade de um ponto de vista racional, a racionalidade própria, para o próprio, constitutiva da pessoalidade, nuclear para a identidade pessoal.

O artigo *Conditions of Personhood* culmina assim com a defesa da ideia segundo a qual apenas em situação de autocontrolo meta-reflexivo a 'ordem que está lá', 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.

#### 4.4. 'Quererei eu realmente ser aquilo que sou?': Os metaproblemas difíceis do controlo, a avaliação forte e a sorte moral.

"Thinking affects 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 they are trying to inspect!)"

Marvin Minsky, *The Society of Mind*, 6.3<sup>96</sup>

"Nothing is entirely one's own making, unless one is God"

Daniel Dennett<sup>97</sup>

O capítulo *Self-made-selves* de ER<sup>98</sup> tem como propósito estabelecer a ligação entre a constituição unificadora e virtual do Eu, ela própria apoiada sobre a organização do *self*, e a identidade pessoal. Noutras palavras, trata-se de reportar a constituição da identidade pessoal no sentido moralmente relevante, à possibilidade de estratégias de controlo racional reflexivo sucessivamente mais sofisticadas.

---

<sup>94</sup> LOCKE 1999: 345 (Livro II, XXI, 48).

<sup>95</sup> ROVANE 1994.

<sup>96</sup> MINSKY 1985, *The Society of Mind*.

<sup>97</sup> DENNETT 1983: 85.

<sup>98</sup> DENNETT 1983: 74.

Se o Eu é uma auto-representação de unificação, constituída nas operações de um sistema físico, e uma representação que possibilita processos de metacontrolo deste, o perigo, do ponto de vista do filósofo moral, de uma explicação totalmente cognitiva e funcional, é não restar nada nem ninguém acima do funcionamento cognitivo, nenhum eu para ser 'responsável pelo eu' nas palavras de Charles Taylor<sup>99</sup>. Se as coisas acontecem, e ninguém faz alguma coisa, se o 'eu' é algo de virtual, um mero funcionamento unificante tendencialmente inimputável, não existem realmente agentes morais. Mas não é certamente como um eu superficial e irreal 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*). No entanto, nada garante à partida que esse apercebimento não seja totalmente ilusório.

Mas mesmo que a doutrina da causação do agente seja uma má doutrina, é certo no entanto que existem agentes e embora os agentes possam não ser algo como primeiros motores não movidos, a ilusão de que o são é bem real. É então preciso saber o que, do ponto de vista cognitivo, causa essa ilusão. Dennett oferece três razões:

1. uma ilusão de escala no nosso auto-apercebimento, causada pela amplificação dos resultados do processamento de informação no sistema nervoso<sup>100</sup>.

2. uma má avaliação natural da fenomenologia da decisão e do controlo, que nos conduz a postular um eu activo.

3. o facto de as identidades pessoais resultarem<sup>101</sup> de um processo de autocriação, i. e., de auto-avaliação.

É este último aspecto que possibilita o cruzamento da questão cognitiva 'mecânica' com a questão do conteúdo. Se a autocriação do agente não é por exemplo uma "escolha radical sartreana" ela não pode senão consistir num *processo de auto-avaliações sucessivas*. Dennett convoca as ideias de C. Taylor acerca de avaliação forte e de responsabilidade pela identidade própria de modo a caracterizar este processo de auto-avaliação. Ao problema da avaliação forte chamar-se-á aqui o problema difícil do controlo<sup>102</sup>.

---

<sup>99</sup> Cf. TAYLOR 1976, *Responsibility for Self*.

<sup>100</sup> Em nós próprios a introspecção mostra-nos 'efeitos dramáticos' e notórios tais como movimentos intencionados do corpo, e portanto grandes gastos de energia armazenada no corpo, iniciados e controlados por processos que usam muito pouca energia como input (e 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.

<sup>101</sup> Resultarem mais ou menos conforme se trate de pessoas mais ou menos 'profundas' (a noção é de C. Taylor e será em seguida caracterizada), e, mais ou menos morais. 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 esta auto reflexão ou avaliação forte que torna alguém responsável pela pessoa que é.

<sup>102</sup> Por analogia com o problema difícil da consciência, na terminologia de David Chalmers.

Taylor contrasta a avaliação forte com o tipo de avaliação característica de qualquer entidade que deseja e que portanto valoriza favoravelmente aquilo que deseja. Neste último caso, para que alguma coisa seja boa basta que seja desejada, enquanto que pela avaliação forte (também chamada por Taylor avaliação qualitativa ou reflexiva) alguém pode julgar como maus ou como indesejáveis certos desejos seus. A avaliação forte é assim uma proposta acerca da natureza de seres que cuidam do seu ser, que o determinam e o avaliam. Este é, segundo C. Taylor, um traço da moralidade ocultado pela predominância da caracterização utilitarista, não qualitativa, dos desejos ou 'preferências' de agentes. O avaliador forte faz mais do que uma pesagem dos seus desejos: ele estabelece uma ligação entre a avaliação dos seus desejos e a sua identidade, tornando-se assim responsável pelo seu próprio ser.

Admita-se, provisoriamente, que a identidade pessoal envolve nestas circunstâncias de avaliação algo como 'representações de si próprio produzidas por um ponto de vista racional único.' De novo, nada garante que a identidade pessoal dependa apenas desta representação e não do mundo<sup>103</sup>. No entanto, a ideia de um 'ponto de vista racional único' abala já os tradicionais critérios da identidade pessoal, o critério físico e o critério fenomenológico. Defender-se-á aqui, a partir de Dennett, que estes se revelam insuficientes e que é necessário um critério adicional, um critério normativo, da identidade pessoal. Este é proposto por exemplo por C. Rovane como um complemento natural da teoria das condições de pessoalidade.

Propor um critério normativo da identidade pessoal traduz-se na suposição de que a identidade pessoal depende do compromisso envolvido num certo exercício da racionalidade a partir de um ponto de vista único. A unificação aí em causa é qualquer coisa mais do que a continuidade física e/ou a continuidade psicológica. Aquilo que faz com que uma pessoa que se identifique como uma e a mesma pessoa é um compromisso com a racionalidade de uma vida mental, aí incluindo a constituição de si por avaliação, que torna esta uma vida mental *própria*, nomeadamente devido ao compromisso de levar em conta episódios mentais de deliberação para a prossecução de fins. É também assim que se estabelece a unidade da pessoa, a identidade pessoal no sentido normativo<sup>104</sup>.

---

<sup>103</sup> Cf. NAGEL 1986:39-40. Segundo Nagel, "I am whatever persisting individual in the objective order underlies the subjective continuities of that mental life that I call mine" (NAGEL 1986:40). Nesse sentido a voluntariedade e a perspectiva interna que adiante se focarão a partir da ideia de ponto de vista racional único não serão suficientes para capturar a identidade pessoal, e por isso "our true nature and the principle of our identity may be partially hidden from us" (NAGEL 1986: 39).

<sup>104</sup> Rovane sugere que serão assim vistas a uma nova luz as pessoas múltiplas e as pessoas grupais (tais como países, organizações, etc.), na medida em que também nestes casos é possível falar de fenómenos de eu no sentido organizacional e funcional – é o que Dennett faz por exemplo na sua análise da desordem de personalidade múltipla (cf. DENNETT & HUMPHREY 1989 e o comentário em ROVANE 1994). Torna-se problemático saber se isso significa a existência de autoconsciência, uma vez que Dennett praticamente identifica esta com a unificação acrescida da prática de racionalidade. É certo que Dennett afirma que os múltiplos (i. e., as múltiplas personalidades num indivíduo) não são pessoas, mas eus. No entanto, quando Rovane desafia Dennett a aceitar as pessoas grupais (ROVANE 1994) ele aceita o desafio (DENNETT 1994); ele

As auto-avaliações supõem portanto a representação e a avaliação das crenças e desejos próprios. Nas auto-avaliações estão envolvidas – na terminologia de H. Frankfurt<sup>105</sup> – volições de segunda ordem. Com a auto-avaliação ou metaproblema difícil do controlo a questão cognitiva da unificação dos agentes e do autocontrolo prolonga-se até ao domínio ético. São os processos de auto-avaliação que possibilitam a personalidade e a identidade pessoal (possibilitam as pessoas profundas, na terminologia de C. Taylor). O fundamento da responsabilidade de uma pessoa pela sua própria identidade é precisamente esta re-avaliação radical. Um avaliador forte é 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. Isto 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 uma voluntariedade e de um compromisso.

Se segundo C. Taylor a responsabilidade pelas avaliações fortes não se identifica com a escolha radical, a escolha que não é fundada em qualquer razão, através da qual por exemplo J.-P. Sartre caracteriza o projecto que cada pessoa é, é porque o agente da escolha radical sartriana não escolhe segundo razões, não sendo portanto um avaliador forte<sup>106</sup>. Como C. Taylor vê, 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, conclui-se que a prática sem limites da auto-avaliação parece tão irracional como a rejeição rígida de qualquer auto-avaliação. É evidentemente bastante problemático saber se uma auto-avaliação sempre crescente melhorará algo a que se poderia chamar o ‘carácter’ do agente. O excesso de autoconhecimento não parece ser bom para nenhum sistema. O autoconhecimento total de entidades naturais que são agentes cognitivos é desde logo impossível: como diz Dennett, é logicamente impossível ter um conhecimento total dos microestados e microtransições que subjazem ao aspectos vários do auto-apercebimento. Algo de semelhante já foi aliás referido a propósito da teoria cognitiva do Eu<sup>107</sup> e C. Taylor<sup>108</sup> considera precisamente que as mais profundas auto-avaliações, aquelas que constituem as pessoas ‘profundas’, são as mais sujeitas a distorção e as

---

tem menos dificuldades em atribuir personalidade às pessoas grupais do que aos múltiplos. Aparentemente obter-se-ia assim uma estranha distribuição da (auto)consciência.

<sup>105</sup> FRANKFURT 1971.

<sup>106</sup> A escolha radical de Sartre é, como se sabe, uma interpretação do pensamento de Heidegger, em *Sein und Zeit*, acerca do sendo em cujo ser se trata do seu próprio ser (*das Seiende, das in seinem Sein um dieses selbst geht*).

<sup>107</sup> Cf. DENNETT 1983:61, acerca da impossibilidade do autoconhecimento completo.

<sup>108</sup> “It is precisely the deepest evaluations which are least clear, least articulated, most easily subject to illusions and distortion. It is those which are closest to what I am as a subject in the sense that shorn of them I would break down as a person, which are among the hardest for me to be clear about” (TAYLOR 1976: 296).

que mais se arriscam a pura e simplesmente provocar o abalo ou o desmoronamento do auto-avaliador. 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 possibilita a constatação do carácter criado e contingente daquilo que se encontra (aquilo que se é). De facto, porque não está disponível uma metalinguagem que permita a avaliação das avaliações mais básicas, o agente moral não pode 'saltar para fora de si próprio' (de resto, como diria D. Hofstadter, este problema da auto-referência não acontece apenas moralmente). É essa a razão por que C. Taylor defende que as avaliações fortes não são descrições mas articulações, tentativas de formular o que é inicialmente não formulado<sup>109</sup>. No entanto, as avaliações fortes não são arbitrárias (i.e., sem razão) como Sartre pretendeu que as escolhas radicais fossem. Elas são arbitrárias apenas no sentido em que não está disponível uma metalinguagem para as avaliar. De acordo com Dennett, a razão da 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.

Esta ideia de criação da identidade pessoal através da avaliação forte é evidentemente vulnerável às críticas que chamam a atenção para o incontornável papel da sorte moral na constituição das pessoas através das suas acções. 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 avaliação do seu próprio ser. A despossessão e a involuntariedade, mais do que a voluntariedade do avaliador forte de C. Taylor, caracterizariam a constituição de qualquer identidade pessoal, de várias maneiras. 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 autocriação acima tratada seria sobretudo questão de sorte. Ora, parece descabido e absurdo falar de culpa ou mérito dos agentes quando está em causa a sorte, i.e., a intervenção de factores que escapam totalmente ao controlo do agente

O motivo da sorte moral foi tratado por B. Williams<sup>110</sup> e T. Nagel<sup>111</sup>. Fala-se de sorte moral (boa ou má) "quando um aspecto significativo daquilo que alguém faz depende de factores fora do seu controlo, e no entanto continuamos a tratar essa pessoa a esse respeito como objecto de juízo moral"<sup>112</sup>. 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 por pessoas):

---

<sup>109</sup> TAYLOR 1976: 295.

<sup>110</sup> WILLIAMS 1981, *Moral Luck*.

<sup>111</sup> NAGEL 1979a, *Moral Luck*.

<sup>112</sup> NAGEL, 19791: 26.

o tipo de pessoa que se é (as inclinações, as capacidades, o 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 antikantiana (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<sup>113</sup>, a atracção do kantismo é precisamente parecer justo, fazendo crer que todo o agente racional está em idênticas condições 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 está evidentemente a afirmar que a justiça do kantismo é totalmente ilusória. O facto é que as acções são moralmente avaliadas de uma forma que inclui factores dependentes da sorte, mesmo que isso pareça injusto ou irracional. Não é uma solução para a teoria moral o isolamento kantiano do momento da acção praticada por dever. Num tribunal, por exemplo<sup>114</sup>, um réu será julgado por um crime diferente (homicídio 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. Recordando um exemplo-chave de B. Williams<sup>115</sup>, a atitude de Gauguin que abandonou a família para pintar será sempre avaliada pela qualidade e beleza da obra que ele assim produziu. Como afirma Williams, apenas o sucesso justificará Gauguin. Se ele falhar, terá agido *erradamente, sem justificação. Se não falhar, as coisas serão diferentes*. No entanto, Gauguin, no momento em que decide, não pode obviamente prever se o sucesso acontecerá ou não<sup>116</sup>. Se a justificação é retrospectiva, as considerações não podem ser aplicadas no momento da deliberação. Trata-se precisamente de sorte.

---

<sup>113</sup> WILLIAMS 1981:21.

<sup>114</sup> O exemplo é de T. Nagel, em NAGÉL 1979a:29.

<sup>115</sup> WILLIAMS 1981:22-26.

<sup>116</sup> Note-se que Williams não supõe que Gauguin possa alguma vez justificar-se perante aqueles que abandonou ou perante si próprio relativamente ao abandono. Ele pretende apenas sublinhar que a avaliação moral da situação será, inevitavelmente, profundamente diferente no caso de existir sucesso.

#### 4.5. Sorte Moral ou Responsabilidade. Deliberação. A prudência e o design do deliberador.

De acordo com Dennett, quando se concede um ênfase excessivo à noção de sorte moral corre-se o risco de esquecer o espaço de manobra entre o sucesso e o fracasso das acções. Em *Elbow Room* Dennett pretende que o argumento baseado na ideia de sorte moral contra a possibilidade de responsabilidade não se sustenta na medida em que a sorte moral e a falta de sorte moral não são situações complementares, não exaurem as possibilidades. É certo que agentes humanos não são totalmente responsáveis por serem responsáveis. No entanto, eles 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 e ele propõe que se trata de possibilidade epistémica. É a possibilidade epistémica – subjectiva pois diz respeito ao que o agente sabe ou não sabe – que importa para a questão da acção e não a possibilidade lógica nem a possibilidade nómica. É a possibilidade epistémica que subjaz ao *design* de um deliberador. Esta é de novo, note-se, uma caracterização subpeçoal.

Afirmar que a possibilidade que subjaz ao *design* do deliberador é a possibilidade *epistémica* é afirmar que a situação de deliberação pressupõe que o deliberador distingue estados de coisas que conhece ou quer conhecer de outros estados de coisas. É esta distinção que 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 possíveis estados incontrolláveis e imprevisíveis, estados controláveis e estados previsíveis.

Assim, do ponto de vista do deliberador 'agir sob a ideia de liberdade' significa em termos concretos que ele:

(1) Prevê o ambiente de forma rápida e confiável. Esta é uma tarefa de tratamento de informação.

(2) Considera no ambiente traços fixos, traços desprezáveis e traços mutáveis dos quais vale a pena manter a pista, sejam previsíveis em certas condições, sejam caóticos.

O ponto da gestão de informação pelo deliberador é chegar a tomar uma decisão que seja boa antes que seja tarde de mais<sup>117</sup>. Essa decisão baseia-se nos desejos do deliberador e em expectativas quanto ao que acontecerá de qualquer modo, e ao que acontecerá se algo não for feito. A autoprevisão é um componente importante nesta deliberação. Também ela faz parte da cena prevista, e tem várias categorias, que cobrem nomeadamente aquilo que é controlável e aquilo que é incontrollável do ponto de vista do deliberador. A autoprevisão que visa a própria decisão futura engloba, é claro, um risco de regressão infinita.

---

<sup>117</sup> Esta questão ligar-se-á ao problema do enquadramento (*frame problem*) na IA.

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 mas também capacidades de agentes cognitivos não humanos. Deliberação e decisão são problemas de *design* de artefactos inteligentes. Nomeadamente, saber como investir o tempo da deliberação *em algo que não seja irrelevante é um problema fulcral para o design* de um 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 fosse pintar um quarto e que para tal considera, juntamente com milhares de outras alternativas, virar este 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. Oportunidades são 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 o que é evitável e o que é inevitável na previsão do deliberador. Todo o deliberador delibera sobre um fundo de antecipação do futuro no quadro de uma determinada imagem manifesta. Só assim oportunidades e evitamentos têm sentido na vida deliberativa de auto-controladores.

Considera-se usualmente que o 'poder ter feito de outra maneira' (*could have done otherwise*) é um critério pacífico da responsabilidade de um agente por uma acção. No entanto, no âmbito da análise dos deliberadores deterministas em ER Dennett defende que o princípio é de alguma forma falso. 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?' após um acto lamentável. Se esta é uma questão metafísica acerca de cursos alternativos do mundo, ela é uma puramente especulativa, irrespondível e esgota-se em si própria. Dennett sugere que o que tem interesse de facto do ponto de vista do agente é a capacidade de determinar o seu próprio carácter de modo a, precisamente, ser um agente tal que *não poderia fazer de outro modo* perante determinadas circunstâncias. A capacidade que os agentes humanos têm de se fazer ser de um modo tal que perante circunstâncias determinadas não poderiam agir de outro modo 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 a liberdade contracausal mas sim 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 heu-

risticos: 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 da racionalidade e da 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 a informação de que idealmente necessitaria para agir é o 'pecado original naturalizado'. No mundo dos deliberadores deterministas os eventos podem ser causalmente determinados não sendo inevitáveis, existem possibilidades epistémica oportunidades, controlo e autocontrolo. Como já se afirmou atrás, está assim constituído espaço de manobra suficiente para a acção livre. Se Fodor, em *The Language of Thought*<sup>118</sup>, aponta como 'argumento' contra a ideia wittgensteiniana de impossibilidade de uma linguagem privada a existência das linguagens de computadores, poder-se-ia aqui apontar os usos das lógica epistémica e temporal na robótica como prova do espaço de manobra que se abre aos deliberadores deterministas. Repita-se que a possibilidade que importa para o deliberador determinista é compreendida quando se evoca uma geração sem padrão e não uma geração não-causada. Um agente moral que é um deliberador determinista capaz de autocontrolo meta-reflexivo e de avaliação forte é certamente capaz de acções livres embora possa não ser culpado no sentido kantiano ou sartriano.

## 5. Manual de primeiros socorros morais e ética da virtude.

Como se afirmou no início do presente artigo, um manual de primeiros socorros morais constitui uma necessidade para agentes que agem em tempo real. A ideia não representa uma cedência na ambição do pensamento moral face aos detalhes imperfeitos da prática mas sim uma tentativa de incorporação na teoria filosófica da racionalidade e da moralidade do conhecimento empírico acerca das imperfeições constitutivas da racionalidade de agentes reais. As ideias de prudência naturalizada e de manual de primeiros socorros morais nascem da constatação de que o processo de *satisficing* é fundamental em grande parte dos processos reais de escolha ou decisão, desde as 'escolhas' da selecção natural que conduziram ao *design* de agentes reais até aos processos humanos conscientes de decisão. Dennett defende que o processo de *satisficing*<sup>119</sup> é a estrutura básica de toda a tomada de decisões, moral, prudencial, económica, evolutiva<sup>120</sup>.

---

<sup>118</sup> FODOR 1975, *The Language of Thought*.

<sup>119</sup> Os processos de *satisficing*, teorizados nomeadamente por Herbert Simon, são processos de decisão que encontram soluções razoavelmente boas para problemas em situações muito complexas, mas não soluções óptimas. Estas são inalcançáveis na medida em que é impossível medir todas as alternativas por uma função de utilidade comum.

<sup>120</sup> DENNETT 1995: 504.

Agentes racionais resultantes de evolução por selecção natural não são agentes perfeitamente racionais, nem poderiam sê-lo. O seu *design*, inclusive o *design* para a cognição, resulta ele próprio de um processo de *satisficing*, e na deliberação e decisão de que são capazes trata-se ainda de um processo de *satisficing*. Se a continuidade evolucionista que liga organismos e processos mentais é um sustentáculo do gradualismo da teoria da acção de Dennett, curiosamente, a atenção concedida às situações de deliberação e de decisão conduz Dennett do darwinismo às éticas da virtude. Como se sabe, as éticas da virtude, ao contrário do que acontece nas éticas da maximização racional, as considerações acerca do carácter dos agentes morais 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. A referência à importância do carácter na deliberação e decisão morais opõe-se à desconexão das acções e decisões relativamente à vida do agente que é característica das éticas da maximização racional. Ora, por imperativos biológicos e por pressão temporal, sugere Dennett, somos constituídos de modo a agir-mos de uma forma que se aproxima mais da situação caracterizada pelas éticas da virtude, i. e., somos regidos nas situações de ponderação prática sob pressão temporal por algo como o carácter e não por maximização racional. É preciso notar que a noção de carácter, que é obviamente uma noção de âmbito pessoal nas éticas da virtude<sup>121</sup>, é aqui de certo modo transposta para o âmbito subpessoal. É possível no entanto estabelecer uma ligação entre caracterizações subpessoais e teorização moral.

Se a visão naturalizada e realista da racionalidade e da responsabilidade 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, ela fá-lo de um ponto de vista quase apenas descritivo, i. e., oferecendo uma teoria cognitiva subpessoal da deliberação. No entanto uma tal caracterização estrutural dos deliberadores reais tem um peso determinado na 'escolha' racional de uma filosofia moral. A consequência que Dennett aponta é precisamente um afastamento das atrás chamadas éticas da maximização racional, por falta de pertinência psicológica.

De modo a pôr em relevo as características do processo a que H. Simon chamou *satisficing*, e que se considera ser fundamental e característico de grande parte das 'escolhas' reais, Dennett propõe o seguinte exemplo de deliberação seguida de decisão<sup>122</sup>. 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, 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?

---

<sup>121</sup> Sendo constituído por uma ligação entre identidade pessoal escolhida (no quadro do que Taylor chama avaliação forte, por hipótese) e deliberação conducente à acção. Aquilo que conduz a decisões seria assim menos um cálculo racional do que aquilo que se é. Esta é uma maneira de contornar as idealizantes e irrealistas propostas das éticas da maximização racional.

<sup>122</sup> Cf. DENNETT 1988.

A primeira coisa a constatar é que é impossível considerar todos os factores relevantes no tempo disponível. Logo, será imprescindível o uso cego e impiedoso de regras grosseiras de selecção e de decisão (no caso, por exemplo um critério não absolutamente confiável mas não irrelevante tal como a média das classificações). Além disso, o decisor abdicará do controlo total do 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 cercar 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 último processo é uma tentativa, resultante de automonitorização, de corrigir o resultado final.

Evidentemente, qualquer decisão tomada nestas circunstâncias é enormemente vulnerável a um olhar retrospectivo. Não existe qualquer garantia de que a alternativa que acabou por ser seleccionada seja a melhor alternativa. 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, de uma forma que está de acordo com o MEM, as características dos processos de deliberação e escolha acontece em deliberadores reais (por exemplo nos seres humanos, por exemplo – mas não apenas – nas suas decisões morais). Ora, estas características da deliberação e da decisão não são usualmente encaradas de forma séria na caracterização da deliberação e decisão moral feita pelas principais correntes da filosofia moral, i. e., quer por utilitaristas quer por kantianos (ou melhor, são afastadas com a consideração de que qualquer situação real está longe de ser aquilo que deveria ser por princípio). De acordo com o utilitarismo, o que é relevante na deliberação e decisão morais é maximizar a utilidade e felicidade do maior número, através da maximização da utilidade esperada, em ordem à satisfação de preferências. Supostamente, ou pelo menos idealmente, deveria existir um algoritmo para a análise das utilidades esperadas que fosse utilizável em situações concretas. A deontologia por seu lado requer que o dever seja feito e portanto que uma regra racional determine o querer. O problema são evidentemente as considerações que conduzem à determinação de uma regra racional, excluindo outras regras candidatas. O utilitarismo e a deontologia concordam num ponto importante: ambos exigem, idealmente, que sejam consideradas todas as coisas anteriormente à decisão. Ora, de acordo com Dennett, “nenhum 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”<sup>123</sup>. Considerar todas as coisas é impossível para agentes reais, e Dennett pensa que é importante que a filosofia moral leve a sério tal impossibilidade. Ela caracteriza a própria natureza dos agentes morais reais e não apenas limitações contingentes. Agentes morais são por natureza e não por uma limitação desprezável do ponto de vista teórico objectivamente limitados. Este é um sentido em que os juízos *all things considered* referidos no ponto anterior como marca distintiva da existência de um ponto de vista racional que constitui o aspecto normativo da identidade pessoal são *realmente* impossíveis. Por esta razão, Dennett sugere que é preferível conceber a deliberação moral como um exercício de disposições (‘virtudes’) mais do que como maximização racional. Entre outras coisas entende-se por maximização racional a seguinte situação numa deliberação que conduz à acção:

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 racionais e de acções guiadas por razões. Dennett defende no entanto que as pessoas não são agentes capazes de considerar todos os factores em jogo numa situação de decisão e que as decisões tomadas são descritas de forma mais realista como resultando de processos heurísticos. Poder-se-ia dizer que as pessoas deverão então ser guiadas por algo como um ‘cuidado não maximizante’ e não por maximização racional, fazendo a todo o momento escolhas não maximizantes. Essa estratégia não é irracional por várias razões, nomeadamente devido à insuficiência da informação e à pluralidade dos ‘bens’ em consideração.

Noutras palavras, aquilo que Dennett caracteriza em termos de *design* do deliberador, e portanto ao nível subpessoal e de um ponto de vista cognitivo, corresponde ao que os filósofos morais caracterizam ao nível pessoal como *carácter e virtude*. Dois tipos de considerações nascem desta consideração da deliberação e decisão racionais. Um primeiro tipo, tratado em seguida, diz respeito à filosofia moral, à questão concreta de encontrar estratégias de melhoramento das decisões. Um segundo tipo diz respeito à natureza da racionalidade. A racionalidade é o ponto de apoio da TSI, a sustentação da teoria do conteúdo e não é enfrentada directamente no presente artigo. No entanto, é pelo menos óbvio que a abordagem naturalista põe em revelo a imperfeição constitutiva daquilo a que se chama racionalidade nos agentes reais. Racionalidade é a melhor adaptação possível do comportamento de agentes racionais ao seu ambiente, consideradas as suas crenças e desejos. No entanto não existe nenhum ponto de vista exterior em função do qual se possa falar de Racionalidade *tout court*, de Racionalidade maiúscula. O erro comum às éticas da maximização racional é precisamente supor que “deve haver

---

<sup>123</sup> DENNETT 1995: 500.

uma única (melhor ou mais alta) perspectiva a partir da qual é possível avaliar a racionalidade ideal<sup>124</sup>.

É relativamente à questão do melhoramento das decisões morais que Dennett propõe o manual de primeiros socorros morais. O manual, que diria o que fazer imediatamente, e que teria como finalidade criar estados disposicionais dos agentes morais, que têm que decidir, sem tempo para calcular todas as implicações do acto, 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 metaproblemas 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 assim 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 (o que supõe, evidentemente, que noções como 'A audiência racional' ou 'A transparência ético-política das sociedades' são duvidosas).

Embora Dennett apresente as suas sugestões sob a forma de investigação do *design* de um deliberador, este deliberador pode obviamente ser um agente moral humano e de facto – embora Dennett rejeite explicitamente a aproximação<sup>125</sup> – muitas das estratégias para um certo controlo da deliberação apontadas no manual de primeiros socorros morais aproximam-se de características propostas nas éticas da virtude. Como se sabe, as éticas da virtudes 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

---

<sup>124</sup> DENNETT 1988: 139.

<sup>125</sup> DENNETT 1988.

mais do que reflexos de características estruturais do deliberador real. Sendo esse o caso, será 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 subpessoais de deliberadores pode conduzir a propostas substantivas em filosofia moral. Por exemplo, pode-se imaginar uma situação em que além da sua utilidade estrutural os interruptores de considerações tivessem uma motivação moral. Por outro lado, a opção pelo abandono do ideal da maximização racional pode traduzir-se na proposta de uma concepção não racionalista de moralidade. De facto, o que seria um agente moral totalmente racional? É razoável pensar que um tal ser não gozaria sequer a vida própria nem gozaria da existência de outros agentes<sup>126</sup>. Uma última ideia moral, substantiva, possível seria a seguinte: se existe uma pluralidade de objectos de cuidado, um cuidado não maximizante é uma estratégia racionalmente sã, acompanhado da admissão de que comparações de valores plurais são intratáveis.

## 6. Problemas de fundo.

O presente artigo teve como objectivo geral identificar e examinar algumas consequências da Teoria dos Sistemas Intencionais e do Modelo dos Esboços Múltiplos nas noções de pessoa e de acção e portanto na filosofia moral. Para isso foram escolhidos três pontos focais de análise: liberdade e agência, unidade e identidade pessoal, deliberação e escolha. Do ponto de vista subpessoal e de acordo com caracterizações do âmbito do MEM caracterizou-se os 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 relativos à constituição da identidade pessoal.

Em função desta dupla aproximação procurou-se explicar porquê, em que circunstâncias e até que ponto determinados eventos constituem acções de agentes autoconscientes, sendo só então possível falar propriamente de pessoas e também em que medida a autoconsciência como condição máxima da pessoalidade supõe uma auto-avaliação racional.

A análise dennetiana tem no entanto várias imperfeições e pontos cegos. Por exemplo, fez-se notar desde logo, a partir de C. Rovane, que faltava algo à análise das condições da pessoalidade a partir da TSI, algo que permitisse 'individualizar' a aplicação dessas condições, i. e., conceber a sua aplicação a uma pessoa individual. Nesse sentido e de acordo com Rovane<sup>127</sup> além da dimensão normativa das ideias de pessoa e acção

---

<sup>126</sup> Por exemplo em termos de amor ou amizade. Estas duas críticas aplicam-se, repare-se, quer à deontologia quer ao utilitarismo precisamente tanto quanto estas são éticas da maximização racional.

<sup>127</sup> ROVANE 1994.

Dennett deveria considerar, dentro do espírito da sua própria proposta acerca de condições de pessoalidade, uma dimensão normativa da identidade pessoal ligada ao exercício da racionalidade pelo próprio sistema cognitivo em causa<sup>128</sup> acerca da sua própria vida mental. A esta situação chamou-se a dimensão normativa da identidade pessoal, traduzida num compromisso com a vida racional própria (o qual envolve resolução de conflitos entre crenças, aceitação das consequências das crenças próprias, ordenação de preferências, etc.), compromisso esse que precisamente unifica a vida racional de um agente consciente como *própria*. São essas características estruturais, ainda cognitivas – ou uma tendência para elas, mesmo que nunca sejam perfeitamente alcançadas – que permitem um ponto de vista racional, uma pessoa individual enquanto vida mental consciente e racional *pessoal*. Noutras palavras, a existência de uma pessoa supõe o compromisso ‘a partir de dentro’, na interpretação de si, da vida mental própria, com normas de racionalidade (de acordo com Rovane um compromisso com a formulação de juízos acompanhados de uma cláusula ‘uma vez considerados todos os factores’ (*all things considered judgments*), juízos esses que guiarão a acção da pessoa). Só assim é possível definir aquilo em que consiste pensar a partir de um ponto de vista. Sem uma tal caracterização não se poderá falar justificadamente de racionalidade reflexiva e de identidade pessoal. Na auto-avaliação trata-se portanto ao mesmo tempo de responsabilidade por uma ‘personalidade’, pela criação e interpretação daquilo que se é, e de responsabilidade pela racionalidade da vida mental consciente própria. De um ponto de vista estrutural estrito é certo que as características mencionadas (a coerência total das crenças, a ordenação coerente das preferências) são impossíveis. No entanto esse é o caso de todas as condições de pessoalidade. Propõe-se portanto, aceitando a crítica de C. Rovane, que a caracterização da pessoalidade do mental não é apenas uma questão relativa às continuidades física e/ou fenomenológica mencionadas no início do capítulo. A caracterização ficaria incompleta sem aquilo a que se poderia chamar um compromisso com a racionalidade da vida mental própria importante para a identidade pessoal<sup>129</sup>. Este é um factor de instituição da identidade pessoal que está para além de, não se identifica com, as continuidades usualmente criteriosais da identidade pessoal. Estas podem estar ‘no lugar’, por hipótese, sem que no entanto exista identidade pessoal.

Repare-se que este complemento à noção de identidade pessoal significa que a sexta condição de pessoalidade, a autoconsciência, envolve um outro uso da racionalidade sobre si próprio que não tem a ver apenas com a auto-avaliação mas com o apoderar-se dos processos de pensamento de uma vida mental consciente. Em ambos os casos se trata de usar a EI sobre si, de se visar sob a perspectiva da EI de modo a controlar

---

<sup>128</sup> De acordo com C. Rovane, “Wherever there exists a commitment to satisfying the normative ideal of overall rational unity within a rational point of view, there too exists an individual person” (ROVANE 1994:354).

<sup>129</sup> Dennett não apenas aceita as sugestões de Carol Rovane como considera que ela aperfeiçoa as suas ideias (DENNETT 1994).

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 autoconsciência não é simplesmente auto-apercebimento (*self-awareness*). 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. Rovane propõe mesmo que se deve considerar – apesar do gradualismo e do naturalismo da teoria dennetiana das pessoas e das acções, que ela subscreve – que existe uma divisão real entre as três primeiras condições de pessoalidade e as três últimas, sendo que as entidades que satisfaçam as condições 4 e 5 já são pessoas no sentido envolvido no compromisso de racionalidade. Em suma, segundo Rovane, as condições 4 e 5 não poderiam ser satisfeitas sem que a condição 6 estivesse 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<sup>130</sup>. Obtém-se deste modo uma concepção qualificada daquilo que se entende por reciprocidade que afasta a possibilidade de existirem intenções comunicacionais griceanas – que fazem necessariamente apelo ao ponto de vista da audiência – sem autoconsciência. 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. Rovane aponta correctamente uma deficiência no pensamento de Dennett acerca da pessoalidade; sugere-se aqui a adopção das correcções propostas como última palavra acerca da pessoalidade dentro da TSI.

Sublinhe-se de novo e finalmente que tudo aquilo que foi dito supõe que apenas é possível dizer o que é uma pessoa ou uma acção com apoio numa teoria normativa do conteúdo e perante sistemas cognitivos com uma estrutura determinada. Aquilo que é assim descrito de forma mentalista é, do ponto de vista cognitivo, subpessoal, uma mecânica do controlo e da (auto) representação de uma unidade global de agentes especialistas, susceptível de abalo por factores completamente extrínsecos a toda a ordem normativa. O verdadeiro desafio colocado por esta convergência imperfeita da teoria do conteúdo e da teoria da consciência é a possibilidade de pessoas e acções não serem absolutamente determinadas, se nada no mundo descrito de forma mentalista o é. Se a descrição intencional à qual as razões pertencem é sempre aproximativa,

---

<sup>130</sup> Com o seu critério ético de pessoalidade Rovane pretende corrigir (e Dennett aceita a correcção) um preconceito nas formulações de *Conditions of Personhood* relativamente a certos ideais kantianos rawlsianos, preconceito que poderia ter a consequência indesejável de negar a pessoalidade aos humanos que demonstrassem más performances relativamente a tais ideais, por exemplo por subscreverem outras teorias morais que não a kantiana, por exemplo uma teoria consequencialista. Daí que o critério ético de pessoalidade de Rovane seja formulado de modo a permitir divergência de opiniões morais.

poderá dar-se o caso de não haver resolução completa de movimentos em acções, de causação em razões, de corpos humanos ao longo do tempo em pessoas.

Note-se para concluir que nesta caracterização radicalmente anti-essencialista de pessoa e acção feita partir de Dennett, a autoconsciência é aparentemente caracterizável como desprovida de conteúdo sentido ou sentimento de si. Este é um dos maiores defeitos da abordagem apresentada, um defeito que não pode ser ultrapassado a partir do pensamento do autor aqui considerado. Idealmente, a análise apresentada a partir de Dennett permitirá isolar certas insuficiências desse mesmo ponto de vista.

## Bibliografia

BECHTEL, W. & GRAHAM, G., 1998, *A Companion to Cognitive Science*, Oxford, Blackwell.

DAMÁSIO, A., 1994, *O Erro de Descartes*, Lisboa, Europa-América.

DAMASIO, A., 1999, *The Feeling of What Happens*, New York, Harcourt Brace.

DANTO, A., 1986, *Review of Elbow Room*, *Mind*, 95.

DENNETT, D., 1986 (1.ª ed. 1969), *Content and Consciousness*, London, Routledge.

DENNETT, D., 1981 (1ª ed. 1978), *Brainstorms*, Cambridge MA, MIT Press.

DENNETT, D., 1981a, "Mechanism and Responsibility", in DENNETT 1981.

DENNETT, D., 1981b, "Conditions of Personhood", in DENNETT 1981.

DENNETT, D., 1981c, "How to Change Your Mind", in DENNETT 1981.

DENNETT, D., 1981d, "The abilities of men and machines", in DENNETT 1981.

DENNETT, D., 1978e, "On Giving Libertarians What They Say They Want", in DENNETT 1981.

DENNETT, D., 1983, *Elbow Room*, Cambridge MA, MIT Press.

DENNETT, D., 1987, *The Intentional Stance*, Cambridge MA, MIT Press.

DENNETT, D., 1988, "The Moral First Aid Manual", in S. McMurrin ed, *Tanner Lectures on Human Values*, III, Salt Lake City, University of Utah Press.

DENNETT, D., 1989, "The Origin of Selves", *Cogito*, 3, 163-173.

DENNETT, D., 1991, "Real Patterns", in *Journal of Philosophy*, 88 (1), 27-51.

DENNETT, D., 1991, *Consciousness Explained*, Boston, Little Brown.

DENNETT, D., 1992, "The Self as a Center of Narrative Gravity", in F. Kessel et al., (eds), *Self and Consciousness*, Hillsdale, NJ, Erlbaum.

DENNETT, D., 1994, "Get Real", *Philosophical Topics*, 22, 1-2.

DENNETT, D., 1995, *Darwin's Dangerous Idea* (New York, Touchstone).

DENNETT, D., 1996, *Kinds of Minds*, New York, Basic Books.

DENNETT, D. & HOFSTADTER, D., 1981, *The Mind's I*, New York, Bantam Books.

DENNETT, D. & HUMPHREY, N., 1989, "Speaking for Ourselves", *Raritan*, 9 (1), 68-98.

FRANKFURT, H., 1971, "Freedom of the Will and the Concept of a Person", in *Journal of Philosophy*, 68, 5-20.

FODOR, J., 1975, *The Language of Thought*, Cambridge MA, Harvard University Press.

GRAHAM, G. & STEPHENS, G.H., 1994, *Philosophical Psychopathology*, Cambridge MA, MIT Press.

GRICE, H.P., 1957, "Meaning", *Philosophical Review* 56, 377-388.

HOFSTADTER, D., 1979, *Gödel, Escher, Bach*, London, Penguin.

- HUME, D., 1985, [1739], *A Treatise of Human Nature*, London, Penguin.
- JAYNES, J., 1976, *The Origins of Consciousness in the Breakdown of the Bicameral Mind*, Boston, Houghton Mifflin.
- LOCKE, J., 1999 [1690], *Ensaio sobre o Entendimento Humano*, Lisboa, Fundação Calouste Gulbenkian.
- MIGUENS, S., 2001, *Uma teoria fiscalista do conteúdo e da consciência*, Dissertação de Doutoramento, Faculdade de Letras da Universidade do Porto.
- MINSKY, M., 1985, *The Society of Mind*, New York, Simon and Schuster.
- NAGEL, T., 1979, *Mortal Questions*, Cambridge, CUP.
- NAGEL, T., 1979a, "Moral Luck," in NAGEL 1979.
- NAGEL, T., 1986, *The View From Nowhere*, Oxford, OUP.
- NOZICK, R., 1981, *Philosophical Explanations*, Cambridge MA, Harvard University Press.
- PARFIT, D., 1984, *Reasons and Persons*, Oxford, Clarendon Press.
- PENROSE, R., 1989, *The Emperor's New Mind*, Oxford, OUP.
- PENROSE, R., 1994, *Shadows of the Mind*, Oxford, OUP.
- PINKER, S., 1994, *The Language Instinct*, London, Penguin.
- RAWLS, J., 1971, *A Theory of Justice*, Cambridge MA, MIT Press.
- RORTY, A. (ed.), 1976, *The Identities of Persons*, Berkeley, University of California Press.
- ROVANE, C., 1994, "The Personal Stance," in *Philosophical Topics*, 22, 1-2, 351-396.
- SOUSA, R. de, "How to Give a Piece of Your Mind," in *Review of Metaphysics*, September 1971.
- SCHICK, F., 1997, *Decision Making*, Cambridge, CUP.
- STICH, S., 1993, *The Fragmentation of Reason*, Cambridge MA, MIT Press.
- TAYLOR, C., 1976, "Responsibility for self," in RORTY 1976.
- WILLIAMS, B., 1973, *Problems of the Self*, Cambridge, CUP.
- WILLIAMS, B., 1981, *Moral Luck*, Cambridge, CUP.
- WILLIAMS, B., 1985, *Ethics and the Limits of Philosophy*, London, Fontana Press.
- YATES, J.F. & ESTIN, P.A., 1998, "Decision Making," in BECHTEL, W. & GRAHAM, G., 1998.