

O problema do auto-conhecimento (self-knowledge)(1)

Sofia Miguens

Resumo: Neste artigo procuro formular os princípios de um modelo para a investigação do auto-conhecimento em termos de entendimento linguístico que ultrapasse as limitações dos modelos cartesiano (observacional) e wittgenteiniano (expressivo). Subjacente a essa proposta está a ideia segundo a qual o problema do auto-conhecimento não deve ser tratado independentemente de questões gerais acerca da natureza do pensamento e da linguagem. Ao longo do artigo procuro mostrar que embora o problema do auto-conhecimento seja formulável como um problema epistemológico acerca do que podemos saber e da forma como adquirimos um certo tipo de conhecimento, parte do seu interesse reside no facto de intersectar questões com fortes implicações cognitivas tais como a introspecção, a consciência, a autoridade de primeira pessoa, a agência e o auto-engano.

1. Por que é que o problema é um problema

Por que é que o auto-conhecimento há-de ser sequer um problema filosófico? De uma perspectiva cartesiana, parece ser mesmo o mais seguro que temos. Enquanto seres mentais, pensantes, temos conhecimento directo e íntimo dos nossos pensamentos, desejos, emoções, percepções, sentimentos. Mais ninguém tem este acesso directo que eu tenho a esta mente, a minha mente, e por isso, à primeira vista, nada existe de problemático no meu conhecimento da minha mente. O problema (do conhecimento da mente) pôr-se-ia antes em relação às outras mentes (como é que eu sei que elas sequer existem?), não em relação à minha. Mesmo sendo, todos nós, susceptíveis de formas várias de auto-engano (2), e, porventura, de estados mentais inconscientes, ainda assim a imediatidade e a autoridade de primeira pessoa parecem características inegáveis da forma como a nossa mente - a de cada um - se nos dá. Nós não temos que esperar para observar o nosso próprio comportamento, ou perguntar aos outros, para sabermos o que pensamos e sentimos. Parece-nos que nós estamos sempre lá quando a nossa mente está lá. Mais ninguém nos é tão constantemente presente. Não conseguimos imaginar o que seria alguém chegar ao pé de nós e afirmar 'Tu pensas que pensas o que pensas, mas eu sei que de facto não pensas (o que pensas que pensas)'. Por exemplo, eu penso 'Penso que está a chover', ou 'Penso que estou a ver uma parede amarela', ou 'Sei que desejo possuir aquele objecto', ou 'Estou enjoada'. E vem alguém que diz 'Não pensas nada disso, apenas pensas que pensas'. Isso não pode ser. As nossas pretensões não podem, aí, ser superadas ou vencidas - nós é que sabemos, somos nós que temos acesso privilegiado à nossa própria mente e autoridade de primeira pessoa quanto ao que pensamos e sentimos (3). Mesmo que se venha a provar que estamos, por exemplo, errados ou desadequados, no que pensamos e sentimos, e que não somos incorrigíveis, ainda assim conhecemo-nos melhor a nós próprios do que conhecemos os outros ou eles nos conhecem a nós. De qualquer forma, o ponto do cogito cartesiano (4), tal como o próprio Descartes o formulou, nunca foi a infalibilidade quanto a conteúdos conhecidos mas sim a certeza (podemos vir a verificar que os nossos pensamentos são falsos mas temos a certeza de que são o que são). É essa certeza acerca do que pensamos, desejamos, sentimos, que a ideia de autoridade de primeira pessoa procura capturar. O que me interessa aqui sublinhar pode ser posto em relevo ainda de uma outra maneira. De um ponto de vista céptico, o problema do auto-conhecimento, ao contrário do problema das outras mentes e do problema do mundo real exterior, não parece um

problema: não temos razão para perguntar 'como sabemos?' (o que pensamos, desejamos, sentimos) porque pura e simplesmente sabemos. E sabemos porque estamos em posição para saber directamente, e mais ninguém está. O núcleo do problema do auto-conhecimento é compreender que privilégio é este que temos relativamente à nossa própria vida mental.

Curiosamente, o problema das outras mentes pode hoje parecer-nos, de certa forma menos bicudo do que o problema do auto-conhecimento. Parte da alteração de perspectiva que se produziu tem a ver com o facto de que conhecer as outras mentes, i.e. atribuir mentalidade a outros seres no mundo, parece ser algo que estamos cognitivamente preparados para fazer, tanto quanto somos agentes humanos normais. É disso que se trata nomeadamente na psicologia cognitiva e do desenvolvimento sob o título de *theory of mind* - um funcionamento cognitivo, resultante de desenvolvimento psicológico, obviamente não controlado pelo próprio agente, com timings específicos (5). Mesmo se as crianças parecem ter um entendimento dos aspectos mentais das coisas desde muito cedo, o que é facto é que há um desenvolvimento dessa capacidade de 'ler mentalidade' fora de si. Esse desenvolvimento pode ser capturado considerando, nomeadamente, a forma como elas são capazes de conceber a relação de crenças de outros agentes com a realidade. Um teste decisivo, como é bem conhecido, é o teste das crenças falsas - é ou não concebível, para uma criança, que um outro agente, uma outra pessoa, pode ter uma representação falsa da realidade, uma crença falsa? O que é necessário que esteja no lugar, cognitivamente, para ser possível atribuir uma crença falsa a alguém, fazendo previsões correctas acerca do comportamento dessa pessoa, assumindo que o comportamento da outra pessoa é guiado pela crença falsa? O mais seguro aqui são os factos da psicologia do desenvolvimento: apenas a partir de uma certa idade - entre os três e os cinco anos de idade - as crianças são em geral capazes de 'passar' os testes de crenças falsas, i.e., são capazes de resolver problemas cuja resolução envolve a consideração de que outro agente tem uma crença falsa acerca da realidade (6). O que isso envolve em termos de teoria da mente é outra questão (7).

A razão pela qual evoquei a investigação empírica das teorias da mentalidade num contexto em que pretendo abordar o auto-conhecimento foi apenas esta: acreditar que outras mentes estão lá, considerá-las no controle do comportamento próprio, não resulta de nenhuma resposta filosófica ao problema filosófico céptico acerca das outras mentes. A existência de outras mentes não é a conclusão de um argumento que formulamos, é antes resultado de uma habilidade cognitiva, sujeita a desenvolvimento, que nos propicia perícia em determinadas situações no mundo. E isto ocorre de uma forma que é opaca ao próprio agente mas que pura e simplesmente funciona (8). Será que algo de análogo acontece com o conhecimento da mente própria?

De alguma forma o auto-conhecimento parece ainda mais misterioso do que o conhecimento das outras mentes, sobretudo se considerarmos a grande difusão do externalismo nas teorias da mente e da linguagem, e se tendermos a aceitá-lo. De facto, o externalismo dá-nos boas razões para pensar que o que dá conteúdo a pensamentos e linguagem não está 'no interior', i.e. e não sobrevém a propriedades intrínsecas de pensadores (9), mas sim no exterior do agente, no mundo. Se o externalismo for verdadeiro, a imediatidade e autoridade que marcam o auto-conhecimento parecem ainda mais inexplicáveis. Temos, em suma, no auto-conhecimento, algo de muito imediato e seguro mas muito difícil de explicar (10). Admitindo a imediatidade e a autoridade como características do auto-conhecimento, o que as explica? Por que é que nos aparecemos mentalmente, a nós próprios, dessa forma? A tentativa de obedecer ao imperativo de Delfos ('Conhece-te a ti mesmo') é tão ou mais antiga que a filosofia, e no entanto o problema do auto-conhecimento é uma questão perfeitamente em aberto.

2. Duas soluções para o problema: solução observacional cartesiana,

solução expressiva wittgensteiniana

2.1 Solução cartesiana

O cartesianismo é a primeira solução à mão (podemos chamar-lhe uma solução observacional) para o problema do auto-conhecimento. Porque é que o cartesianismo é uma solução, e uma solução atraente, para o problema do auto-conhecimento? Porque nos faz olhar para uma mente como um interior totalmente privado e subjectivo, acessível unicamente ao próprio e inacessível a outrem, constituindo o auto-conhecimento a observação desse interior (a introspecção, um 'mecanismo de olhar para o interior', seria o método utilizado). John McDowell, discutindo com Crispin Wright a herança wittgensteiniana na forma de rejeitar esta concepção cartesiana de mente (11), aponta ainda uma outra razão para a capacidade de atracção da resposta cartesiana ao problema do auto-conhecimento: a resposta cartesiana é atraente devida a uma transposição do Mito do Dado para o interior. 'Olhamos' introspectivamente e 'há lá coisas mentais', vemos o que já está lá, estruturado, antes de nos apoderarmos do que quer que seja. Pensar no auto-conhecimento como observação não mediada, ao contrário da percepção sensorial (12), não-inferencial e metafisicamente directa parece-nos, apesar de tudo, à primeira vista, uma boa ideia. Mas os problemas são muitos e insuperáveis.

Quais são então os limites do modelo cartesiano? Desde logo se não aceitamos o compromisso ontológico dualista do cartesianismo a pretensão fica mais frágil (ou mesmo informulável) - o que será esse interior privado e subjectivo, acessível unicamente ao próprio, que conhecemos imediata e transparentemente se não é uma alma imaterial, uma substância pensante ontologicamente distinta da matéria? O que é isso que está, supostamente, a ser observado directamente? E quem é o observador? Que agência, que iniciação, que meios de conhecimento são esses de que dispõe? Falar, em abstracto, de introspecção, é, obviamente, uma metáfora visual, que dá, além do mais, por garantida, sem mais justificação, algo como uma separação entre espaço interior e mundo exterior, sem lidar com as questões metafísicas e epistemológicas que tal separação envolve. E a distinção entre o sujeito e tudo o resto arrisca-se a fazer do auto-conhecimento não apenas um tipo especial de conhecimento mas mesmo o único conhecimento possível.

Para além deste esboço de pontos gerais a criticar, resumo em seguida algumas objecções comuns ao modelo observacional cartesiano. A primeira objecção diz respeito às pretensões epistémicas fortes, especialmente atraentes para os defensores das teorias fundacionalistas da justificação epistémica, que precisam de crenças auto-justificadas para servir de fundamento último às outras crenças. As crenças de auto-conhecimento seriam por excelência crenças auto-justificadas. É no entanto muito pouco claro em que consiste o seu estatuto epistémico de excepção (a certeza, a transparência, a infalibilidade). Para dizer mais claramente: será que essas características são sequer possíveis em algum tipo de conhecimento? Poder-se-á falar, de resto, de 'conhecimento', numa situação em que parece não existir a possibilidade de erro (recorde-se o argumento wittgensteiniano da Linguagem Privada, nas Investigações)? Dever-se-á falar de conhecimento quando não há qualquer esforço cognitivo? Uma segunda objecção é a seguinte: não é claro em que consiste a 'observação não mediada' que assegura a imediatidade metafísica. Um terceiro problema é o seguinte: como pode o modelo dar conta dos conteúdos relacionadamente determinados de muitos dos pensamentos interiormente observados (por exemplo 'Eu sei que vejo uma árvore verde' sendo que a dita árvore está supostamente lá fora no mundo), se aquilo que é observado internamente são apenas traços intrínsecos dos estados mentais? (13) Uma última observação: percepções de outras pessoas podem constituir uma fonte valiosa de auto-conhecimento (por exemplo, alguém pode convencer-nos da falta de honorabilidade das nossas intenções, ou faz-nos reconhecer os motivos que lhes subjazem) - ora, um modelo cartesiano não tem como incorporar tais 'contribuições externas' para o auto-conhecimento.

De qualquer modo, a ideia nuclear do modelo é que há algum traço epistémico diferenciador do auto-conhecimento - parte da sua atracção tem certamente a ver com isso.

A ideia de uma estrita observação não mediada, de um modelo cartesiano puro como solução para o problema do auto-conhecimento, é, como é óbvio, mais uma caricatura do que outra coisa. A questão mais séria é saber que privilégio é este, que admitimos que existe, de imediatidade e autoridade. Como podemos explicá-lo sem pressupormos um reino privado, subjectivo, delimitado, de pensamento, pessoalidade e agência? Alternativas mais próximas, qualificadas e portanto bem mais defensáveis, para compreender a 'estrutura do privilégio' são por exemplo os modelos, avançados em filosofia da mente no âmbito de discussões acerca da consciência, que configuram a introspecção como um sentido interno, semelhante à percepção. Tais modelos são frequentemente construídos, mesmo por filósofos, fazendo apelo a considerações sub-pessoais, i.e. relativas a mecanismos cognitivos. A ideia é que observações estruturais e funcionais sobre o auto-acesso e a auto-monitorização que ocorreriam num sistema cognitivo e produziriam estados mentais de segunda ordem (i.e. estados mentais que são acerca de estados mentais de primeira ordem do mesmo agente) são a pista de que precisamos aqui. Muitas vezes o assunto é abordado por filósofos de forma semelhante ao que faria um cientista empírico, por exemplo um psicólogo cognitivo (com a diferença de o filósofo avançar o modelo antes, ou independentemente, da recolha dos dados). Como já afirmei, normalmente é a consciência (e a auto-consciência), mais do que propriamente o auto-conhecimento, aquilo que é visado nestas teorias. Mas, precisamente, as questões da consciência e do auto-conhecimento cruzam-se inevitavelmente: constitui uma pretensão legítima e merecedora de consideração no âmbito das discussões acerca da natureza da consciência a ideia segundo a qual a auto-consciência, num sistema cognitivo, deve ser explicada como auto-conhecimento, resultante de inspecção por um sentido interno (14). Exemplos salientes de propostas deste tipo na filosofia da mente contemporânea são as teorias da consciência avançadas por autores como D. Armstrong, W. Lycan, D. Rosenthal e D. Dennett (15). Podemos chamar-lhes em geral higher-order theories (vou traduzir a expressão por 'teorias de segunda ordem'). Estas são normalmente divididas em 'teorias HOP' (teorias da higher-order perception, i.e. aquelas que concebem os estados mentais de segunda ordem, em função dos quais se explicaria a consciência, como percepções) e 'teorias HOT' (teorias do higher-order thought, aquelas que vêem os estados mentais de segunda ordem, em função dos quais se explicaria a consciência, como crenças).

Como disse, o problema visado por Armstrong e Lycan, e que conduz ambos a propor um sentido interno (16), é o problema da consciência, e não directamente o problema do auto-conhecimento. De acordo com o modelo de Armstrong, a chave para compreender a consciência numa criatura é pensar na introspecção como um processo de auto-monitorização, no cérebro, a ser considerado em termos causais, cujo resultado do processo são estados mentais perceptivos que são acerca de outros estados mentais. Também aqui o auto-conhecimento pode ser visto como não inferencial, embora exista uma diferença entre estado / processo que monitoriza e estado / processo monitorizado, devendo a relação entre estes ser concebida em termos causais. A criatura na qual estes estados de crença acerca dos seus próprios estados mentais ocorrem é uma criatura consciente. Temos assim que a consciência é um apercebimento perceptivo de segunda ordem dos estados actuais da mente própria.

Também W. Lycan pretende explicar a consciência através do apelo à introspecção. De acordo com o seu modelo, a consciência é uma representação perceptiva dos estados mentais próprios sujeitos a atenção, representação resultante da operação de um sentido interno, de um mecanismo de monitorização.

A ideia de modelos do sentido interno como os de Armstrong e Lycan é, claramente, substituir a observação metaforicamente descrita do modelo cartesiano puro pela introspecção concebida como um

mecanismo cognitivo dirigido para o interior do agente, mecanismo do qual é legítimo investigar o funcionamento sub-pessoal. A questão do auto-conhecimento passa assim a ser tratada como não apenas epistémica mas também empírica. De qualquer forma, o âmbito da investigação da natureza do auto-conhecimento restringe-se, nestes modelos, ao indivíduo, ao agente isolado.

Os modelos de David Rosenthal e Daniel Dennett são usualmente classificados como modelos HOT. De acordo com o modelo de Rosenthal, um estado mental é um estado mental consciente se for acompanhado por um pensamento de ordem superior (higher order thought, ou HOT) acerca dele. As ideias básicas da teoria da consciência de Dennett (desenvolvida em *Brainstorms*, 1978, e em *Consciousness Explained*, 1992)(17) recaem (não apenas mas) também numa explicação da consciência através de crenças acerca dos estados mentais próprios, numa criatura. Dennett parte do princípio de que o apercebimento (awareness) de um sistema cognitivo é mais alargado do que a 'experiência consciente' (consciousness). A existência deste apercebimento é simplesmente identificável com disponibilidade na memória do sistema de um saber-que. Para compreender a consciência, é preciso compreender algo mais do que essa disponibilidade na memória do sistema de um saber-que: é preciso compreender a forma como a expressão ou 'publicação' de qualquer conteúdo disponível envolve um acesso sub-pessoal entre memória e mecanismos cognitivos para a reportabilidade. Sendo o conteúdo do apercebimento tudo aquilo que é registado na memória, esse conteúdo pode degradar-se ou sofrer interferências antes de ser recuperado para acesso público, de ser expresso. Ser expresso, ser objecto de acesso 'público', pode significar simplesmente acesso de si para si no sistema. Por contraste com os acessos sub-pessoais (mecanismos de memória e reportabilidade), o acesso pessoal é já público - é reportabilidade de si para si, acontecendo evidentemente sem qualquer observação interna. É este acesso pessoal-público que Dennett identifica com a consciência. A autoridade introspectiva é, neste quadro, evidentemente, apenas um sentimento (não uma garantia), que resulta da forma como, ao nível sub-pessoal, se dão as intenções semânticas. A situação, repito, tem como base uma ligação entre controlo e reportabilidade, e é explicada em termos de mecanismos cognitivos sub-pessoais.

Afirmo que neste quadro o apercebimento tem um âmbito muito maior do que a consciência (considerada como reportabilidade dos conteúdos para um Eu, i.e. acesso pessoal). Esse âmbito deverá ser determinado pela investigação empírica - isto significa que a 'autoridade' subjectiva, a consciência-de-acordo-com-Dennett, assinala apenas um certo tipo de acesso, e não a presença de um género natural que seria 'a consciência'. Os géneros naturais são da competência da investigação empírica, e nada garante a priori que a consciência seja um. Desde logo, não há autoridade subjectiva quanto a géneros naturais, nomeadamente quanto ao que seria 'a consciência'. A parte da teoria do mental que é teoria da consciência é, assim, uma teoria do apercebimento-memória, da forma como os conteúdos estão presentes, se dão unificadamente no acesso de si a si de um sistema cognitivo. Relativamente aos mecanismos sub-pessoais, os modelos de agentes (como os de *Brainstorms* e *Consciousness Explained*) postulam vários homúnculos e assim disseminam, até a dispensar, a observação.

A célebre crítica dennettiana aos qualia que acampanha o modelo de *Consciousness Explained* não é senão uma consequência desta abordagem: Dennett não aceita que algo possa ser 'objectivamente subjectivo', i.e. possa fazer parte da experiência do sujeito, mesmo que este não o soubesse-que (seria melhor dizer: ser uma entidade de experiência sem ser apercebida por algum sujeito, sendo que o apercebimento aqui em causa é uma crença). A ideia, defendida por Dennett, de que a consciência de um estímulo na ausência de crença nessa consciência é impossível por princípio é uma ideia verificacionista. O verificacionismo é ainda especificado pela ideia de que seria impossível por princípio ao sujeito decidir, relativamente à sua consciência de x, entre aparência e realidade. A pertinência deste verificacionismo em primeira pessoa relativamente à experiência interior é justificada pelo facto de o em-si da experiência, a existência

de experiência, só poder ser acedido como um para-mim e pelo facto de só a própria entidade que se sente ser poder saber e exprimir como é para si ser. Daí a consciência ou autoridade do sujeito (a ambiguidade, inevitável se lermos a posição de Dennett de forma suficientemente sofisticada, é interessante).

A partir de uma perspectiva como a de Dennett as relações entre consciência, introspecção e auto-conhecimento tornam-se relativamente claras. No primeiro livro de Dennett, *Content and Consciousness* (1969) a análise da introspecção constitui mesmo o primeiro passo no sentido de mostrar que através da atenção concedida ao nível sub-pessoal se torna claro ao teórico da mente que a consciência não é um traço da mente mas vários.

Dennett baseia-se nas ideias funcionalistas de H. Putnam para esclarecer o estatuto do acesso introspectivo. A questão do acesso introspectivo é absolutamente central na filosofia de Dennett também porque os relatos introspectivos (relatos linguísticos, portanto) são a matéria-prima, os dados brutos, da teoria da consciência. De acordo com Dennett, o auto-acesso deve ser analisado levando em conta uma distinção que se impõe entre dois tipos de apercebimento (*awareness*). O primeiro tipo de apercebimento nomeia a possibilidade de expressão linguística por um sistema, o segundo nomeia o controle por um sistema das suas actividades. A estes sentidos do apercebimento chama apercebimento-1 e apercebimento-2. O ponto de Dennett é que não existe um terceiro apercebimento, que seria a consciência (fenomenal), a qual teria sido deixada de fora nas explicações do auto-acesso.

O que resta compreender para compreender a consciência uma vez 'eliminada' a suposta consciência fenomenal é o sentimento de infalibilidade relativamente ao conteúdo dos nossos pensamentos, que tenho vindo a considerar marca do auto-conhecimento. Dennett afirma aí que autores como Wittgenstein, Ryle e Anscombe deram um primeiro passo acertado para a avaliação dessa nossa condição aparentemente cartesiana ao considerarem que se as asserções de experiência interior são infalíveis é porque elas não são relatos factuais. Não são relatos factuais porque qualquer relato factual pode estar errado, e as asserções de experiência interior não podem estar erradas. O atestar introspectivo de estados mentais não é assim um conhecimento (ao qual corresponderiam relatos factuais) mas uma condição prática, uma capacidade, um comportamento. Uma consequência natural deste tipo de posição é a negação da utilidade epistemológica dos relatos de experiência interior - e isso é precisamente o que alguém como Dennett se recusa a aceitar. Embora os relatos do auto-acesso não possam ser tomados como uma expressão da 'intuição-clara-e-distinta-daquilo-que-é', eles devem ter alguma utilidade epistemológica. Como lida Dennett com esta situação? A diferença da sua posição relativamente às abordagens características dos três autores referidos é precisamente determinada pela atenção prestada ao nível sub-pessoal. De facto, e como os três autores defendem, ao nível pessoal as explicações páram. Eles estão no entanto enganados, de acordo com Dennett, quando pensam que o problema da incorrigibilidade da introspecção tem solução ao nível pessoal. Dennett sugere que é ao nível sub-pessoal que a incorrigibilidade da introspecção deve ser investigada. E é nesse sentido que utiliza a distinção putmaniana entre estados físicos e estados lógicos.

Um estado lógico de um sistema é, por definição, aquilo que é em virtude das suas relações com inputs e outputs e com outros estados e não em virtudes de quaisquer características físicas específicas. Assim sendo, a situação na qual um sistema se assegura que está num estado lógico só pode ser concebida como a situação em que o sistema se comporta como se estivesse nesse estado lógico. E se o sistema se comporta como se estivesse nesse estado lógico é porque o sistema está nesse estado lógico, já que estar num estado lógico consiste exactamente em 'comportar-se' de uma dada maneira. Então, quando um sistema relata que está num estado lógico, esse relato sai directamente do estar nesse estado lógico, não implica computação ou inspecção de indícios para chegar à conclusão de que está nesse

estado lógico. Se alguém perguntar como é que o sistema sabe que está num estado lógico, a única resposta possível é, como diz Dennett na sequência da exposição de Putnam em *Minds and Machines*, 'degenerada': o sistema sabe que está num estado lógico determinado estando nesse estado lógico. O contraste que Dennett quer realçar é o seguinte: os humanos, e também as máquinas, são capazes de alguma monitorização dos seus estados internos, nomeadamente dos seus estados físicos internos. Quando um humano monitoriza ou sonda os seus estados físicos internos e relata as condições a que acede, faz sentido querer saber como é que ele se assegura daquilo que reporta. O humano poderá responder a uma questão com esse teor dizendo qual foi a sucessão de estados por que passou. No entanto quando o estado relatado pelo sistema é um estado lógico, i.e., funcionalmente individuado segundo Putnam, a mesma questão não faz qualquer sentido. De entre os relatos de estados sub-pessoais feitos por uma pessoa, são os relatos de estados lógicos de um sistema que são invulneráveis a erros, e não os relatos de estados físicos. É exactamente esse o caso dos relatos introspectivos humanos. A incorrigibilidade deve portanto ser explicada pelo estatuto dos estados lógicos. Como Putnam vira, certos relatos são incorrigíveis precisamente porque são relatos de estados lógicos. Os estados lógicos não podem ser erradamente identificados, simplesmente porque não têm que ser (nem podem ser) primeiramente identificados.

Numa situação em que não existe outro interior que não seja o próprio funcionamento (exactamente como o nosso cérebro), não pode haver engano acerca do que parece ser. Ela pode 'estar errada' acerca do que é o mundo e pode cometer erros 'verbais', mas não pode identificar erroneamente o output dos processamentos internos, pois o output dos processamentos internos é simplesmente o mesmo estado lógico que o input do centro de fala. Logo, ela não pode ser 'responsabilizada' por input fraudulento ou por erros de análise de input.

A identidade entre o output dos processamentos internos e o input do centro da fala é obviamente postulada, assumindo já um ponto de vista funcionalista. É no entanto essa identidade que traz, segundo Dennett, a autoridade do auto-conhecimento. Esta resulta não da ponderada apreciação dos estados mentais próprios num retiro interior, não de um 'parecer' da pessoa, mas da arquitetura cognitiva no cérebro. A situação do auto-acesso de um humano que produz um relato introspectivo não é assim nem apreciação ponderada pelo pensador dos seus estados interiores, nem descrição de eventos nervosos, mas sim expressão (do conteúdo desses eventos). Essa expressão está sujeita a erros verbais mas não pode por princípio ser uma má descrição ou uma má identificação, pois aquilo que é dito num relato introspectivo deriva directamente da circunstância de o sistema estar nesse estado. Mais: se se chamar pensamentos àquilo que é relatado por um humano (entenda-se: mesmo relatado para si próprio) a partir do auto-acesso, esses pensamentos não poderão ser identificados com o que quer que seja ao nível sub-pessoal. Não existem entidades determinadas (os pensamentos) que estariam a ser relatadas (expressas no sentido de estarem anteriormente no interior sendo depois exteriorizadas). O facto de as pessoas serem capazes de dizer o que estão a pensar não obriga a teoria da mente a um compromisso com a existência de pensamentos. Não existem sequer 'pensamentos' de modo a que se possa afirmar que 'aquilo que as pessoas exprimem ou exteriorizam são os seus pensamentos'.

A forma como Dennett pretende explicar o privilégio inerente ao auto-conhecimento é interessante mas falha. Dennett defende que o facto de as pessoas saberem e serem capazes de dizer o que estão a pensar, com autoridade, é explicável através do estatuto de estados lógicos no nível sub-pessoal, nível no qual não existem nem pessoas nem pensamentos. Vários pontos são criticáveis aqui.

Antes de mais, a tentativa de compreender a autoridade de primeira pessoa através do nível sub-pessoal, através do apelo a abordagens empíricas, supõe rejeitar que se trata de uma questão epistémica. Considerações como as de Dennett são, no máximo, apenas uma parte da resposta ao problema do auto-conhecimento. O problema do auto-

conhecimento tem aspectos epistemológicos e ontológicos (relativos a crenças, evidência, justificação, estatuto da imediatidade pensada) não apenas aspectos cognitivos (relativos a arquitectura e processos). Em segundo lugar, Dennett considera desde logo que o auto-apercebimento que interessa é, só pode ser, linguístico. A capacidade de apercebimento humano sofisticado é ligada à capacidade de fazer relatos introspectivos linguísticos. Mas o conhecimento não inferencial, imediato que nos interessa para o problema do auto-conhecimento não é necessariamente introspecção linguística sofisticada. É duvidoso que a introspecção linguística sofisticada possua um estatuto exemplar no pensamento humano, o auto-conhecimento é mais lato do que a introspecção linguística sofisticada. Dennett rebate o apercebimento humano, a consciência, sobre a capacidade de falar acerca da mente própria (i.e. de levar a cabo introspecção linguística). O apelo de Dennett à conexão directa para explicar a incorrigibilidade pressupõe assim, vê-se aqui, um cartesianismo, materialista embora.

Como se vê através destes últimos modelos a que chamei também observacionais-cartesianos, uma vez cruzando o problema do auto-conhecimento com questões relativas a consciência e introspecção, inevitavelmente a discussão do problema do auto-conhecimento converge com as discussões em torno das teorias filosóficas da percepção e da consciência. Do que foi dito até aqui, é fácil ver como podem algumas dessa teorias da consciência pecar por cartesianismo, restrição ao eu, ao agente, para pensar na natureza da mente e do pensamento, sobrevalorização da consciência. De forma a mostrar o que pode ser uma saída de tal cartesianismo, vale a pena mencionar um outro modelo, cuja diferente forma de abordar o auto-conhecimento se justifica pela rejeição das teorias da consciência de 2ª ordem, substituindo-as por uma teoria de 1ª ordem. O modelo é o de Fred Dretske: para Dretske, o auto-conhecimento, ou melhor o auto-apercebimento (self-awareness) é uma forma de conhecimento que é obtido via o apercebimento de objectos não mentais. De acordo com Dretske, o que explica o auto-conhecimento é o seguinte: quando nos apercebemos de um objecto não mental - uma flor amarela à minha frente, por exemplo - podemos inferir que estamos num estado mental que representa 'flor amarela à minha frente'. Neste modelo, note-se, o conhecimento introspectivo é inferencial, e nesse sentido 'chega mais tarde' do que outro apercebimento não consciente (daí que alguém como Dretske possa e deva perguntar, com toda a relevância, algo como: *How do i know i am not a zombie?* (18)) Estamos já a afastar-nos, por isso, de soluções 'cartesianas' para o problema do auto-conhecimento, embora o problema em análise seja ainda saber o que deve ser o caso para que exista uma auto-adscrição de mentalidade, um auto-conhecimento de algo como mental (19).

O ponto a que quero chegar e que me parece constituir o principal defeito do cartesianismo na abordagem do problema do auto-conhecimento é o seguinte. Talvez não tenha ainda sido perfeitamente claro, mas todas estas ideias se reportam a indivíduos ou mentes isoladas e algumas aos processos subpessoais de tais agentes: a marca do cartesianismo é a desconsideração da terceira pessoa, da dimensão pública, e da importância que esta tem na natureza do mental e do pensamento. Noutras palavras, muito embora nós tenhamos conhecimento directo e íntimo dos nossos pensamentos, desejos, emoções, percepções, sentimentos, e mais ninguém tenha este acesso directo que eu tenho a esta mente, isso não constitui por si razão suficiente para presumirmos que esse privilégio é explicável a partir do interior material ou imaterial do sujeito ou agente. Precisamos aqui de um ponto fregeano para introduzir uma nova dimensão na discussão: os nossos pensamentos são partilháveis, não são totalmente inacessíveis a outrem (recordo os exemplos do início deste artigo: 'Penso que está a chover', ou 'Penso que estou a ver uma parede amarela', ou 'Sei que desejo possuir aquele objecto', ou 'Estou enjoada'). Nós não somos totalmente inabordáveis a partir de fora. O facto de nos comportarmos e exprimirmos linguisticamente assegura-o. Ora, o comportamento linguístico não pode possivelmente ser compreendido a partir do agente em estado isolado, e ele é essencial para compreender o mental, e o conhecimento da mente própria, i.e. o auto-conhecimento. Em resumo: modelos cartesianos do auto-conhecimento falham porque os nossos pensamentos, cujo conhecimento directo e íntimo está em causa no problema do auto-

conhecimento, não podem ser compreendidos exclusivamente a partir da privacidade, ou do nível sub-pessoal, tendo como referência o agente isolado. Precisamos do exterior.

2.2 Solução wittgensteiniana

Wittgenstein dá-nos uma outra solução para o problema do auto-conhecimento, uma solução anti-cartesiana, que não passa tão pouco pelo apelo a condições empíricas como alguns dos últimos modelos mencionados. Alguns autores chamam-lhe uma solução 'não epistémica' (Gertler 2003). Podemos chamar-lhe uma solução expressiva. Nesta solução não presumimos que o interior privado e subjectivo que conhecemos imediata e transparentemente seja uma alma imaterial, substância pensante ontologicamente distinta da matéria, não presumimos que observamos - quem seria o observador? - nem que algo que já está lá é objecto de algum tipo de percepção. A sugestão de Wittgenstein é que devemos pensar no 'interior' (20) em termos de *Ausserungen*, *Ausdrücke*, *avowals*. Vou traduzir estes termos por 'expressões' - de qualquer forma o que se pretende capturar é algo como saber com autoridade, imediatidade, transparência e sem fundamento, sem bases. Podemos ver a situação assim: o problema do auto-conhecimento diz respeito a auto-adscrições de mentalidade. Estas não são observação de coisas mentais que estariam lá, de estados e processos mentais que o sujeito descreve olhando para dentro, mas sim performances de expressão. Performances de expressão não podem ser verdadeiras ou falsas - é essa a (não) solução wittgensteiniana para o problema do auto-conhecimento (21). A *groundlessness* (ausência de fundamento) dessas auto-adscrições de mentalidade que são as asserções de experiência interior, i.e. a forma como acreditamos aquilo que acreditamos acerca do que mentalmente somos, parece bem capturada pela ideia de expressão porque expressão não é nem observação, nem resultado de inferências baseada em razões.

Sendo a expressão a proposta nuclear do modelo wittgensteiniano, a forma mais directa de tentar capturar os limites deste é analisar o estatuto das expressões, e a utilidade que a ideia pode ter no tratamento do problema do auto-conhecimento. Já aludi atrás àquele que parece ser o principal limite do modelo: nesta forma não epistémica de tomar as auto-atribuições mentais joga-se qualquer coisa como a negação da utilidade epistemológica dos relatos de experiência interior. Esta posição está evidentemente ligada à mais geral rejeição, por Wittgenstein, de qualquer postura explicativa em filosofia.

Mas será que se deve de facto defender que as descrições que os sujeitos podem fazer dos seus próprios estados mentais são epistemologicamente inúteis? Significará isso que o interior é um 'nada', ou melhor, que não há nada de interior ou pelo menos que o interior não serve para nada quando queremos tratar questões gerais como a natureza do pensamento e da linguagem?

É importante antes de mais ter em mente que Wittgenstein não é um behaviorista, a sua teoria da mente e da linguagem não elide pura e simplesmente o 'interior' - é preferível ver Wittgenstein precisamente como um teórico do interior, que aborda a partir da assimetria entre 1ª e 3ª pessoa. Wittgenstein está, também ele, interessado em compreender a natureza da autoridade da 1ª pessoa (22). Mas como capturar, no contexto dessa assimetria, o facto de as expressões não serem - parafraseando Wittgenstein - um 'nada' epistemicamente? A solução expressiva de Wittgenstein faz-nos correr o risco de ficar sem nada para dizer acerca desta questão (23).

Retomo a ideia de expressão e alguns dos exemplos que tenho vindo a utilizar: 'Penso que está a chover', 'Penso que estou a ver uma parede amarela', 'Sei que desejo possuir aquele objecto'. Desde logo, a ideia nuclear do modelo é que as asserções de experiência interior não são relatos, com pretensões de verdade, susceptíveis de serem verdadeiros ou falsos em função de alguma observação de objectos, estados ou processos interiores, mas sim expressões. Uma solução

expressivista para o problema do auto-conhecimento abdica da ideia de que nos observamos enquanto interior mental privado: dizer que há expressões, é dizer que (só) damos por nós (já) assim. Expressar é de algum modo estar já cá fora, mesmo para nós próprios: é como se nos dessemos, desde logo, a nós próprios como coisa exterior. Mas ainda assim com um certo tipo de autoridade. O problema é saber o que tem Wittgenstein a dizer de produtivo acerca dessa autoridade. (Convém recordar que no caso específico de Wittgenstein, o argumento segundo o qual os relatos factuais podem precisamente e por definição, estar errados e os relatos introspectivos não podem estar errados, não sendo portanto os relatos introspectivos relatos factuais, foi um ponto importante para o abandono do solipsismo e do verificacionismo defendidos em escritos anteriores às Investigações Filosóficas).

Para situar melhor o âmbito da questão em discussão, convém recordar a conhecida distinção introduzida por Wittgenstein (Livro Azul e Livro Castanho) entre dois diferentes casos de uso da palavra 'eu'. Trata-se do uso como sujeito ('tento levantar o meu braço', 'vejo uma parede amarela') e o uso como objecto ('cresci seis centímetros', 'tenho o cabelo muito comprido'). Com a distinção, Wittgenstein pretende chamar a atenção para o facto de a imunidade ao erro através de identificação errónea (misidentification) ser uma característica do primeiro uso (não posso gemer de dor por erro, quando quem está a sentir dor é outro e não eu) e não do segundo (24). Podemos considerar que problema do auto-conhecimento fica assim 'dividido em dois', ou, pelo menos, é por aqui que ele se prolonga em problemas acerca de crenças acerca do eu (self-beliefs) e questões de identidade pessoal, mas o facto é que aquilo de que temos estados aqui a falar é do uso do eu como sujeito. É relativamente a esse uso que se põe a questão da natureza da autoridade e da imediatidade, e é aqui que a ideia de expressão pode constituir uma via de exploração.

De qualquer forma, e apesar de algumas teses positivas - tais como a ideia segundo a qual a autoridade a atribuir ao sujeitos que 'professam a sua mentalidade de uma determinada maneira' é primitiva, é um princípio constitutivo, e não consequência de estados mentais alguns - para muitos autores a 'solução' expressivista wittgensteiniana é pura e simplesmente um non-starter, não se consegue trabalhar a partir daí (25). Porquê? Simplesmente não resta muito espaço de manobra para desenvolver uma filosofia da linguagem e do pensamento, no sentido em que se perguntarmos por exemplo o que explica a significação das palavras usadas nas 'expressões do interior' ficaremos em afirmações tais como 'é a gramática, a gramática não é sustentada por nada, a gramática é história natural, há um jogo de linguagem que é jogado, trata-se de seguir regras, seguir regras é uma praxis' e não nos movemos daí, de uma presunção exteriorista e anti-explicativa geral. Como afirma C. Wright, trata-se de 'descartar suposições de explicabilidade' (26). Quanto ao problema do auto-conhecimento, Wittgenstein quer 'destruí-lo': parte do que vai com a proposta da noção de expressão é a sugestão de que nos desfaçamos da quase irresistível analogia entre as asserções de experiência interior como 'estou a ver uma parede amarela' e os relatos observacionais, como 'A pedra que está na minha frente está pintada de azul'. As coisas que podemos perguntar acerca de relatos observacionais como esse não as devemos perguntar acerca de asserções de experiência interior, dado o seu estatuto de expressões.

Mas será que uma teoria da significação pode ficar pelas afirmações exterioristas atrás citadas? Será que nada se passa na mente dos indivíduos que importa para questões de significação? Será que não podemos mesmo procurar uma explicação do estatuto de expressão das asserções de experiência interior? (27) Wittgenstein responde demasiado rapidamente que não a estas duas últimas perguntas, e o modelo expressivo incorre deste modo no risco de jogar pela janela as possíveis aportações epistemológicas da primeira pessoa no tratamento do problema que aqui nos interessa.

3. Para além da observação e da expressão: entender o entendimento linguístico, e nesse contexto, a autoridade de primeira pessoa.

A crítica fundamental aos dois modelos até aqui considerados pode ser formulada a partir de uma terceira perspectiva. A crítica é a seguinte: os dois modelos do auto-conhecimento até aqui considerados capturam essencialmente ou a dimensão de privacidade do pensamento (o modelo cartesiano), ou a dimensão pública deste (o modelo wittgensteiniano). Nenhum captura as duas em conjunto (28). Ora esse é precisamente o principal desafio que se coloca. Admitir isto implica, obviamente, admitir que esclarecer a natureza do pensamento constitui uma tarefa prévia a qualquer posição perante o problema do auto-conhecimento. Mas a verdade é que se não aceitamos esta tarefa prévia, não teremos como evitar as pretensões não justificadas quer do modelo cartesiano quer do modelo wittgensteiniano: no modelo cartesiano é como se já tivéssemos o direito de falar epistemologicamente acerca do interior, no modelo wittgensteiniano não chegamos a ter possibilidade de falar epistemologicamente acerca da experiência interior, ou melhor, o interior não importa quando se trata de falar epistemologicamente. São essas pretensões não justificadas que tornam os dois modelos vulneráveis.

A terceira perspectiva, que apresento em seguida, propõe uma estratégia linguística de abordagem do problema do auto-conhecimento, ou melhor, propõe uma teoria do entendimento linguístico como chave de uma investigação do auto-conhecimento. Penso que apenas uma tal estratégia nos permitirá ultrapassar as limitações dos modelos cartesiano (observacional) e wittgensteiniano (expressivo). Esta terceira perspectiva assume desde logo que o problema do auto-conhecimento não pode ser tratado independentemente de questões fundamentais acerca da natureza do pensamento e da linguagem, e é por isso formulada em diálogo com alguns dos autores contemporâneos que tratam directamente tais questões.

O autor a que me reporto a partir de agora, para formular os princípios da terceira proposta perante o problema do auto-conhecimento, é o filósofo inglês da mente e da linguagem Barry Smith. A abordagem do entendimento linguístico defendida por B. Smith é influenciada por autores como N. Chomsky, D. Davidson e J. McDowell, e procura superar as insuficiências do modelo expressivo wittgensteiniano. Tenta também evitar os defeitos de projectos que têm de alguma forma intenções semelhantes ao seu, tais como a interpretação radical quineana e a tradução radical davidsoniana. Segundo Smith, autores como Quine e Davidson incorrem, nos seus projectos, em 'excessos de terceira pessoa', negligenciando a fenomenologia do pensamento e da significação. Não se pode dizer que Wittgenstein negligencie a fenomenologia do pensamento e da significação, no entanto a sua abordagem não tem certamente o poder explicativo dos projectos de Quine e Davidson. Quanto a McDowell, Smith pretende servir-se do insights fenomenológicos deste acerca de entendimento linguístico, embora não partilhe as suas posições metafísicas acerca de um 'espalhamento exteriorista' da significatividade de enunciações (significance of utterances) no comportamento linguístico das pessoas.

Vou para todos os efeitos considerar que o propósito global de Barry Smith é chegar aos princípios de uma "language-based epistemology of mind" (29), uma epistemologia da mente baseada na linguagem (30). O mero facto de Smith falar prontamente de 'entendimento' revela uma divergência relativamente a autores como Quine e Davidson. Estes falam de 'interpretação' e 'tradução', como se os problemas da natureza da mente e da linguagem pudessem ser abordados desde logo em terceira pessoa. Ao tomar o entendimento linguístico como chave para pensar na natureza da mente e da linguagem Smith mostra pretender considerar não apenas o acesso directo e privado do pensador aos seus pensamentos como também a natureza pública desses pensamentos. No entanto, a consideração da natureza pública dos pensamentos não tem por que conduzir-nos a elaborar uma teoria da mente e da linguagem 'em terceira pessoa'. Para usar os termos de B. Smith, é fundamental aqui encarar o problema da reconciliação: a dificuldade das questões em causa prende-se com a necessidade de conjugar a imediatidade dos pensamentos ao pensador, a dimensão interna e consciente do mental, com a dimensão pública e comunitária da linguagem e do pensamento. (Smith admite que o pensamento de Wittgenstein é já

uma luta com o problema da reconciliação; não a considera no entanto bem sucedida).

Voltemos ao problema do auto-conhecimento para vermos de que forma a abordagem de Smith pode fornecer uma nova luz. O auto-conhecimento é conhecimento da mente própria. Frequentemente pensamos que estamos numa posição muito diferente conhecendo a nossa própria mente (algo que fazemos directamente e privadamente) e conhecendo as mentes de outros (algo que fazemos usando como evidência algo de exterior, a sua linguagem e comportamento - este é o foco dos acima referidos projectos de Quine e Davidson).

Consideramos assim a linguagem essencial para sabermos o que os outros pensam, admitimos que ela é essencial para os outros saberem o que pensamos - mas não será a linguagem também importante no nosso conhecimento da nossa própria mente? Smith defende que sim, e que por isso mesmo será essencial investigar "a ligação entre o domínio que temos da linguagem e a relação epistémica que mantemos com as nossas vidas interiores" (31). Só neste quadro poderá vir a ser tratado o problema do auto-conhecimento. Embora o entendimento linguístico nos possa servir como chave para o tratamento do problema do auto-conhecimento, não podemos simplesmente usar uma noção não explicada de entendimento linguístico para falarmos do auto-conhecimento, temos que analisar essa noção.

Compreender o domínio que temos da linguagem passa por procurar saber como passamos a ter entendimento linguístico, sendo que depois acederemos às nossas mentes com esses meios. Para capturar o domínio que temos da linguagem, começemos pela forma linguística da autoridade de primeira pessoa, a 'autoridade do falante'. A autoridade do falante é este sabermos o que queremos-dizer (32) quando dizemos o que queremos dizer. É isto que temos que compreender se queremos compreender a relação epistémica que mantemos com as nossas vidas interiores.

O que é este domínio que temos da linguagem? É muito importante ter em mente que quando Smith pergunta pelo nosso domínio da linguagem ele não está sobretudo interessado em algo como a competência chomskyana, envolvendo mecanismos cognitivos sub-pessoais de reconhecimento do que é linguístico e parsing gramatical (embora essa questão - que é acerca do que se passa no interior do agente isolado - não seja de todo desprezável). O foco de interesse de Smith é o facto de sabermos o que queremos-dizer (what we mean) quando dizemos o que queremos dizer. Quando digo 'Penso que está a chover', ou 'Penso que estou a ver uma parede amarela', eu sei o que quero-dizer (i know what i mean). Como é que isso é possível? Como é possível este domínio do querer-dizer, este facto de aparentemente, interiormente, nós sabemos o que pensamos, sabermos o que queremos-dizer, se não podemos supor qualquer coisa como significação estável e fixada nas linguagens naturais, disponível para ser tomada e usada pelos falantes (ou sequer a existência de línguas como 'o português, ou 'o inglês' (33)? Eu sei o que as minhas palavras e frases significam. Ora, o facto de eu querer-dizer alguma coisa com as minhas próprias palavras não é nem um assunto puramente privado (no sentido em que não encontraremos algo como uma iniciação solipsística e arbitraria do querer-dizer), nem um assunto puramente público (não há entidades abstractas que seriam as línguas naturais independentemente dos falantes, com significados fixados, que seriam aprendidos, convocados e instanciados pelos falantes nas enunciações). Quanto ao que é dito, o que é dito é privado ma non troppo: grande parte da nossa vida mental, daquilo que preenche a nossa subjectividade, é constituída por pensamentos com conteúdo objectivo, partilhável, para falarmos como Frege (pensamentos no sentido que aqui interessa envolvem pretensões epistémicas: não estamos a falar de qualia, de subjectividade puramente fenoménica - teríamos no mínimo que argumentar primeiro que é isso que a subjectividade essencialmente é). Como podemos capturar o querer-dizer concretizado nesses pensamentos? Produzindo linguagem e compreendendo-a. Ora compreender a linguagem que nós produzimos é algo que é feito a partir de fora pelos outros relativamente a nós, tendo em conta o nosso comportamento (explicar como isso supostamente se passa, como se passa de evidência

comportamental exterior para significação e mentalidade, é o projecto da interpretação radical de Quine e o projecto da tradução radical de Davidson). No entanto, nós próprios, aparentemente, não temos que nos interpretar a nós mesmos a partir do nosso comportamento, como os outros fazem, para sabermos o que queremos-dizer. Simplesmente sabemos 'a partir de dentro'. Ainda assim, temos que admitir que os frutos do nosso querer-dizer podem ser exteriormente capturados. Inclusivamente, o querer-dizer as nossas auto-atribuições de mentalidade pode ser exteriormente capturado (recordemos: 'Penso que está a chover', ou 'Penso que estou a ver uma parede amarela', ou 'Sei que desejo possuir aquele objecto', ou 'Estou enjoada'.) Estas não são enunciações incompreensíveis por outrem. São até atribuições de mentalidade que outros podem fazer-nos. Como é então possível que as nossas auto-atribuições mentais coincidam com as atribuições que nos são feitas por outrem a partir da observação do nosso comportamento? (e o facto é que podem, e coincidem).

Estas são as perguntas de Smith. Nas respostas que desenvolve, Smith parte de pressupostos externalistas (o mental, o significado, estão na mente, mas não dentro das nossas cabeças) ao mesmo tempo que pretende evitar aquilo a que chama a (inconcebível, 'mágica') exteriorização que caracteriza a forma como alguém como McDowell trata a questão da significatividade de enunciações: localizar, como faz McDowell, o significado das palavras das pessoas à 'superfície das próprias palavras' (34) parece a Smith uma exteriorização demasiada, uma radicalização injustificada do exteriorismo necessário ao tratamento da questão em causa. Nos termos de Smith, devemos, é certo, observar um 'constrangimento de publicidade' no tratamento do problema da reconciliação. Este constrangimento de publicidade não tem, no entanto, que nos conduzir a uma leitura exteriorizante radical como a de McDowell (ou como poderia a significação, nestas circunstâncias, estar sequer na mente dos falantes que querem-dizer o que querem-dizer com as suas palavras?).

A ideia de Smith é que (usando agora o termo de Davidson) a intersubjectividade é importante aqui. A subjectividade do entendimento linguístico que nos interessa aqui será 'ganha' em grande medida pela 'partilha de um mundo' com outrem. O passo em frente do modelo sob análise relativamente ao modelo wittgensteiniano (entendendo por modelo wittgensteiniano um puro exteriorismo de regras e um nada epistémico da primeira pessoa) é o seguinte: enquanto criaturas linguísticas, o nosso mundo interior é parte do mundo dos outros - trata-se de procurar compreender como e porquê isso vem a ser assim.

Como afirmei, localizar o significado das palavras das pessoas à superfície das palavras parece exteriorização demasiada a Smith. Como evitar a exteriorização excessiva no tratamento da significação em que recaem Wittgenstein e também McDowell utilizando esta nova sugestão da intersubjectividade? Analisemos de novo a situação em que eu digo o que quero-dizer e os outros dizem o que querem-dizer. Vamos imaginar - embora sabendo que essa é uma pretensão não fundamentada - que existe de facto uma língua comum, o português, cuja natureza e propriedades são fixadas por uma particular comunidade. Seria fácil imaginar como os falantes comunicam entre si: eles dão os mesmos significados às mesmas palavras, seguem as mesmas regras. Mas como poderíamos explicar o conhecimento de primeira pessoa do querer-dizer alguma coisa com alguma palavra P? Essa autoridade sem esforço, evidência ou inferência, quando eu digo 'eu amo-te' ou 'dá-me uma colher' deve parecer-nos extremamente estranha se os significados são fixados a partir de fora pelas praticas da comunidade (como Wittgenstein especificamente quis, com a ideia 'o significado é o uso').

Sabemos o que os outros querem-dizer baseados em evidência acerca da forma como usam palavras e em inferência, i.e. através do seu comportamento linguístico. Mas o facto de o comportamento linguísticos de outrem ser importante para pensarmos na significação não deve persuadir-nos de que os 'relatos de terceira pessoa' de Quine e Davidson, a interpretação radical e a tradução radical, estão a usar

dados suficiente e estão no bom caminho para compreender a mente e a significação. Precisamente, eles deixam de fora o entendimento em primeira pessoa de que os falantes gozam: é como se os sujeitos de Quine e Davidson, na interpretação radical e na tradução radical, fossem unicamente 'comportamento'. A situação que ambos, Quine e Davidson, consideram é uma situação em que o teórico - o intérprete, o tradutor - observa comportamento de humanos e as circunstâncias em que estes produzem enunciações. Ora, em que é que Quine e Davidson têm razão e em que é que eles não têm razão ao fazerem isto assim? Eles têm razão em considerar o comportamento, e especificamente o comportamento linguístico, essencial para pensar na natureza da representação e do pensamento; têm razão em insistir na publicidade e partilhabilidade da significação. Mas eles estão errados de acordo com Smith em pensar que não há nada na significação para além daquilo que é observável no comportamento aberto. Mas que fundamento temos para afirmar tal coisa? Basta-nos comparar um intérprete radical quineano e um tradutor radical davidsoniano, por um lado, e uma criança por outro. Eles não estão na mesma circunstância quando se trata de ser exposto ao comportamento linguístico de outrem. O intérprete formula hipóteses acerca do comportamento de outrem falando já ele próprio uma linguagem: fora de si ocorrem asserções não interpretadas, comportamento de outros que ele considera como acontecimento físicos. Mas ele fala já uma linguagem. Com a criança não é assim, ela não fala linguagem nenhuma, ela vai adquirir uma linguagem, com base nas estruturas de que é inatamente dotada e na evidência a que é exposta. Segundo Smith, se queremos entender o entendimento linguístico, a criança deve interessar-nos mais do que o intérprete ou tradutor radical, com a presunção de 'exterioridade sapiente já linguística' que esses conceitos envolvem. Ora compreender a criança implica compreender a importância da comunicação e da intersubjectividade na teoria da mente e do pensamento.

Recapitulando: queremos saber como pôde vir a ser o caso, para esta criança que hoje fala português, que ela tenha investido sons e marcas determinados de significação. Ela hoje inevitavelmente ouve como palavras portuguesas ocorrências de sons e marcas que para uma outra criança, que fala chinês, continuam a ser insignificantes. Segundo Smith, McDowell oferece-nos uma boa caracterização fenomenológica da situação: para a criança que fala português palavras portuguesas são imediatamente significativas. Nós próprios, quando dominamos uma língua, 'ouvimos' significação sem qualquer esforço voluntário activo, 'ouvimos' o que as outras pessoas querem-dizer nas suas palavras mesmas, não temos que interpretar o que quer que seja (35). Nada disto ocorre com um outra pessoa que não tem o domínio dessa língua. É isto que o domínio de uma língua nos dá - e o projecto de Smith é acomodar tais factos numa teoria do entendimento linguístico que obedeça ao constrangimento de publicidade e resista ao movimento exteriorizante.

O constrangimento de publicidade traduz-se, basicamente, em considerar que tudo o que uma pessoa pode querer-dizer é o que outros, convenientemente equipados, podem saber que ela quer-dizer. A exteriorização é uma interpretação excessivamente radical do constrangimento de publicidade, à qual é preciso resistir. De acordo com a descrição atrás, se dominamos uma língua, nós não temos que interpretar nada perante ocorrências determinadas de sons e marcas: a nossa atenção é imediatamente capturada por alguma coisa presente nas palavras que só pode ser vista ou ouvida nelas por aqueles que entendem a língua. Se temos o domínio da língua, vemos e ouvimos o conteúdo ou significado das palavras, e não os meros sons ou marcas, que é o que todos os outros humanos, que não dominam a língua, vêm e ouvem. Se isto é assim, devemos defender que o significado de asserções de outrem numa língua está em geral, à vista, em factos publicamente disponíveis? Como podem estas entidades fenomenológicas estar literalmente lá, à superfície do comportamento? E como é que é possível que estejam lá e ao mesmo tempo eu saiba o que quero-dizer quando digo o que quero-dizer? Smith tem dúvidas quanto à interpretação que McDowell faz da sua própria descrição da fenomenologia de significar, e pensa que uma leitura correcta do constrangimento de publicidade não nos obriga a chegar a concluir que a significação está publicamente manifesta no

comportamento linguístico. Qual é a alternativa?

Parte do que está em causa é restaurar a autoridade da primeira pessoa, sem a confundir com infalibilidade, i.e. sem fazer dela uma autoridade total, única, quanto ao querer-dizer. Uma abordagem genética da questão pode mostrar-nos o caminho. Antes de mais, para compreendermos o que é, para crianças, quererem-dizer alguma coisa com coisas de linguagem, devemos olhar para a aquisição de linguagem, e não para situações do género quineano ou davidsoniano.

Adquirir linguagem é vir a dotar de significação aquilo que começa por ser desprovido dela, por exemplo uma palavra P (adquirir linguagem não é reconhecer um significado que sempre esteve lá, na palavra). Uma proposta forte quanto à forma de as crianças adquirirem linguagem é aquela de acordo com a qual as crianças dotam P de significação, pela atenção conjunta (joint attention) a objectos percebidos em comum com outrem. A atenção conjunta é um tema recentemente muito estudado por filósofos e psicólogos (36), precisamente pela sua importância nestas questões da linguagem, comunicação, atribuição de mentalidade a outrem e pensamento acerca de coisas objectivas no mundo. O debate pode também ser formulado em termos mais davidsonianos como um debate sobre triangulação, ou triangulações. É muito importante notar que o facto de Davidson colocar o problema do pensamento e da linguagem em termos de triangulação nos seus escritos mais recentes representa uma inflexão interiorizante no seu pensamento. De resto Davidson fornece-nos contribuições importante para o problema que aqui nos interessa. Ele tratou directamente as duas questões cuja relação está aqui em causa (o conhecimento da nossa própria mente e o adquirir de significação de enunciações linguísticas) e fê-lo no contexto de uma investigação acerca dos tipos de triangulação que têm que estar no lugar para o pensamento objectivo ser possível (37). A ideia de Davidson é que o pensamento objectivo requer uma interacção entre pelo menos dois sujeitos e o objecto, ou objectos, no mundo. Numa 'primeira triangulação' há duas criaturas e um objecto, e de acordo com Davidson pode haver condicionamento mas não representação. Numa segunda triangulação, a exploração do conceito de verdade objectiva pelo sujeito está envolvida na interpretação e intenções comunicativas (38).

Mas deixando de lado a interessante, e relevante, discussão das triangulações segundo Davidson, voltemos a Smith. O ponto essencial que Smith quer sublinhar é que se não pressupusermos uma linguagem pública com significados já existentes, tudo o que pode servir-nos para alguma coisa ser dotada de significado são os entendimentos de pessoas. Ora uma palavra pode servir como etiqueta para um objecto comumente percebido por duas pessoas. Quando falamos, estamos a tentar moldar e coordenar a experiência de outrem de forma a este pensar num assunto da mesma maneira que nós; trata-se de usar dispositivos de coordenação de experiências. As mentes ficam assim ligadas umas às outras no 'investimento' de sons e marcas por significação. Vir a significar alguma coisa, para alguma coisa, não é, assim, um acontecimento num e para um indivíduo isolado, mas um acontecimento que tem que ser visto nesta situação de triangulação, só assim pode acontecer.

Um outro facto é também importante: uma pessoa só pode ouvir em palavras os significados que ela própria lhes atribui. O caso de primeira pessoa será, desta forma, básico nos funcionamentos da significação. Eu baseio-me no caso de primeira pessoa para saber o que outra pessoa quer-dizer com aquilo que diz. Quando falo, tomo as minhas palavras como significando tal e tal. Quando ouço palavras, tomo-as como significando aquilo que eu próprio quero-dizer com elas. Noutras palavras, eu entendo-te a ti pela forma como me entendo a mim.

O problema torna-se saber como seria uma teoria da significação satisfatória que tornasse o caso da primeira pessoa assim básico. Que género de teoria seria essa? Os passos envolvidos são claros: temos que começar por explicar o que é, para palavras, significarem alguma coisa, temos que explicar como, e por que razão, temos direito

epistémico de pretender, com base no nosso próprio entendimento, que há entendimento dos outros, e temos que explicar como é que tudo isso se passa. Está aqui em causa, de forma fundamental, uma relação entre primeira e segunda pessoa, e Smith propõe uma analogia entre olhar e ver para capturar o teor dessa relação.

O que é então, para palavras, significar alguma coisa? Olhemos para a aquisição de linguagem. As crianças adquirem linguagem pela atenção conjunta (joint attention) a objectos percebidos em comum. Trata-se de uma coordenação de mentes, das mentes do falante e do ouvinte, e ela assegura que os significados investidos dependem do que está na mente do outro. Assim, o facto de eu querer-dizer o que digo com as minhas palavras não depende apenas de mim: eu tomo a palavra P como significando S, quero-dizer S com P, sou eu que quero-dizer mas que tudo funcione depende da satisfação de condições que envolvem outras pessoa e partes partilhadas do ambiente. Satisfeitas estas condições, quando digo o que quero-dizer, digo o que quero-dizer.

E por que razão temos direito epistémico de pretender, com base no nosso próprio entendimento, que há entendimento linguístico nos outros? Eu assumo que qualquer pessoa que usa a palavra P quer dizer com ela o que eu quero-dizer. Mas como é que eu posso ter a certeza disso? Ou por que devo sequer esperar tal coisa?

Wittgenstein, em *Da Certeza*, tal como é interpretado por Crispin Wright, diria que sem essa pressuposição presente desde o início as práticas comunicativas nem sequer funcionariam; trata-se de uma suposição estratégica, por defeito, não sustentada empiricamente. Smith concorda que há uma suposição por defeito, pensa que a suposição por defeito é de facto Quem quer que use a palavra P quer-dizer com ela aquilo que eu quero-dizer mas defende que ela é baseada na situação inicial de aprendizagem partilhada e atenção conjunta.

É claro que adoptando esta explicação genética ficamos com um novo problema, o problema de saber qual é a motivação psicológica para fazermos essa suposição por defeito. Smith avança então uma analogia, tomada da percepção visual, para nos ajudar a compreender esta. Suponhamos que alguém olha, e vê algo que não vemos. Nós olhamos para onde essa pessoa olha e vemos também. Fazemos isto tão automaticamente que nem pensamos. O mesmo ocorre com a linguagem, e com a forma como as coisas de linguagem significam. Temos então o seguinte: ao contrário do que acontece com as outras pessoas em relação a mim, eu não observo nem interpreto os meus usos de coisas de linguagem, apenas os entendo. Eu observo os usos de outrem, e não posso ter a experiência de entendimento que têm. No entanto uso e entendimento vão juntos nas nossas mentes, como olhar e ver.

Nestas condições de explicação do entendimento linguístico, o que é a autoridade de primeira pessoa característica do auto-conhecimento? Não é qualquer coisa como uma forma de contacto, ou observação, ou expressão; é apenas "uma assimetria, entre atribuições de mentalidade em primeira e terceira pessoas" (39). Mas a aportação epistémica da primeira pessoa é muito mais relevante do que no modelo expressivo.

Vou, também aqui, recordar o que Davidson defende acerca de auto-conhecimento (40), para depois, por ajustamentos, chegar ao que B. Smith defende e que me parece mais correcto. Para Davidson, estar num estado mental é estar num estado físico com uma determinada descrição intencional a partir do comportamento. Ora se estar num estado mental p é isso, como é que uma pessoa sabe, ela própria, o que está a pensar? Não há espaço interior, não se trata de uma questão de acesso ao espaço interior, partimos do princípio de que é a linguagem que nos relaciona com a nossa vida mental. Mas como podemos então explicar a autoridade de primeira pessoa quando dizemos 'Penso que está a chover', ou 'Penso que estou a ver uma parede amarela', ou 'Sei que desejo possuir aquele objecto', ou 'Estou enjoada'? De acordo com Davidson, apenas através da linguagem

podemos individualar os nossos próprios estados mentais. Mas o ponto principal aqui é que para fazermos isso, não temos que nos interpretar, usualmente sabemos o que queremos-dizer com as nossas palavras. Ora a verdade é que Davidson não nos diz como, apenas diz que conhecemos as nossas mentes porque sabemos o que dizemos quando dizemos o que pensamos. Mas como sabemos? - podemos nós perguntar de novo. E Davidson responde à pergunta céptica com um argumento transcendental: 'é isso que é ser um sujeito, não podíamos não saber o que queremos dizer e ainda assim sermos falantes' (41). A intenção de Smith é evitar o excesso de terceira pessoa e este argumento transcendental - é por isso que se ele propõe, como acabámos de ver, enfrentar a questão prévia do entendimento linguístico em termos genéticos.

Conclusão

Penso que é no contexto de uma investigação acerca da natureza do entendimento linguístico que o problema do auto-conhecimento deve ser tratado. Esta é a única forma de evitar as pretensões não justificadas dos modelos observacional cartesiano e expressivo wittgensteiniano. Uma investigação acerca da natureza do entendimento linguístico permite-nos não rejeitar nem as aportações da terceira pessoa nem as aportações da primeira pessoa numa teoria da linguagem e do pensamento (esta teoria da linguagem e do pensamento impõe-se como tarefa prévia ao tratamento do problema do auto-conhecimento). Outras soluções para o problema do auto-conhecimento que não foram aqui abordadas, tais como os chamados 'modelo do compromisso' e 'modelo da racionalidade' (Gertler 2003), que não rejeitam as pretensões epistémicas do auto-conhecimento, poderão ser desenvolvidas a partir de esclarecimento básicos quanto ao entendimento linguístico. O mesmo de resto quanto à tão discutida questão de saber se é necessário a um agente possuir conceitos de crenças e desejos para ser auto-consciente. Para decidirmos se um agente pode ou não ser 'cego a si próprio' (self-blind) de várias maneiras precisamos de saber se a linguagem é constitutiva, necessária ou apenas suficiente para conhecermos a nossa própria mente, e a discussão do entendimento linguístico é um passo essencial nesse sentido (42).

Referências

- Bermudez, J.L., 1998, *The Paradox of Self-Consciousness*, Cambridge MA, MIT Press.
- Burge, Tyler, 1988, Individualism and Self-Knowledge, *Journal of Philosophy* 85, 649-663.
- Carruthers, Peter, 2001, Higher order theories of consciousness, *Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/entries/consciousness-higher/#2>
- Davidson, Donald, 1986, A Nice Derangement of Epitaphs, in R. Grandy & R. Warner eds., *Philosophical Grounds of Rationality*, Oxford, Clarendon Press, 157-174.
- Davidson, Donald, 1994, Knowing One's Own Mind, in Q. Cassam (ed.), *Self-Knowledge*, Oxford, Oxford University Press, 1994.
- Davidson, Donald, 2001, The emergence of thought, in D. Davidson, 2001, *Subjective, Intersubjective, Objective*, Oxford, Oxford University Press.
- Davidson, Donald, 2004, Paradoxes of irrationality, in D. Davidson, 2004, *Problems of Rationality*, Oxford, Oxford University Press.
- Dretske, Fred, 1999, The Mind's Awareness of Itself, *Philosophical Studies* 95, 103-124.
- Eilan, Naomi, 2005, Joint Attention, Communication and Mind, in Eilan, N. (ed.), *Joint Attention - Communication and Other Minds - issues in philosophy and psychology*, Cambridge, Cambridge University Press.
- Gertler, Brie, 2003 *Self-Knowledge*, *Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu>.
- Gulick, Robert van, 1999, Self-Knowledge, in Wilson, R. & Keil, F., 1999, *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge MA, MIT Press.
- Gopnik, Alison, 1999, Theory of mind, in Wilson, R. & Keil, F., 1999, *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge MA, MIT Press.

Guttenplan, Self-deception, in Guttenplan 1994, A Companion to the philosophy of mind, Oxford, Blackwell.

Martin, M.G.F., 1997, European Journal of Philosophy

Marques, António, 2003, O interior - linguagem e mente em Wittgenstein, Lisboa, FCT / Fundação Calouste Gulbenkian.

McDowell, John, 1981, Anti-realism and the epistemology of understanding, in McDowell, John, 1998, Meaning, Knowledge and Reality, Harvard, Harvard University Press.

McDowell, John, 1986, In Defence of Modesty, in McDowell, John, 1998, Meaning, Knowledge and Reality, Harvard, Harvard University Press.

McDowell, John, 1998, Response to Crispin Wright, in Wright, Crispin, Smith, Barry and MacDonald, Cynthia, 1998.

Miguens, Sofia, 2002, Uma Teoria Fisicalista do Conteúdo e da Consciência - D. Dennett e os debates da filosofia da mente, Porto, Campo das Letras.

Miguens, S., 2003, Teorias representacionais de 1ª e 2ª ordem na filosofia da mente, *Análise*, 23.

Miguens, S., 2004, Racionalidade, Porto, Campo das Letras.

Miguens, S., 2005, J. Fodor e os problemas da filosofia da mente, *Intelectu* 11.

Morando, Claro, 2005, Teorias da Mentalidade, *Intelectu* 11.

Perry, John, 1998, Myself and I, in Marcelo Stamm ed, *Philosophie in Synthetischer Absicht*, Stuttgart, Klett-Cotta, 83-103.

Perry, John, 1990, Self-notions, *Logos*, 17-31.

Perry, John, 1979, The Problem of the essential indexical, *Nous*, 13, 1, 3-21

Putnam, Hilary, 1975, The Meaning of Meaning, *Philosophical Papers* 2, Cambridge, Cambridge University Press, 215-271.

Smith, B., 1998, On Knowing One's Own Language, in Wright, Crispin, Smith, Barry and MacDonald, Cynthia, 1998.

Smith, B., no prelo, Knowing What You Mean - the active role of the listener - Comunicação apresentada no Workshop Meaning and Communication, Lisboa 3-5 Junho 2005, Instituto de Filosofia da Linguagem da Universidade Nova de Lisboa (a publicar nas Actas do Workshop).

Wimmer, H. & Perner, J., 1983, Beliefs about beliefs: representations and constraining function of wrong beliefs in young children's understanding of deception, *Cognition* 13: 103-128.

Wright, Crispin, 1998, Self-Knowledge: the wittgensteinian legacy, in Wright, Crispin, Smith, Barry and MacDonald, Cynthia, 1998.

Wright, Crispin, Smith, Barry & MacDonald, Cynthia, 1998, *Knowing Our Own Minds*, Oxford, Oxford University Press.

Notas:

(1) Enquadramento: O presente trabalho enquadra-se no Projecto de Investigação Rationality Belief Desire II - from cognitive science to philosophy (POCI/FIL/55555/2004). Um dos objectivos centrais do Projecto é compreender a diferença entre as abordagens filosóficas dos problemas da natureza da mente e da racionalidade e as abordagens características da ciência cognitiva. Assumindo que existe um objecto de investigação comum à filosofia e à ciência cognitiva (a mente, a sua natureza e processos), assumindo que ambas as abordagens são legítimas e não se excluem uma à outra, a nossa hipótese é que a diferença entre a filosofia e a ciência cognitiva se relaciona com o lugar concedido, na filosofia, à perspectiva de primeira pessoa e ao entendimento naquilo que se entende por 'teoria da mente'. O tratamento do problema do auto-conhecimento aqui apresentado inscreve-se na tentativa de elucidar a natureza da 'perspectiva de primeira pessoa' e do 'entendimento'. Estas questões são frequentemente deixadas por tratar nas abordagens da racionalidade feitas em termos de agentes, crenças e desejos. Aqui pretende-se ir 'ao interior do agente' e compreender como é que este sequer sabe e pensa que tem crenças e desejos.

(2) O auto-engano supõe o intuito deliberado (por parte de um eu) de se fazer acreditar algo que sabe ao mesmo tempo que é falso, como se os desejos alcançassem aí o 'direito' de moldar as crenças. Por que seremos susceptíveis de auto-engano? Relacionar-se-á isto com o facto de os humanos serem melhores no auto-conhecimento do que outros agentes cognitivos, mesmo aqueles que são também

presumivelmente conscientes, tais como animais de outras espécies? A possibilidade de auto-engano é obviamente um desafio para a teoria da mente, na medida em que põe em causa a própria unidade do eu, pelo menos tanto quanto esta envolve um compromisso com a coerência das crenças próprias (cf. DAVIDSON 2004, Paradoxes of Irrationality), bem como o sentido de autoria das acções e pensamentos próprios. O auto-engano põe ainda em causa o que está envolvido nos chamados Paradoxos de Moore (a impossibilidade de situações como aquela em que um agente professa a seguinte crença 'Está a chover mas eu acredito que não está a chover'). A um nível consciente, isto parece ser assim; essa pode no entanto não ser a última palavra.

(3) É verdade, no entanto, que o ponto de vista de outros sobre o que pensamos e sentimos, bem como o conhecimento que estes podem ter do nosso comportamento, pode ajudar a tornar "os lapsos no nosso auto-conhecimento inteligíveis" (SMITH 1998: 394).

(4) O que aqui me interessa no cogito cartesiano é a posse, reconhecida pelo pensador, dos seus próprios pensamentos.

(5) GOPNIK 1999.

(6) Cf. o clássico WIMMER & PERNER 1983. Os elementos de que se necessita para compreender o que está em causa encontram-se na seguinte situação. Crianças observam, numa peça de marionetas, esta situação: Maxi, o protagonista, põe um chocolate numa caixa e sai da sala. Enquanto Maxi está fora, a sua mãe remove o chocolate para um armário. A pergunta feita às crianças-observadoras é a seguinte: onde é que Maxi vai procurar o chocolate quando entrar outra vez na sala? De forma a responder acertadamente, a criança tem que ser capaz de distinguir entre o que ela própria sabe ser verdadeiro acerca do actual estado do mundo (o chocolate está no armário) e aquilo que ela sabe sobre os estados mentais de Maxi (ele pensa que o chocolate está na caixa). Crianças com três anos falham muito, com quatro anos são normalmente bem sucedidas, o que mostra que são capazes de atribuir crenças falsas a outros agentes e pensar nelas como guiando o comportamento desses agentes.

(7) Cf. MORANDO 2005, *Intelectu*11.

(8) Pode obviamente falhar, como por exemplo no caso do autismo, mas o ponto é exactamente o mesmo: também aí não se trata de qualquer coisa como um argumento (a favor da existência de outras mentes), que seria inválido, mas sim de um défice cognitivo. Quando a discussão acerca de determinados problemas envolve cientistas empíricos e filósofos, equívocos são inevitavelmente gerados. Isso acontece no caso do auto-conhecimento. É comum dizer que a marca do auto-conhecimento é a autoridade da primeira pessoa. É no entanto conveniente ter ideias claras acerca do que é e do que não é a autoridade de primeira pessoa. No fim deste artigo defender-se-á que "A autoridade de primeira pessoa não é uma forma de contacto. É apenas uma assimetria, entre atribuições de mentalidade em primeira e terceira pessoas". De qualquer forma desde já é possível compreender que por exemplo uma afirmação como a da psicóloga Alison Gopnik, segundo a qual 'pesquisas empíricas mostram que a doutrina filosófica da autoridade de primeira pessoa está errada' não me pareça aceitável: se a autoridade de primeira é, como acima propus, certeza e não algum tipo de infalibilidade, pesquisas empíricas não poderiam conduzir a uma tal conclusão. Há no entanto algo de interessante na afirmação de Gopnik, embora dependente da sua particular posição relativamente ao estatuto da teoria da mente, posição essa que não temos que partilhar. Gopnik defende a chamada *theory theory*, ideia segundo a qual o desenvolvimento cognitivo dever ser explicado por analogia com a mudança teórica em ciência - assim, as crianças desenvolveriam uma sucessão de teorias da mente para explicar a experiência e o comportamento próprios e de outros (como em teorias científicas, seriam postuladas entidades e leis, previsões tornar-se-iam possíveis, e as teorias mudariam mediante evidência contrária e experimentação). O ponto interessante - mesmo que se

discorde da theory theory quanto ao estatuto da perícia da leitura de mentes - é que Gopnik considera que isto se aplica à leitura quer de outras mentes quer da mente própria: o conhecimento da mente própria é tão teórico como o conhecimento da mente de outrem. Essa é a razão pela qual Gopnik afirma que 'pesquisas empíricas mostram que a doutrina filosófica da autoridade de primeira pessoa está errada'. Gopnik está no entanto a pressupor, injustificadamente, que 'a doutrina filosófica da autoridade de primeira pessoa' implica a ideia segundo a qual aquilo de que o sujeito pensante tem conhecimento íntimo e directo é verdadeiro.

(9) Autores como Tyler Burge e Hilary Putnam preocuparam-se com este assunto (cf. BURGE 1988, PUTNAM 1975), apresentando conhecidos argumentos e experiências de pensamento como a Terra Gémea; mais em geral teorias naturalistas do conteúdo como as de Dretske, Millikan, Dennett, Fodor, vão no mesmo sentido. O ponto é o seguinte: assim como não encontramos o que dá significado à palavra 'cão' nas letras 'c', 'a', 'o' não encontramos olhando para os neurónios do nosso cérebro aquilo que eles representam - isto não é negar que representem, mas dizer que temos que olhar para fora (da cabeça, do cérebro) para sabermos o que o que está dentro representa.

(10) WRIGHT, SMITH & McDONALD 1998, Introduction, 13-14.

(11) McDOWELL 1998: 48-49.

(12) O que quer dizer qualquer coisa como: não há nada entre o estado observacional e o estado observado.

(13) Esta objecção assume obviamente o externalismo acerca de conteúdos mentais, e por isso não poderia ser por exemplo endereçada ao próprio Descartes.

(14) Trazendo a discussão para o âmbito da filosofia da mente, o que está em última análise em causa é a disputa entre teorias da consciência fenomenal e teorias representacionais. Os modelos de que falo só fazem sentido uma vez tendo optado por uma teoria representacional.

(15) Para uma comparação destes autores cf. MIGUENS 2003, Teorias Representacionais de 1ª e 2ª ordem.

(16) Esta é, como é sabido, uma ideia desenvolvida por Kant.

(17) Dennett é, de resto, um dos autores cuja obra pretendemos manter como referência no presente Projecto.

(18) DRETSKE 1999, The Mind's Awareness of itself.

(19) Uma outra forma de afastarmos o cartesianismo é a ideia defendida por exemplo por Davidson, relativa ao auto-conhecimento de algo como mental segundo a qual a posse dos conceitos de crença e desejo é condição necessária para a auto-consciência. Como tal proposta passa pelo papel da linguagem no auto-conhecimento, deixo-a para o ponto em que trato do entendimento linguístico.

(20) MARQUES 2003.

(21) Os paralelos com o expressivismo em ética são claros. Cf. MIGUENS 2004, Cap. 4, Natureza das pretensões normativas éticas: expressivismo.

(22) É esse o foco de análise de MARQUES 2003.

(23) WRIGHT 1998.

(24) Casos como os de John Perry estão...entre um e outro (John Perry tem uma sequência de textos bem conhecidos sobre este assunto, cf. PERRY 1979, PERRY 1990, PERRY 1998).

(25) WRIGHT 1998: 35.

(26) WRIGHT 1998: 44.

(27) Esta pergunta vem de Dennett - cf. ponto 2 do presente artigo.

(28) É certo que Wittgenstein tenta fazer precisamente isso, a questão é saber se consegue.

(29) SMITH 1998: 404.

(30) Esta ideia - a ideia segundo a qual uma epistemologia da mente deve ser baseada na linguagem, i.e. não é possível tratar a questão da natureza da mente sem tratar a linguagem - é obviamente relevante para o problema discutido em MIGUENS, S., 2005, J. Fodor e os problemas da filosofia da mente, neste mesmo número da *Intelectu* (o problema geral do artigo sobre Fodor é saber quais são os problemas da filosofia da mente).

(31) SMITH 1998: 392.

(32) Vou usar as expressões querer-dizer, quero-dizer onde em inglês se diria 'I mean...'.
(33) Em *A Nice Derangement of Epitaphs* (DAVIDSON 1986) Donald Davidson chegou a afirmar provocatoriamente 'There's no such thing as a language'. Esse desafio consiste na prática em constatar que a ideia de uma língua-E determinada como o português ou o inglês é mais uma decisão política do que qualquer outra coisa: não está 'lá fora no mundo' nada, auto-subsistindo para ser aprendido pelos falantes que viremos a caracterizar como 'falando uma mesma língua', a que devemos chamar 'o português' ou 'o inglês'. O desafio merece ser considerado e é de resto levado a sério pelos linguistas empíricos. De qualquer modo aqui é uma questão absolutamente incontornável.

(34) McDOWELL 1981.

(35) McDOWELL 1986.

(36) Cf. EILAN 2005.

(37) DAVIDSON 2001

(38) A primeira triangulação ocorre em animais não humanos e crianças pequenas - as criaturas envolvidas aprendem a correlacionar as reacções de outras criaturas com mudanças ou objectos no mundo aos quais a própria criatura também reage (Davidson dá aqui vários exemplos de comportamento animal coordenado). A ideia é que tudo pode acontecer por condicionamento e sem representação. Esta primeira triangulação é importante porque certos aspectos do que virá a ser pensamento (nomeadamente a objectividade do pensamento) não poderiam ocorrer sem tal pré-condição (dizer que o pensamento é objectivo é dizer que é tornado verdadeiro por um mundo independente daquilo que o pensador pensa). Só a interacção envolvida na triangulação dará a possibilidade de pensar que pensamentos podem ser verdadeiros e falsos. Para Davidson, esta 'triangulação pré-cognitiva' é necessária mas não suficiente para a existência de pensamento. De facto, para Davidson, apenas criaturas linguísticas serão capazes de pensamento objectivo. Apenas

enunciações linguísticas são comportamento suficientemente estruturado para constituírem evidência do mundo comum e da natureza deste. Isto envolve, obviamente, defender que não há pensamento sem linguagem, e que só há sujeitos pensando sofisticadamente, e objectivamente, sobre o mundo quando há uma teoria da mente ela própria sofisticada.

(39) SMITH, Barry, 1998, in Wright, Smith MacDonald.

(40) DAVIDSON 1987.

(41) SMITH 1998: 416.

(42) SMITH 1998: 405.