

Compreender a mente e o conhecimento é um propósito subjacente a tudo o mais que se pretenda fazer em filosofia. A reunião dos textos deste volume teve como objectivo disponibilizar um trajecto possível para a exploração de questões relativas à mente e ao conhecimento, correspondente àquela que tem sido a minha prática na Faculdade de Letras da Universidade do Porto desde 1996, tendo por um lado a teoria da mente e do conhecimento tal como ela se apresenta contemporaneamente e por outro lado a história da filosofia do conhecimento. Estes textos, que tiveram uma vida anterior, na sua maioria na Internet, tratam sobretudo de temas e autores contemporâneos. Pretendi com a Introdução que lhes acrescentei exemplificar a comunicação que penso existir entre os tópicos e autores tratados e a história da filosofia, em particular autores como Descartes, Hume e Kant. Do ponto de vista pedagógico, este livro prolonga a experiência iniciada com o manual de filosofia da linguagem publicado em 2007.

**Sofia Miguens** é Professora Associada do Departamento de Filosofia da Faculdade de Letras da Universidade do Porto e Investigadora no Instituto de Filosofia da mesma faculdade, onde coordena o MLAG (Mind, Language and Action Group). Tem ensinado sobretudo nas áreas de epistemologia, filosofia da linguagem e filosofia da mente e dirigido projectos de investigação nas mesmas áreas. Publicou anteriormente *Uma Teoria Fisicalista do Conteúdo e da Consciência—D. Dennett e os debates da filosofia da mente* (2002), *Racionalidade* (2004), *Filosofia da Linguagem—uma introdução* (2007) e *Será que a minha mente está dentro da minha cabeça? Da ciência cognitiva à filosofia (ensaios)* (2008), bem como numerosos artigos.

**Compreender a Mente e o Conhecimento**  
Sofia Miguens

**Compreender  
a Mente e o Conhecimento**  
Sofia Miguens

# COMPREENDER A MENTE E O CONHECIMENTO

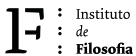
SOFIA MIGUENS

Porto  
2009

## Edição



## Apoio



## Ficha Técnica

*Título*

*Compreender a Mente e o Conhecimento*

*Autor*

*Sofia Miguens (n. 1969)*

*Capa*

*Manuela Teles*

*Impressão*

*Greca - Artes Gráficas*

*Depósito Legal*

*304024/09*

*ISBN*

*978-972-8932-50-3*

*ISSN*

*1646-6527*

*À memória de Fernando Gil  
(1937-2006)*



«Se pensares nisso verás que o interior da tua própria mente é a única coisa de que podes estar certo. Seja o que for em que acredites (...) é baseado nas tuas experiências e pensamentos, sentimentos e impressões dos sentidos. É só a isso que tens acesso directo, quer vejas o livro nas tuas mãos, sintas o chão debaixo dos teus pés, ou te lembres de que D. Afonso

Henriques foi o primeiro rei de Portugal, ou que a água é H<sub>2</sub>O. Tudo o mais está mais afastado de ti do que as tuas experiências e pensamentos internos e é só através destes que te alcança» T. Nagel, *O que quer dizer tudo isto?*

«Pensar? Aqui descubro: o pensamento é; apenas este atributo não pode ser separado de mim. Eu sou, eu existo, isto é certo. Mas por quanto tempo? Certamente enquanto penso, porque pode porventura acontecer que se eu cessasse totalmente de pensar deixaria, desde logo, inteiramente de ser. Agora não admito nada que não seja necessariamente verdadeiro: portanto eu sou, por precisão, apenas uma coisa pensante, isto é, um espírito ou uma alma ou um intelecto ou uma razão, termos cuja significação ignorava antes (...) e assim reconheço que nenhuma das coisas que posso perceber graças à imaginação pertence ao conhecimento que tenho de mim mesmo e que devo pelo contrário desviar o espírito delas, para conhecer bem distintamente a natureza deste. Mas que sou eu então? Uma coisa pensante. O que quer isto dizer? Quer dizer: uma coisa que duvida, que compreende, que afirma, que nega, que quer, que não quer, que também imagina e que sente» R. Descartes, *Meditações sobre a Filosofia Primeira, 2ª Meditação*

«Imagine o leitor que existe uma máquina cuja estrutura produz os pensamentos, os sentimentos, as percepções. Agora imaginemos essa máquina aumentada embora preservando as mesmas proporções, de modo a podermos entrar dentro dela. O que observaríamos? Nada, senão partes que empurram e movem outras partes e nada que pudesse explicar a percepção, o pensamento, os sentimentos» G. Leibniz, *Monadologia*, Parágrafo 17



# ÍNDICE

<i>PREFÁCIO</i>	9
<i>ORIGEM DOS ARTIGOS</i>	11
<i>COMPREENDER A MENTE E O CONHECIMENTO</i>	13
Que forma deve ter uma introdução à epistemologia?	15
No centro de todas as coisas: as <i>Meditações</i>	17
O cepticismo humeano: ‘nada afastará estas dúvidas’	26
Kant: o transcendental como alternativa ao naturalismo humeano	37
Conclusão	54
Apêndice 1 – Leibniz e os mundos possíveis	57
Apêndice 2 – Hume e a religião	64
Apêndice 3 – Por trás da concepção kantiana de espaço e tempo	66
<i>CONHECIMENTO</i>	
1. Introdução à Teoria da Mente e do Conhecimento – Parte I	71
2. Introdução à Teoria da Mente e do Conhecimento – Parte II (aspectos históricos)	87
3. A filosofia e o pensamento como ‘medida do mundo’	107
<i>MENTE</i>	
4. A mente segundo Chomsky	135
5. Daniel Dennett e a Inteligência Artificial	141
6. Alguns problemas de filosofia da Inteligência Artificial	145
7. Critérios de experiência (Wittgenstein, Malcolm e Dennett acerca dos sonhos)	155



8. <i>Qualia</i> e razões – uma história da interioridade natural	167
9. Recensão a António Damásio, <i>The Feeling of What Happens</i> (1999)	203
10. Models of understanding – minds and machines	211
11. Heterofenomenologia	223

#### *IDENTIDADE PESSOAL E ACÇÃO*

12. Problemas de identidade pessoal	235
13. A identidade pessoal e a posição original rawlsiana	263
14. Liberdade, cognição e acção	295

#### *FILOSOFIA E LINGUÍSTICA*

15. As ciências cognitivas e a naturalização do simbólico – a mente computacional e a mente fenomenológica	315
16. Linguistas e filósofos: maneiras de fazer teoria da mente	351
17. Language and thought (the nature of mind from G. Frege and J. Fodor to cognitive linguistics)	385

#### *RACIONALIDADE E FILOSOFIA CONTINENTAL*

18. Um certo desdém de Habermas face a Derrida	399
--	-----

## PREFÁCIO

Os artigos aqui reunidos tocam temas tais como as crenças, o conhecimento, a história da filosofia do conhecimento, a identidade pessoal, a consciência, a acção, a racionalidade e as relações entre a filosofia e a linguística no que respeita à teoria da mente. Estão organizados tematicamente sob os títulos ‘Conhecimento’, ‘Mente’, ‘Identidade Pessoal e Acção’, ‘Filosofia e Linguística’ e ‘Racionalidade e filosofia continental’. Todos eles tiveram uma vida anterior, na sua maior parte na Internet. Ora, a vida electrónica de textos é sujeita a vicissitudes específicas. Nomeadamente, os *sites* podem ter um tempo de vida delimitado – essa é uma das razões que me levam a reunir aqui textos de mais de uma década (1997 a 2009). Apesar dessa limitação de existência, parece-me que as coisas que escrevi e que foram publicadas *on-line* chegaram a muito mais lugares do que aquelas que tiveram apenas vida impressa. Chegaram desde logo a estudantes de Filosofia, em Portugal e no Brasil, e onde quer que o Português seja a chave para aquilo que se quer ler e isso sempre me deu muita satisfação. Tenho que agradecer muito particularmente à Sara Bizarro, a criadora da revista electrónica *Intelectu*, com a qual colaborei desde 1999 e onde foram inicialmente publicados muitos dos textos aqui reunidos, por me ter dado essa possibilidade de difusão alargada.

Outra vida destes textos e outra razão para os incluir neste volume é o ensino na FLUP, nas áreas que sempre me ocuparam, nomeadamente a filosofia do conhecimento e a filosofia da mente, com uma ligação à filosofia moral e à filosofia da acção, especificamente através das questões da identidade pessoal e da racionalidade. É assim também com uma finalidade didáctica que aqui publico estes textos, procurando ligar o passado e o futuro do ensino na FLUP, e por isso juntei aos artigos uma Introdução. Penso sobretudo no ensino em áreas ancoradas na introdução inicial à epistemologia (‘Filosofia do Conhecimento’), a disciplina que lecciono

desde 1996 na licenciatura em Filosofia. Espero assim continuar, embora com bem menor ambição sistemática, o trabalho que iniciei com o manual de filosofia da linguagem publicado em 2007.

Algumas notas práticas: mantive as línguas originais em que os artigos foram escritos – tem sido nossa política no trabalho de investigação que fazemos na FLUP que se escreve e fala ou em português ou em inglês, sem mais discussões acerca das razões para tal. Mantive também o formato original: alguns artigos são precedidos por resumos, outros não, uma vez que uns foram publicados em revistas ou apresentados em colóquios que assim o exigiam e outros não. Devo dizer que acredito na colagem de materiais de forma a manter a história destes, por isso não introduzi uniformidade nem fiz cortes de maior monta nos textos que aqui se reúnem, o que significa que haverá certamente repetições de conteúdo entre eles. No entanto, e porque eles foram escritos com finalidades práticas diferentes, acredito que isso poderá ajudar a compreender as questões discutidas e as propostas avançadas.

Ronda, Andalucia, Verão de 2009

# ORIGEM DOS ARTIGOS

## CONHECIMENTO

1. Introdução à Teoria da Mente e do Conhecimento – Parte I, *Intelectu* 8, Maio de 2003.
2. Introdução à Teoria da Mente e do Conhecimento – Parte II (aspectos históricos), *Intelectu* 8, Maio de 2003.
3. A Filosofia e o Pensamento como medida do Mundo. Conferência do Ciclo *Qual é a medida do Mundo? – A Escala de Abel Salazar*, Reitoria da Universidade do Porto 2004 (comunicações publicadas *on-line* no site do IRICUP).

## MENTE

4. A mente segundo Chomsky, *Intelectu* 1, Fevereiro de 1999.
5. Daniel Dennett e a Inteligência Artificial, *Intelectu* 1, Fevereiro de 1999.
6. Alguns problemas de filosofia da IA, *Intelectu* 3, Fevereiro de 2000.
7. Critérios de Experiência (Wittgenstein, Malcolm e Dennett acerca dos sonhos), *Intelectu* 4, Setembro de 2000. Também publicado em Maria José Cantista (ed.), *A Dor e o Sofrimento – Abordagens*, Porto, Campo das Letras, 2001, pp. 373-385.
8. Qualia e Razões – Uma História da Interioridade, em Maria José Cantista (ed.), *Dor e Sofrimento – Uma perspectiva interdisciplinar*, Porto, Campo das Letras, 2001, pp. 235-275.
9. Recensão de António Damásio, *The Feeling of What Happens*, *Revista da Faculdade de Letras-Série de Filosofia*, nº17, 2001, pp. 187-195.
10. Models of understanding – minds and machines, Conferência no Colóquio *Interdisciplinaridade*, Universidade Nova de Lisboa (RICI), 27 de Fevereiro de 2004, organizado por Fernando Gil e também seminário no âmbito do Grupo de Discussão em Ciência Cognitiva da

Universidade do Porto, organizado por Fernando Santos (Maio de 2004). Publicado *on-line* no primeiro *site* do MLAG entre 2005 e 2008.

11. Heterofenomenologia, in *Ciência e Vida, número especial*, Filosofia e ciência cognitiva, (J. Teixeira org.), São Paulo, 2009.

#### IDENTIDADE PESSOAL E ACÇÃO

12. Problemas de Identidade Pessoal, *Intelectu* 6, Dezembro de 2001. Também publicado na *Revista da Faculdade de Letras-Série de Filosofia*, nº18, 2002, pp. 139-163.
13. A Identidade Pessoal e a Posição Original Rawlsiana, *Revista Portuguesa de Filosofia*, LIX, 1, 2003, pp.139-170.
14. Liberdade, Cognição, Acção, conferência apresentada no 3º Encontro Nacional de Professores de Filosofia (Lisboa, Liceu Maria Amália, 7-8 Abril 2005, comunicações *on-line* em [www.spfil.pt](http://www.spfil.pt))

#### FILOSOFIA E LINGUÍSTICA

15. As Ciências Cognitivas e a Naturalização do Simbólico - A mente computacional e a mente fenomenológica, *Revista da Faculdade de Letras-Série de Filosofia*, nº14, pp. 385-427, 1997
16. Linguistas e Filósofos: maneiras de fazer teoria da mente, *Revista da Faculdade de Letras-Série de Filosofia*, nº15-16, 1998-1999, pp. 327-365.
17. Language and Thought (the nature of mind from G. Frege and J. Fodor to cognitive linguistics)", em Augusto Soares da Silva, Amadeu Torres, Miguel Gonçalves (orgs.), *Linguagem, Cultura e Cognição*, Coimbra, Almedina, 2004, pp. 657-667.

#### RACIONALIDADE E FILOSOFIA CONTINENTAL

18. Um certo desdém de Habermas face a Derrida, *Trabalhos de Antropologia e Etnologia*, vol. 39 (1-2), 1999, pp.11-32.

**COMPREENDER A MENTE  
E O CONHECIMENTO**



## Que forma deve ter uma introdução à epistemologia?

Resolvi chamar a esta colecção de artigos *Compreender a mente e o conhecimento*. O pretexto próximo para o título são os dois primeiros textos reunidos sob o título ‘Conhecimento’, os artigos *Introdução à Teoria da Mente e do Conhecimento – Parte I* e *Introdução à Teoria da Mente e do Conhecimento – Parte II (aspectos históricos)*. Estes dois artigos são as iniciações didácticas que utilizo há mais de dez anos como apoio às disciplinas de Filosofia do Conhecimento na FLUP. Eles são aqui reproduzidos sem alterações de maior às versões publicadas na revista *Intellectu* e reflectem – embora talvez não de forma inteiramente explícita, já que se trata de textos introdutórios, tanto quanto possível de fácil leitura, e essencialmente pontos de partida para o desenvolvimento dos cursos – a minha particular decisão quanto à forma de uma introdução à epistemologia. A ideia é fazer uma introdução à epistemologia que seja por um lado uma iniciação à terminologia e aos problemas da área em textos contemporâneos (problemas relativos a justificação de crenças, tipos diversos de cepticismo, teorias da justificação epistémica tais como o fundacionalismo, o coerentismo e o fiabilismo, questões relativas à definição do conhecimento e aos problemas de Gettier, natureza do apriori, problema (s) da indução, internalismo e externalismo, contextualismo, etc) e por outro uma história da filosofia do conhecimento, cujos núcleos são autores-chave tais como Descartes, Leibniz, Hume, Kant e Quine.

Desde o tempo em que a Filosofia do Conhecimento era uma disciplina anual na FLUP, as leituras das *Meditações sobre a Filosofia Primeira*, da *Monadologia*, do *Tratado da Natureza Humana*, da *Investigação sobre o Entendimento Humano*, da *Crítica da Razão Pura* e de *Dois Dogmas do Empirismo* sempre nela ocuparam um lugar central, servindo de banca de ensaio para os problemas tratados. Hoje os textos de referência continuam



a ser os mesmos<sup>1</sup>, embora decididamente não sejam lidos de forma tão alargada. Em disciplinas semestrais de Filosofia do Conhecimento como as actuais é muitas vezes através de temas específicos, por exemplo o tema da percepção, que essa vocação histórica da introdução à epistemologia se mantém (no caso da percepção, através de leituras seleccionadas de autores como J. Locke, G. Berkeley, D. Hume, T. Reid, B. Russell e J. L. Austin).

A forma como a introdução é concebida conduz naturalmente a temas relativos à mente – falando de conhecimento fala-se naturalmente de crenças e de tipos de mentes. Na história da filosofia a comparação clássica para falar da natureza da inteligência e da natureza daquilo que é conhecido dava-se entre as mentes humanas e a mente divina, nos textos contemporâneos mentes animais e mentes artificiais reúnem-se às mentes humanas. Este é o fio condutor dos textos reunidos sob o título ‘Mente’ (*A mente segundo Chomsky, Daniel Dennett e a Inteligência Artificial, Alguns problemas de Inteligência Artificial, Critérios de experiência, Qualia e razões, Recensão a António Damásio, The Feeling of What Happens, Models of understanding, Heterofenomenologia*). Os temas desses artigos – que vão desde a inteligência, à consciência, à memória, às relações consciência-inconsciente, sensiência-intelecção, corpo-mente – são de facto temas clássicos da história da epistemologia, aqui tratados a partir de autores contemporâneos como Chomsky, Dennett, Damásio e outros.

‘Identidade pessoal e acção’ são dois temas que nas aulas de epistemologia foram, durante alguns anos, tópicos específicos a tratar nos autores da história da teoria do conhecimento directamente analisados, nomeadamente em Descartes, Leibniz, Hume e Kant. Em todos eles se encontra concepções do eu, da relação do eu com o corpo e com a acção e a iniciação da acção. As obras destes autores atrás identificadas eram lidas em busca (entre outras coisas) de uma teoria do eu e da relação deste com o corpo e o pensamento. Mas o tema da natureza do eu não é apenas uma questão cognitiva, epistemológica e metafísica: ele tem ramificações em áreas da filosofia aparentemente distantes destas. Prolonga-se também, por exemplo, pela filosofia política, o que fica aqui evidenciado com o artigo sobre os reflexos de uma determinada concepção de identidade pessoal na filosofia de Rawls (*Identidade Pessoal e posição original rawlsiana*), incluído nesta secção. O artigo sobre liberdade e acção (*Liberdade, cognição e acção*), embora focando um autor contemporâneo sobre o qual tenho trabalhado

<sup>1</sup> Com excepção da *Monadologia*, infelizmente. A introdução da terminologia dos mundos possíveis para tratar questões relativas a necessidade e contingência foi sempre a grande razão para o estudo do autor na cadeira de epistemologia, no ponto dos programas em que os tópicos (respectivamente do âmbito epistemológico, metafísico e semântico) do a priori/a posteriori, necessidade/contingência, analítico/sintético eram expostos. Outros aspectos da metafísica e da teoria da mente leibnizianas foram muitas vezes também introduzidos, sobretudo para traçar uma oposição à ideia cartesiana de mente como consciência.

bastante, D. Dennett, retoma o tópico da acção no contexto das relações mente-corpo, tratado ao longo dos anos nas aulas de história da filosofia do conhecimento em Descartes, Leibniz, Hume e Kant.

Os textos reunidos sob o título ‘Filosofia e Linguística’ e ‘Racionalidade e filosofia continental’ são aqueles que têm menos relação directa com o ensino da epistemologia na FLUP ao longo dos anos. Eles representam respectivamente aquela que foi a minha própria porta de entrada na filosofia da mente – as implicações de teorias linguísticas numa concepção da natureza da mente, tal como estas foram pensadas por pessoas como por exemplo J. Fodor e N. Chomsky – e o meu interesse inicial pelo tópico da racionalidade, que veio a ocupar livros e projectos futuros<sup>2</sup>. Um dos artigos mais antigos aqui incluído, *Um certo desdém de Habermas face a Derrida* (1998) mostrou-me, ao olhar de novo para ele passado tanto tempo, que o problema que então me preocupava, lendo autores ‘continentais’, o tema da racionalidade e dos limites desta, não é assim tão diferente da questão que ocupou projectos posteriores, desenhados a partir da perspectiva da filosofia analítica.

### **No centro de todas as coisas: as *Meditações***

Sendo todos os textos aqui reunidos unificados pelo desafio geral de compreender a natureza da mente e do conhecimento, e pensando agora na forma de iniciar alguém à compreensão filosófica de tais tópicos, olhando para os esforços pedagógicos de mais de uma década no ensino da Filosofia do Conhecimento, não posso deixar de notar a posição centralíssima das *Meditações sobre a Filosofia Primeira* de Descartes na forma de enfrentar o desafio. Descartes foi sempre um autor essencial nas disciplinas de Filosofia do Conhecimento, desde os tempos em que a leitura integral das *Meditações* era feita nas aulas até hoje, em que a obra aparece de forma mais fragmentada (a propósito do cepticismo acerca do mundo real exterior, como proposta de referência acerca da natureza do pensamento e da natureza da matéria, como exemplo de proposta ‘fundacionalista’ clássica, baseada no cogito e em contraste com o fundacionalismo empirista, ou ainda como um exemplo de representacionismo, pela sua ideia de ideia, funcionando aí como prolegómeno a J. Locke, no tratamento do tema da percepção). Ora essa posição central de Descartes (ou, mais propriamente, das *Meditações*), parece-me certa, e algo a manter. Mas por que será que tal acontece? Por que é que as *Meditações* merecem uma posição tão central

<sup>2</sup> Cf. Miguens 2004, *Racionalidade e Miguens & Mauro 2006, Perspectives on Rationality*.

na tentativa de compreender a mente e o conhecimento? Antes de mais, pelo mergulho vívido no cepticismo: os sonhos, o deus enganador, o génio maligno, a possibilidade de o mundo real exterior ser uma ilusão da minha mente – estas são hoje questões reconduzíveis nos espíritos dos estudantes ao mundo da Matrix, aos mundos virtuais, e como tal perfeitamente pertinentes; o itinerário cartesiano é reproduzível no espírito de cada um. Mas há uma outra razão prática para essa posição central: nas *Meditações* Descartes dá respostas de referência a praticamente todos os problemas que interessa tratar ao procurar compreender a mente e o conhecimento. Esses são problemas quanto ao fundamento do conhecimento, quanto à natureza do pensamento e da matéria, quanto às relações entre pensamento e matéria, quanto à ideia de ‘ideia’, quanto à natureza da ‘substância’, quanto à natureza da ‘verdade’, quanto à aplicação antropológica das relações entre pensamento e matéria numa visão da relação mente-corpo, estendendo-se essa teoria do ‘composto humano’, i.e. do indivíduo humano agora e aqui, cada um de nós, a uma teoria do eu e das acções. As suas opções fundamentais por uma des-naturalização da alma (ou consciência ou espírito ou mente), por uma descontinuidade abrupta entre consciência e não-consciência, pela consideração da união alma-corpo como facto bruto, a sua convicção de que o espírito poderia talvez existir não apenas sem o corpo como também sem as paixões (estados do espírito derivados da união deste com o corpo), a sua ‘intelectualização’ da noção de mente, as suas concepções acerca das relações entre vontade e juízo, são incontornáveis.

Assim, ler as *Meditações* uma a uma, reproduzir em si próprio o trajecto do meditador, será sempre uma introdução ideal à epistemologia (e à metafísica, e à filosofia da mente, de resto). Para alguns, aqueles com mais instintos históricos, isto pode ser feito com mais pendor para o sabor epocal e local, compreendendo detalhes e contingências. Para outros a argumentação a seco bastará – se as *Meditações* são ainda tão importantes é porque podem fazer sentido para cada um de nós, directamente, sem mais delongas<sup>3</sup>.

Uma proposta de leitura seria começar pela ‘Saudação aos Doutores da Sorbonne’ (os ‘Sapientíssimos e Ilustríssimos Senhores Decano e Doutores da Faculdade de Teologia de Paris’) e ouvir Descartes defender que a Filosofia é diferente da Teologia e a religião natural diferente da religião revelada. Ao filósofo cabe o caminho da racionalidade, não o da revelação, mesmo que os tópicos sejam os mesmos, neste caso deus e a alma. Interessa

<sup>3</sup> O que se segue é uma glosa – as referências poderão ser facilmente encontradas em qualquer edição das *Meditações*. As edições portuguesas mais frequentemente utilizadas são as da Livraria Almedina (Coimbra, 1985, Introdução, tradução e notas de Gustavo de Fraga) e da Rés (Porto, 2003, supervisão de Helena Sousa Pereira)

perceber como podem existir – e existem – esses estranhos seres, os ateus. Serão realmente irracionais? Ou será que a existência de deus não é racionalmente demonstrável? Perceber o porquê de haver ateus é parte do que conduz à necessidade de demonstrar racionalmente a existência de Deus e a distinção espírito-corpo que ocupa as *Meditações*. A nós que ‘somos fiéis’, diz Descartes, pode bastar a fé, mas não podemos certamente pensar que assim converteremos ou convenceremos alguém. A aposta deve ser na ‘razão natural’, e esta passará, naquilo que se segue, por um vertiginoso trajecto de dúvida.

A 1ª Meditação intitula-se ‘Das coisas que se podem pôr em dúvida’. O mote é que ‘Não existe nada de que não seja lícito duvidar (.....) não por leviandade mas por certas e fundamentadas razões’. O meditador decide evitar (cuidadosamente, metodicamente) dar o seu assentimento ‘não menos ao meramente duvidoso do que ao abertamente falso’, sabendo no entanto, que ‘não basta eu advertir-me de que devo considerar o duvidoso exactamente como o falso’ pois ‘as opiniões habituais regressam com frequência e apoderam-se da minha credulidade’.

Mas qual será a utilidade de uma dúvida tão geral e tão pouco natural? A utilidade é máxima: ela permite destruir, naquilo em que eu acredito, o duvidoso, o não fundamentado, o irreflectidamente aceite, e esta destruição é absolutamente necessária, uma vez que ‘desde cedo se recebe tantas coisas falsas como verdadeiras’. Só assim será possível ‘desviar o espírito dos sentidos e preparar caminho para o exercício da razão’. São assim enumeradas e analisadas as razões para duvidar, desde as razões comuns (erros e ilusões dos sentidos, enganos nos raciocínios, a loucura, a indistinção entre a vigília e o sonho) às razões metafísicas (o Deus Enganador e o Génio Maligno, que só deixará de ‘existir’ na 4ª Meditação).

O exemplo forte daquilo de que (não) se pode duvidar é algo que se espera que fique no espírito de todo o leitor da obra, pela sua absoluta normalidade – é isso que radicaliza o desafio. Eu estou aqui, sentado junto à lareira, vestido com um roupão de inverno, declara o meditador, e acredito que estas mãos são minhas, que este corpo é meu (este exemplo – este corpo, o meu corpo – será retomado na 6ª Meditação). Será possível duvidar disto? Sim, é possível, e isso é importante para a tarefa de análise e fundamentação do conhecimento. Desde logo, já me aconteceu, reflecte o meditador, acreditar tais coisas (que estou aqui, sentado junto à lareira, com o roupão vestido) quando me encontrava estendido na cama e despidendo (dormindo). E há ainda aqueles loucos que ‘acreditam serem reis, quando são paupérrimos’, ou que são cabaças, ou que são feitos de vidro.

No itinerário do meditador, algo, no entanto, sobrevive ao primeiro estado de dúvida, provocado pelas possibilidades da loucura ou do sonho.

Trata-se das ‘ciências do simples’: a aritmética, a geometria, erguidas sobre as noções primitivas – a imaginação e o sonho não criam os elementos simples que entram nas suas composições. Esse é o ponto de entrada da experiência mental do Génio Maligno. Tratar-se-á depois de saber o que resiste ao Génio Maligno, essa variação de Deus, que mantém todos os predicados deste (muito particularmente a onipotência) mas que é desprovido de bondade. Ele é suficientemente poderoso para me fazer acreditar que tudo existe tal como me aparece (o céu, a terra, os corpos...) sem que nada seja como parece ou sem que nada exista. Dado o Génio Maligno, a própria matemática é duvidosa: mesmo quando faço a adição  $2+3$  ou conto os lados de um triângulo eu posso estar a enganar-me.

O título da 2ª Meditação esclarece-se a si próprio: ‘Da Natureza do Espírito Humano que se conhece melhor do que o corpo’. Neste título decide-se já o dualismo substancial, ontológico e antropológico que será reafirmado e analisado na 6ª Meditação (a 6ª Meditação é a meditação mais interessante no que diz respeito ao problema mente-corpo, pois é escrita em torno do carácter ‘híbrido’ da natureza humana, uma vez resolvidos os problemas epistemológicos e metafísicos mais prementes, problemas quanto à fundamentação do conhecimento e quanto à natureza do pensamento e da matéria).

Para irmos ao núcleo da 2ª Meditação uma pergunta simples pode ajudar: o que é que o meditador não consegue pensar? Ele não consegue pensar que não existe, é esse o limite da possibilidade pensada: penso, logo existo – eis a descoberta do cogito. Mas não é por não conseguir pensar que não existe que ele sabe o que é, ele que sabe, assim, que é, necessariamente. Como investiga então o meditador aquilo que é? O relato do meditador é o mais directo possível: ‘Antes julgava ser um homem, um animal racional, um corpo que anda e se alimenta mas tudo isto é duvidoso (...) descobro: o pensamento é, só este atributo não pode ser separado de mim’. Que espécie de coisa é este ser, que não pode garantir ser, por exemplo, o clássico ‘animal racional’? É uma coisa pensante, uma essência: *res cogitans*. E o que é *res cogitans*? É «Um espírito, uma alma, uma coisa que duvida, que compreende, que afirma, que nega que imagina e que sente» (note-se: até ‘sentir’ é ‘pensamento’). O meditador tem assim alguma certeza, a certeza de algo. Mas não abdica da pergunta perturbadora: quanto dura esta certeza? ‘Eu sou, eu existo, mas por quanto tempo?’ E a resposta que se dá é: existo enquanto penso. A certeza é curiosamente temporária, mas a proposição é necessariamente verdadeira sempre que for pensada. E essa certeza primeira, o cogito, será o fundamento e o modelo de todas as outras certezas. Em termos de conhecimento, é aqui que se começa a entrar em terreno seguro.

A 3ª Meditação (‘De Deus, que existe’) afasta muitos leitores contemporâneos: não é fácil aceitar que Deus caia assim do céu para dar cartas nos

problemas do fundamento, da certeza, da natureza do espírito, do raciocínio e da vontade. O que é que Deus vem fazer? A questão é saber como é que o meditador pode vir a conhecer alguma coisa além do cogito. Em que ponto estão as coisas? Fechando os olhos e os ouvidos, inspeccionando o eu, encontra-se as ideias. Mesmo que os modos de pensar encontrados (as ideias) não existam ‘lá fora’, enquanto existem ‘em mim’ eles são certos. Tem-se então na mão pelo menos qualquer coisa como um critério da certeza: o que se requer para haver certeza é uma compreensão clara e distinta da coisa enquanto pensada. O meditador conclui agora que deve pensar acerca de coisas que antes tinha admitido como certas (as coisas recebidas pelos sentidos, as matemáticas), que só delas compreendia as ideias ou pensamentos. Este é um passo importante e definidor, que conduz a responder à pergunta ‘A que é que se chama afinal ‘ideia’?’ Ideia é tudo aquilo que pode ser imediatamente apercebido pelo espírito, e que em si mesmo e não referido a algo não pode ser nem verdadeiro nem falso. Apenas nos juízos pode existir o verdadeiro e o falso. O que resta ao meditador fazer para analisar as suas ideias? Ele experimenta classificá-las em inatas, adventícias e factícias<sup>4</sup>, mas não vai longe por aí. Perguntando-se por que acredita numa semelhança entre as ideias e o que existe no exterior da mente, o meditador verifica que é por ‘uma certa inclinação’, e não por ‘luz natural’. A luz natural seria uma intuição intelectual não enganadora, enquanto as inclinações frequentemente conduzem a escolher o pior. As ideias não procedem da vontade do meditador, é certo, mas daí não se segue que decorram do que há lá fora (até porque há no ser pensante uma faculdade produtora de ideias não muito bem conhecida). Não foi portanto por juízo certo mas por uma mera tendência não fiável que o meditador acreditou que existiam coisas distintas de si. Em suma, a investigação pela origem das ideias é um caminho sem saída. Mas que outra via estará disponível para investigar se existem coisas fora do pensamento do meditador? Como pode o meditador alguma vez recuperar o direito de falar do seu corpo e das coisas materiais? Nova análise: entre as ideias como modos de pensar não há desigualdade, é certo. No entanto essa ‘desigualdade’ existe na medida em que umas representam uma coisa e outras outra. É sob esse prisma (da *realidade objectiva* das ideias, na linguagem de Descartes) que elas devem ser comparadas. É preciso portanto distinguir, quanto às ideias, a sua realidade objectiva (a ideia enquanto representando x), a sua realidade formal (a própria coisa, na sua existência actual) e a sua realidade material (a sua existência psicológica, como operações do entendimento). É por aqui que

<sup>4</sup> Respectivamente existindo desde sempre na alma, vindas do exterior e resultantes de variação da imaginação, feitas por mim próprio.

Deus vai entrar no sistema metafísico de Descartes. A primeira referência feita a Deus é uma referência a uma ideia entre outras ideias, mas com mais realidade objectiva. O que é que isto significa? Qualquer coisa como uma ideia que é ideia mas que ‘representa mais’ do que as outras ideias. A parada é alta: como é que o meditador pode saber que Deus existe? E como pode todo o sistema do conhecimento humano vir a depender de Deus se Deus está para além da compreensão humana?

Nas *Meditações* o meditador ensaia (e aceita) provas várias da existência de Deus. Os pontos de partida das provas a posteriori são a realidade objectiva da ideia de Deus ‘em mim’, i.e., no meditador, a própria existência do meditador, enquanto pensante e a sua permanência temporal. O ponto de partida da prova a priori da 5ª Meditação é uma ‘impossibilidade de pensar Deus sem a perfeição’: a ideia de uma essência na qual todas as perfeições estão reunidas não é concebível sem a perfeição da existência. No primeiro caso temos qualquer coisa como: tenho a ideia de Deus em mim, eu não posso ser o autor dessa ideia, não pode haver efeito sem causa<sup>5</sup>, Deus deve ser a causa deste efeito em mim, Deus existe.

A entrada de Deus em cena vai significar a viragem da destruição céptica para a construção de um sistema metafísico, mas os problemas dessa entrada são muitos e alguns deles tratados nas próprias *Meditações*. Por exemplo: como é que o meditador saberá que não existe por si ou por causas menos perfeitas que Deus? Como é que o meditador sabe que Deus não continua a ser o Génio Maligno? De onde vem a veracidade atribuída a Deus e que fará toda a diferença? Será que as verdades de razão são ‘factos’? Deus transcende ou não transcende os princípios da razão? O que significa dizer que Deus ‘cria as verdades eternas’? (significa dizer muito pouco – ou melhor, dizer de uma forma que os rebaixa na sua natureza – quer de Deus quer das verdades eternas, segundo os leibnizianos)<sup>6</sup>. E será sequer que a reunião dos atributos divinos (infinitude, perfeição, ser *causa sui*) produz algo de logicamente coerente?

Deus pode ser problemático mas faz muita coisa num sistema filosófico. Dá-nos imediatamente, por exemplo, uma teoria da liberdade, e da forma como esta se reflecte na nossa natureza de pensadores. Na 4ª Meditação, intitulada ‘Do Verdadeiro e do Falso’, o que encontramos é um texto sobre a liberdade e a vontade e um exemplo de teodiceia, i.e. um tipo de discurso sobre a divindade feito directamente, exactamente o tipo de discurso que deixará de poder ser feito em boa consciência depois de Kant. À pergunta que ocupa a 4ª Meditação e que pode ter leituras quer cognitivas

<sup>5</sup> É totalmente misteriosa a entrada deste princípio.

<sup>6</sup> Cf. Apêndice 1, Leibniz e os mundos possíveis.

quer morais (Porque é que um Deus Perfeito não me criou de modo a que eu não me enganasse?) responde o meditador que ele me deu o máximo que me poderia dar: o erro é um sinal de liberdade, Deus criou-me enquanto ser de vontade e de liberdade à sua imagem. O que acontece é que a minha vontade, tal como por exemplo esta se concretiza na afirmação de um juízo, em muito transcende os meus poderes cognitivos. Por isso, engano-me. Culpa minha, não de Deus. Tudo o que ele fez foi fazer-me livre, dotado desta vontade irrestrita.

Nesta altura dos acontecimentos temos um meditador enquanto espírito e várias certezas já, mas nada de mundo exterior. A questão do mundo exterior, corpóreo, terá que passar por clarificações relativas à essência e à existência deste. Na 5ª Meditação, intitulada ‘Das coisas materiais e de Deus que existe’, o meditador retoma da 2ª Meditação o exemplo do pedaço de cera como ponto de análise da essência da matéria. A conclusão dessa análise tinha sido que se conhece melhor o que pertence à cera (e, generalizadamente, a todas as coisas materiais) se se a conhece pelo espírito e não pelos sentidos. A essência do corpóreo é uma ideia, uma coisa pensada. E é essa essência de matéria que é feita existir por Deus, é essa que é a essência da natureza, tal como esta é, nomeadamente, conhecida pela ciência físico-matemática. O mundo material, cuja essência é a extensão, existe – isso ficará assente. Mas tudo aquilo que nós percebemos desse mundo, nomeadamente pelos sentidos, é muito diferente de tal essência e terá que ter uma outra explicação. Estes serão os tópicos da 6ª Meditação (Da existência das coisas materiais e da distinção real entre alma e corpo).

Na 6ª Meditação entra-se no estranho terreno do humano, aquilo que de certa forma mais valera que não existisse, se o propósito de uma investigação é a metafísica e a ciência – e para essas os dados estão já lançados. O que são então a ‘ciência’ e a ‘metafísica’? A ciência (do mundo físico) é a investigação (dedutiva, *more geometrico*) da essência deste, que é *res extensa*. A metafísica é a teoria dos fundamentos da verdade e uma investigação das relações entre as três substâncias (Deus, espírito, matéria enquanto pensada). Deus é a substância espiritual, infinita, incriada e criadora dos espíritos finitos e da matéria, que como é perfeito, é veraz, e assim funciona como garante do conhecimento. A questão da garantia ou fundamento do conhecimento é na verdade um pouco mais complicada do que isto e envolve o célebre ‘círculo cartesiano’ (o ‘círculo cartesiano’ é o facto de ser necessário, no sistema metafísico cartesiano, provar a existência de Deus para garantir a fiabilidade da mente humana e ao mesmo tempo acreditar na fiabilidade da mente humana para provar a existência de Deus). De qualquer forma, o meditador satisfaz-se com os fundamentos encontrados.

Resta articular essa realidade impura que é o humano. O que poderá haver de bom no facto de uma essência pensante estar ligada intimamente



a um corpo, que de alguma forma a ‘contamina’ naquilo que podem ser os seus modos de pensar? Olhando para o corpo, considerando a sua realidade epistemologicamente inútil, somos levados a admitir que os sentidos têm uma utilidade pragmática e não gnosiológica. Eles servem para nos fazer sobreviver, para sermos este ser, neste mundo, não para conhecermos o que quer que seja, para fazer ciência ou filosofia. E ao termos (ou sermos) este corpo somos espírito invadido por paixões da alma – os únicos modos puros do espírito sendo os juízos e as vontades, tudo o resto resultando da ‘união’ com o corpo. Por exemplo explorando a diferença entre imaginação e intelecção (comparando os actos de pensar num triângulo e pensar num quiliógono, como o meditador faz no início da 6ª Meditação) verifica-se que o corpo se intromete em alguns dos modos do pensar: chama-se a isso imaginação, e ela é algo que não é necessário à minha essência de ser pensante.

É preciso também explorar as razões que o meditador tem para pensar que aquele corpo a que chama seu lhe pertence mais do que todos os outros. Como é que o meditador sabe que é distinto do seu corpo e que pode existir sem ele? Sobretudo porque, reflecte o meditador, ele não está naquele corpo como um piloto no navio (se a alma estivesse no corpo como um piloto num navio, eu não sentiria dor quando o corpo fosse ferido, antes ‘inteligiria’ a partir de fora que acontecera uma lesão). A solução passa por uma distinção real das substâncias corpórea e pensante – tenho ideias claras e distintas, e diferentes uma da outra, de matéria e de pensamento. Cada uma é feita existir enquanto tal. Estranhamente, eu sou as duas coisas, uma união dessas duas substâncias (i.e. dessas duas formas de existir por si, que são feitas ser – por Deus – enquanto distintas). O composto humano é qualquer coisa como uma ‘união de facto’ de duas essências distintas. Mas algo nele sou mais eu: este eu é a minha alma, diz o meditador – é através dela que eu sou. Ela é ‘verdadeiramente distinta do corpo’, ela ‘poderia existir sem ele’. Está no entanto unida a ele, e é isso o humano.

Permanecem, obviamente, problemas. O dualismo provará a imortalidade da alma? Nada é dito. Isto chega para pensar em cada um de nós como uma pessoa, esta pessoa, ou faz de todos nós o ‘mesmo espírito’? É uma questão. A identidade pessoal pura identifica-se com a essência pensante, reduzida aos seus modos centrais. O problema é que essa identidade pode resultar em despersonalização quando a alma fôr ‘desunida’ do corpo: parece perfeitamente possível que desapareça qualquer particularidade do eu ao desligar-se este do corpo. E como explicar a acção, o fazer ser ou acontecer algo no mundo através de um corpo, deste corpo? Para Descartes, a causação mental explica-se de forma interaccionista – estamos

perante mais uma referência histórica para a filosofia da mente. O interaccionismo significa que existe influência mútua entre duas substâncias irreduzíveis, e portanto o seu pano de fundo é o dualismo espírito-corpo. A acção é a influência da vontade sobre a matéria e deve ser contrastada com o caso em que o desejo impele o corpo para movimentos, pois neste caso os modos de pensar são resultantes da união da alma com o corpo. Em termos antropológicos, Descartes faz ainda ‘entrar a alma’ para explicar o pensamento e a linguagem, para além da acção. Mas tem ainda algo mais a dizer acerca de acção: tem a dizer que quando a acção é perfeitamente racional ela é livre e só pode ser aquilo que é, pois resulta de clara concepção. Esta situação opõe-se à indiferença (o poder ser de muitas maneiras). Ao mesmo tempo, a explicação cartesiana da acção admite que a alma é justaposta a um corpo-máquina que funciona exclusivamente devido a princípios internos (convém não esquecer que Descartes, o defensor da imaterialidade da alma, é um mecanicista).

Recapitulando: as *Meditações* apresentam um trajecto que vai desde um cepticismo radical acerca do mundo exterior até uma solução fundacionalista acompanhada por uma defesa da des-naturalização da alma (ou consciência ou espírito ou mente), uma defesa concomitante da descontinuidade abrupta entre consciência e não-consciência, a consideração da união alma-corpo como facto bruto, a proposta de que o espírito poderia existir não apenas sem o corpo como também sem as paixões (derivadas da sua união ao corpo), uma ‘intelectualização’ da noção de mente. Aquilo a que chamamos mente ou espírito ou alma é consciência imaterial, substância criada finita; o poder pelo qual conhecemos é puramente espiritual. Os seus modos centrais são os juízos (intelecções) e as vontades; só esses são puros; mas há outros modos de pensar, aqueles que existem em virtude da união ao corpo: as afecções, as paixões e as inclinações (desejos). Os primeiros revelam a actividade do espírito, os segundos a sua passividade. São os primeiros que importam para o conhecimento. O conhecimento dá-se através da gestão que a alma faz dos modos de pensar que há em si. A referência, o apoio primeiro e insubstituível, é o cogito; quando há evidência na relação espírito/modo de pensar, a verdade do que é concebido é garantida por Deus. ‘Verdade’ é, assim, um nome para a certeza intelectual subjectiva garantida por Deus. Deus garante a verdade devido à sua veracidade (decorrente da ‘perfeição’). O âmbito da racionalidade assim caracterizado, os movimentos de pensamento conducentes à ciência e à metafísica, é totalmente apriorista: as verdades são alcançadas na relação entre o espírito criado (e seus modos de pensar) e o espírito incriado, e não por alguma experiência do mundo.

Mesmo se o pensamento contemporâneo tem muito a contribuir também, e em diálogo directo com aquilo que Descartes discutia, é das

*Meditações* que se pode partir. A relação com as *Meditações* será sempre intuitiva, mais intuitiva do que aquilo que alguma vez se consegue com Kant ou com Quine, os outros dois autores igualmente centrais nas disciplinas de Filosofia do Conhecimento ao longo dos anos. Sendo igualmente incontornáveis (o transcendentalismo kantiano e a epistemologia naturalizada quineana devem aparecer, penso, como grandes opções gerais para considerar a natureza da mente, do conhecimento da mente e do conhecimento produzido, por exemplo conhecimento matemático ou físico ou outro) e mesmo contemporaneamente mais directamente pertinentes, os estudantes (ou outros interessados em epistemologia) lidam usualmente de forma muito menos imediata com Kant e Quine do que com Descartes. A ideia é portanto começar pelas *Meditações*, e usá-las para erguer uma referência nos estudos da epistemologia, da metafísica, da filosofia da mente. O único caso análogo ao das *Meditações*, em termos de adesão dos espíritos numa primeira abordagem aos problemas da epistemologia é o caso de Hume, mas propriamente o caso do Adão do *Abstract* do *Tratado da Natureza Humana*<sup>7</sup> – o Adão que não sabe, porque foi criado em pleno vigor do seu entendimento, mas sem qualquer experiência, se o fogo queima, se a água molha, se o pão alimenta e se o sol nascerá de novo amanhã. Ele faz parte também, penso, dos casos para uma teoria da mente e do conhecimento que nos confrontam directamente, sem necessidade de grandes circunlóquios históricos.

### O cepticismo humeano: ‘nada afastará estas dúvidas’

No trajecto pela história da teoria do conhecimento o cepticismo humeano aparece como uma alternativa incontornável a contrastar com o cepticismo cartesiano e com o cepticismo clássico quando pela primeira vez se identifica e caracteriza tipos de cepticismo. Mas como se deve compreender o cepticismo humeano? O que é que o distingue do cepticismo cartesiano que acabamos de ver culminar em garantidas certezas? Talvez o núcleo essencial dessa compreensão deva ser a ideia humeana de uma solução céptica das dúvidas cépticas e o que resta ainda de cepticismo nessa solução. É isso que, além do mais, pode ser ligado ao naturalismo tão difundido na epistemologia contemporânea e do qual David Hume pode ser considerado um introdutor. Em termos de história da teoria do conhe-

<sup>7</sup> O *Abstract* [1740] é um resumo da obra maior de Hume, o *Tratado da Natureza Humana* [1739]. Com o *Abstract*, cujo título completo é ‘An Abstract of a Book lately Published, Entitled A Treatise of Human Nature’, Hume procurou tornar mais compreensível o *Tratado*, resumindo as principais ideias. É aí que aparece a experiência mental do Adão, criado sem experiência e no pleno vigor do seu entendimento e é colocada a questão acerca do que ele poderia inferir acerca do curso do mundo.

cimento, Hume aparece também como aquele cujo desafio céptico vai ser levado tão a sério por Kant, que toda a teoria transcendental do sujeito (com sucesso ou não) lhe responde. Hume é obviamente também a referência para o tratamento do problema (clássico<sup>8</sup>) da indução.

Se as *Meditações* de Descartes usualmente ‘falam’ aos estudantes enquanto texto, a figura pessoal de Hume tem também, penso, atraído muitos. Olhemos por isso um pouco para a pessoa de David Hume. Eis alguém que se preocupou a certo ponto com a sua própria vaidade por se considerar indubitavelmente mais inteligente do que todos os seus companheiros de estudos... A mãe, que o educou, achava-o ‘*uncommonly wake-minded*’ e ‘*acute*’. Apesar dessas preocupações de vaidade, em termos de estilo, segundo o (à partida insuspeito) filósofo francês Gilles Deleuze, no livro que dedicou a Hume, intitulado *Empirisme et Subjectivité*<sup>9</sup>, Hume foi responsável pela introdução de um novo tom em filosofia, um tom de simplicidade e de firmeza. Este tom de simplicidade e firmeza corresponde à ideia segundo a qual há qualquer coisa de natural nestas nossas inquirições intelectuais, a teoria é acerca daquilo que nós fazemos; a teoria é uma prática, um *inquiry*, no qual estamos já de qualquer forma lançados. E o que é que essa prática e a teoria dessa prática têm a ver com a razão? Ainda segundo G. Deleuze, se Hume, enquanto autor e escritor, foi sobretudo um historiador, um moralista e um pensador político<sup>10</sup>, isso não aconteceu sem motivos. Hume foi sobretudo um historiador, um moralista e um pensador político porque a prática, a história, a moral são pela sua própria natureza indiferentes à razão; ora, esta indiferença da prática à razão é o ponto nuclear de todo o pensamento de Hume. É porque a razão não é co-extensiva à realidade que se pode (e deve) colocar a questão da sua natureza, i.e. que se deve perguntar pela razão da razão, vendo a razão como o efeito de alguma coisa e não como o simples aparecer da verdadeira natureza da realidade, de um real racional, como os filósofos racionalistas anteriores faziam. Esta é, em filosofia, a linhagem de Hume mas também a de Nietzsche ou de Freud.

<sup>8</sup> Por contraste com o novo problema da indução tal como foi formulado por Nelson Goodman, em *Fact, Fiction, Forecast*.

<sup>9</sup> Deleuze 1973.

<sup>10</sup> Na história da filosofia e aos olhos de filósofos, Hume aparece sobretudo como o autor do *Tratado da Natureza Humana* e das três investigações que deste vieram a sair. Mas se Hume foi um escritor conhecido no seu tempo deveu-o sobretudo a escritos de outra natureza, ensaística e histórica. Os seus *Essays, Moral, Political* (1741) depois publicados como *Essays, Moral, Political and Literary* (1758) tinham temas variadíssimos, desde comentários políticos sobre governação, parlamento, liberdade de imprensa, partidos, a comentários económicos, comentários sobre as artes e as ciências, a peças sobre o Suicídio, a Superstição e o Entusiasmo, etc. A sua *History of England* (1754-1762) foi provavelmente a obra pela qual foi mais conhecido. Escrevendo história Hume estava em parte a criar e em parte a responder a um novo mercado e a uma nova apetência. A história parecia-lhe ser a leitura preferida das pessoas— e se em meados do século dezoito houve uma expansão enorme do público com hábitos de leitura em Inglaterra, Hume foi em parte responsável por tal fenómeno. A história era para Hume a melhor maneira de alguém se familiarizar com os assuntos humanos e ele recomendava-a fortemente a esse novo público—em larga medida feminino e de classe média.

Para Hume, a razão não é senão um maravilhoso e ininteligível instinto das nossas mentes, que nos conduz para certas sequências de ideias e as dota de qualidades particulares. A razão é um hábito, uma natureza, nada mais. Essa é a ideia central de Hume e ele passará toda a vida a estudar as implicações de tal ideia em âmbitos vários.

Aos dezasseis anos Hume estava já empenhado na escrita da obra *A Treatise of Human Nature*, que virá a publicar aos vinte e sete. Depois passará toda a vida a cuidar deste tratado a que chamou nado-morto, e cujo conteúdo virá a gerar a *Investigação sobre o Entendimento Humano*, a *Investigação sobre os Princípios da Moral* e a *Dissertação sobre as Paixões*<sup>11</sup>. Quase no fim da vida (e já utilizando o tempo passado...) escreveu acerca de si próprio: «Eu era um homem de um humor aberto, alegre e sociável, pouco susceptível às inimizades, com grande moderação nas paixões. Mesmo o meu amor pela fama literária nunca me estragou a disposição. A minha companhia não era inaceitável nem para os jovens e descuidados nem para os estudiosos e literatos. Tinha um particular prazer na companhia das mulheres e não tenho razão para estar descontente com a recepção que encontrei por partes delas» (*My Own Life*)<sup>12</sup>. Mas mesmo se Hume foi de facto pouco dado às inimizades, muitos foram aqueles que se deram à inimizade para com ele. Hume nunca conseguiu por exemplo ensinar na Universidade de Edimburgo, pois tinha inúmeros inimigos, especialmente entre os clérigos. Era considerado ateu, embora sempre tenha afirmado não o ser. Bertrand Russell comentou: «Se Hume não era ateu, quem o será...». Mas essa é talvez uma opinião demasiado taxativa. Hume foi provavelmente demasiado céptico para ser ateu, provavelmente pensou que não havia razões suficientes para um céptico negar a possibilidade longínqua de um Deus

<sup>11</sup> *O Treatise of Human Nature* (1739) foi descrito por Hume como um nado-morto por ter sido recebido com críticas destrutivas e incompreensão. É a partir dele que surgirão as Investigações sobre entendimento humano, moral e paixões. O *Tratado* pretende elaborar uma 'Ciência da Natureza Humana', considerando para isso nomeadamente a origem das percepções da mente humana (impressões e ideias), a imaginação como natureza do espírito humano, os princípios da imaginação como leis da natureza, a natureza e as causas da crença. São ainda tratados vários problemas cépticos e é feita uma apologia do cepticismo. Hume passará toda a vida a tratar do *Tratado* e a reformular e retocar as argumentações apresentadas. O *Essay Concerning the Human Understanding* (1748) [*Investigação sobre o Entendimento Humano*] recompõe os materiais do primeiro Livro do *Tratado*, intitulado 'Do Entendimento'. É retomada a teoria da crença como teor da vida mental e a diferença entre relações de ideias (*relations of ideas*) e questões de facto (*matters of fact*), para tratar a natureza de diferentes tipos de conhecimento. O livro contém ainda a teoria da indução e as célebres e provocadoras secções X e XI sobre os Milagres e sobre uma Providência Particular e um Estado Futuro. O segundo e o terceiro livros do *Tratado* [*Das Paixões e Da Moral*] são sobre as 'impressões de reflexão'. No segundo Livro discute-se as paixões, no terceiro o sentimento moral como uma particular espécie de paixão. Hume tenta resgatar uma parte da natureza humana, as paixões, que foi objecto de avaliações negativas ao longo dos tempos. A moralidade, objecto do Livro terceiro, é para Hume uma questão totalmente humana, fundada na natureza humana e nas circunstâncias da vida. *An Inquiry Concerning the Principles of Morals* (1751) [*Investigação sobre os Princípios da Moral*] é o título do volume onde Hume reelabora o Livro segundo do *Tratado*; *Dissertation on the Passions* (1757) [*Dissertação sobre as Paixões*] é o volume em que Hume recompõe os materiais do Livro terceiro do *Tratado*. Existem traduções recentes em português de todas estas obras de David Hume [cf. referência finais do presente artigo]

<sup>12</sup> Esta autobiografia foi publicada em 1777 como 'The Life of David Hume'. Quando a escreveu Hume estava já doente, tendo morrido em 1776. Está incluída na edição de E. Miller dos *Essays* (*Moral, Political and Literary*).

não providente (i.e. de um Deus que tendo criado o mundo abandona este à sua sorte). Mas mesmo se Hume não foi ateu, ele foi sem qualquer dúvida um grande crítico de dois perigos para o espírito que considerava estarem associados à religião, a ‘superstição’ e o ‘entusiasmo’, tendo ainda atribuído à religião a origem de um grande número de práticas imorais.

Aparte os problemas que a sua reputação de céptico e ateu lhe granjeou e que ficaram na história, Hume é usualmente simplesmente classificado como mais um empirista, em conjunto com J. Locke (1632-1704) e G. Berkeley (1685-1753). Mas o pensamento de Hume não deve visto unicamente a partir da tradição empirista em filosofia do conhecimento, a tradição ligada aos nomes de Locke e Berkeley. De facto, Hume pretendeu algo de mais ambicioso: ele pretendeu estender o método experimental a uma ‘ciência da natureza humana’, abarcando questões que vão desde o conhecimento, até à moralidade, à arte, à forma de viver em sociedade, com todos os problemas económicos e políticos inerentes. I. Newton (1642-1727), constituiu o seu modelo ‘metodológico’ para uma ‘abordagem de experiência’. Mas outra figura esclarecedora da figura de Hume é Adam Smith, o autor de *The Wealth of Nations* e o fundador da economia como ciência: David Hume e Adam Smith foram amigos próximos, interlocutores e escreveram sobre inúmeros temas económicos e políticos comuns. Em suma, partilhando certamente com Locke e Berkeley a preocupação por elaborar uma teoria das ideias, e uma teoria da relação da mente com o mundo, a veia naturalista de Hume conduzi-lo-á por caminhos singulares.

Mas quais foram afinal os principais contributos de Hume para a teoria da mente e do conhecimento? Podemos ver as coisas da seguinte forma. Na teoria da mente, de acordo com D. Dennett (*Kinds of Minds*<sup>13</sup>), Hume mostrou-nos a forma como as ideias pensam por si: não há aí um eu a supervisioná-las. Além de ser uma aplicação ao eu de um dos princípios fundamentais da ciência da natureza humana que desenvolverá no *Tratado* (o princípio: não há ideias sem impressões<sup>14</sup>) o associacionismo é, pelo

<sup>13</sup> D. Dennett 1996.

<sup>14</sup> No *Tratado Hume propõe-se fazer com o “mundo interior” aquilo que Newton fez com o mundo físico exterior: encontrar as leis naturais do seu funcionamento. Estas leis de uma “ciência da natureza humana” serão muito importantes pois todas as ciências estão dependentes de uma ciência do homem [i.e. baseiam-se nesta, na medida em que são produto do seu “objecto”, a mente]. Os únicos fundamentos possíveis para esta ciência são a experiência e a observação: desconhecemos tanto a essência da mente como desconhecemos a essência dos corpos externos. Assim não será estranho que venha a ser impossível estabelecer princípios últimos na ciência da natureza humana. Esta impossibilidade é um “defeito comum a todas as ciências” pois “nenhuma delas pode ir para além da experiência ou estabelecer princípios que não se baseiem nessa autoridade”. No início da sua análise da mente Hume nota que todas as percepções da mente humana são ou Impressões ou Ideias. A diferença entre Impressões e Ideias reside nos graus de força e vivacidade com que estas fazem a sua aparição no pensamento ou consciência. As mais fortes e violentas são as impressões [“as sensações, paixões, emoções, quando aparecem pela primeira vez na alma”], enquanto que as ideias serão “as débeis imagens daquelas no pensamento e no raciocínio”. Todos somos aliás capazes de compreender esta diferença entre sentir e pensar, embora não seja impossível que sentir e pensar se aproximem [por exemplo no sonho, na febre, na loucura, as ideias podem aproximar-se de impressões]. As percepções dividem-se ainda em percepções*

menos parcialmente, uma crítica às concepções racionalistas de subjectividade como ‘espontaneidade’, para usar um termo kantiano que adiante se discutirá<sup>15</sup>. Na teoria do conhecimento, de acordo agora com Quine<sup>16</sup>, o problema da indução continua hoje tal qual como Hume o deixou: o impasse humeano é o impasse humano. Noutras palavras: apesar da sua apologia da extensão do método experimental a uma ciência do homem – o propósito global do *Tratado da Natureza Humana* – a verdade é que Hume submeteu o raciocínio experimental a uma severa análise. Pretendeu explicar o que fazemos e o que pensamos que fazemos quando passamos da observação de eventos particulares presentes para generalizações, leis e previsões de eventos futuros. Quis nomeadamente compreender o

*simples (i.e., que não admitem distinção ou separação) e percepções complexas (i.e., nas quais podem ser distinguidas partes). Por exemplo nesta maçã que eu percebo estão em conjunto, uma cor, um sabor e um cheiro. Eles estão unidos, mas são distinguíveis. Ora, Impressões e Ideias são muito semelhantes, excepto no que respeita à sua força e vivacidade. Pareceria, por isso, em geral, que todas as ideias e impressões se assemelham se não fosse pelo facto de termos ideias complexas que nunca tiveram impressões correspondentes (podemos por exemplo imaginar uma "Nova Jerusalém com pavimento de ouro e paredes de rubis" sem termos tido impressões correspondentes). No entanto, toda a ideia simples tem uma impressão simples que se lhe assemelha (é essa a relação por exemplo entre a "ideia de vermelho que concebemos no escuro" e a "impressão que nos atinge em plena luz do sol"). É interrogando-se sobre a prioridade existente entre Impressões e Ideias, que Hume é levado à afirmação do primeiro princípio da ciência da natureza humana, um princípio que não deve ser desprezado apesar da sua aparente simplicidade: segundo Hume todas as ideias são derivadas de, ou procedem de, impressões. As Impressões podem ser Impressões de Sensação (que aparecem na alma originalmente, a partir de causas desconhecidas para a própria alma) e Impressões de Reflexão, que derivam de ideias. No segundo caso, tudo começa com uma impressão que atinge os sentidos. Desta impressão há uma cópia feita pela mente, que permanece depois de a impressão cessar: esta é a ideia. Esta ideia [por exemplo de prazer ou dor] quando retorna à alma produz novas impressões [por exemplo de desejo, aversão, esperança ou medo]. Estas são Impressões de Reflexão. Estas Impressões de Reflexão são por sua vez copiadas pela memória e pela imaginação e tornam-se ideias; estas podem por sua vez dar origem a outras impressões e ideias. Em suma, as Impressões de Reflexão são anteriores às ideias que lhes correspondem, mas posteriores às impressões da sensação. O exame das sensações pertence à filosofia natural (i.e., à física, à fisiologia) e não à filosofia moral. No entanto o exame das Impressões de Reflexão como as paixões, os desejos, as emoções é importantíssimo na filosofia moral. Como as Impressões de Reflexão surgem a partir de ideias, é preciso analisar o "funcionamento" destas.*

<sup>15</sup> Uma Impressão reaparece na mente como ideia essencialmente de duas maneiras: pela memória e pela imaginação. As ideias da memória são muito mais fortes e vividas do que as ideias da imaginação. A grande diferença entre elas é que, embora não haja ideias nem na memória nem na imaginação sem ter havido impressões prévias, as ideias da memória retêm a ordem e a forma das impressões originais, enquanto o mesmo não se passa com as ideias da imaginação. O poder de variação da imaginação é muito grande. Temos portanto o segundo princípio da ciência da natureza humana que se enuncia dizendo que a imaginação tem liberdade para transpor e alterar as suas ideias. Sendo as ideias assim uníveis e separáveis conforme "agrada" à imaginação, essa faculdade seria completamente desregrada e as ideias seriam unidas unicamente pelo acaso, se a imaginação não fosse regida por princípios universais que a tornam afinal uniforme. Há portanto, de acordo com Hume, princípios em acção na união de ideias, mas esses princípios não produzem uma conexão inseparável. Sublinhe-se que Hume fala aqui de *matters of fact* por oposição a *relations of ideas*: nestas últimas o que é pensado conserva a sua certeza e evidência independentemente daquilo que existe. Hume pensa que o contrário de qualquer questão de facto é possível e concebível pela mente com nitidez e facilidade: que o sol não se levantará amanhã não é uma afirmação menos inteligível e não implica maior contradição do que a afirmação de que ele se levantará, defende na *Investigação sobre o Entendimento Humano*. O problema de Hume será então saber qual é a natureza da prova que nos assegura de qualquer existência real e questão de facto, para além do testemunho presente dos sentidos ou dos registos da memória. Os princípios de união que levam a mente de uma ideia a outra são três: a semelhança, a contiguidade no tempo ou lugar e a causa e efeito. As ideias introduzem-se e sucedem-se assim naturalmente umas às outras. Esta sucessão natural resulta de uma "força de atracção" (*Tratado da Natureza Humana*) que tem tido extraordinários efeitos no mundo mental como a força de atracção no mundo natural. É de sublinhar no entanto que Hume aceitou a diferença entre *relations of ideas* e *matters of fact*, tendo considerado que o conhecimento matemático e o conhecimento lógico serão totalmente independentes da experiência, obtidos 'sem olhar para o mundo' [como defenderam os racionalistas clássicos].

<sup>16</sup> W. V. Quine, em *Ontological Relativity*.

estatuto dessa crença matricial, a crença segundo a qual o futuro será como o passado, i.e. a crença numa regularidade do curso do mundo, e sugeriu que essa crença, que fundamenta as nossas inferências indutivas, seria ela própria, resultado de uma inferência indutiva, com todas as fragilidades inerentes. É esse o impasse.

Como pano de fundo da análise humeana das inferências indutivas, encontra-se ainda a sua teoria da crença. Hume tem algo de novo a dizer, em termos de história da filosofia, quanto àquilo que é uma crença, e é construindo sobre essa concepção de crença que se perguntará depois como é que sabemos coisas que parecemos saber por olharmos para o mundo repetidamente, como por exemplo que o sol se vai levantar amanhã, ou que o fogo queima. Como é possível que a partir de eventos observados nós sejamos capazes de formar crenças acerca de futuros não observados e de generalizar? Como é possível acreditarmos que o mundo vai continuar a comportar-se no futuro como se ‘comportou’ no passado?

O que é, então, a crença? A crença, escreve Hume no *Tratado*, é um sentimento, um *feeling to the mind*, só explicável em termos de força, firmeza, peso, influência nas paixões e pensamentos de uma pessoa, gerado pelo costume, que é uma *gentleforce*. A crença é a pedra de toque para o nosso pensamento dos *matters of fact*, das questões de facto ou de experiência. Em questões de facto os contrários são possíveis (eu posso conceber que o sol se levanta e que o sol não se levanta, que o fogo queima e que o fogo não queima), eu posso conceber ambos, nada mais os distingue senão a crença. Mas a crença não depende da vontade, não pode ser manipulada arbitrariamente: eu acredito, não finjo que acredito nem me ordeno que acredite, que o sol se levanta e que o fogo queima. A crença depende de cadeias causais mas parte sempre do presente. Ela acompanhará o nosso ‘instinto de indução’.

Esta forma quase-observacional de falar da crença falando da forma como o nosso espírito se comporta é profundamente distinta da forma de falar de verdades em autores anteriores na história da epistemologia. Mais do que as suas propostas acerca de impressões e ideias – em parte partilhadas com outros empiristas clássicos – ela estabelece um contacto directo de Hume com a epistemologia contemporânea.

Voltemos ao problema da indução. Como se comportam então as nossas crenças sobre questões de facto? Como é que eu sei que o chão não vai desaparecer debaixo dos meus pés daqui a instantes? Como é que eu sei que o céu não vai cair em cima da minha cabeça? Como é que eu sei que a água molha e o fogo queima? O problema humeano da indução é o problema da legitimidade das inferências que partem dos objectos da razão que são *matters of fact* observados para *matters of fact* não observados



(passados, futuros, presentes para além do âmbito estrito do directamente observável). O problema é sério tanto quanto diz respeito a toda a investigação empírica (por oposição ao âmbito das *relations of ideas*, o domínio da lógica e da matemática, domínios que, verdade seja dita, ao contrário do que foi o caso com Descartes e Leibniz, nunca constituíram o foco de Hume em teoria do conhecimento). Se quisermos considerar a questão em termos de argumentos, aquilo que está em causa é essa diferença entre argumentos dedutivos e indutivos que é a vulnerabilidade destes últimos aos novos dados, nomeadamente aos eventos futuros (embora todos os argumentos indutivos sejam vulneráveis a evidência contrária).

Hume centra a sua análise no facto de todo o raciocínio sobre *matters of fact* convocar a relação causa-efeito. O que é isso a que chamamos uma causa? No exemplo emblemático, o exemplo das bolas de bilhar da *Investigação sobre o Entendimento Humano*, temos (i) contiguidade, (ii) prioridade no tempo do evento a que chamamos ‘Causa’ relativamente ao evento a que chamamos ‘Efeito’, (iii) conjunção constante. Mas não há mais nada. Eu vejo, diz Hume na Secção IV da *Investigação sobre o Entendimento Humano*, uma bola de bilhar mover-se em linha recta em direcção a uma outra e o movimento da segunda bola é sugerido à minha imaginação como resultado do seu contacto ou impulso. Mas não seria possível conceber que uma centena de resultados possíveis se seguissem daquela causa? O que é que faz com que o nosso espírito espere que específicos efeitos se sigam a específicas causas? Porque é que o nosso espírito ‘espera’ que a segunda bola se mova, e não que ela caia no nada ou se transforme numa vaca ou numa mesa?

A “experiência mental” do Adão, proposta no *Abstract*, o resumo do *Tratado da Natureza Humana* escrito para tornar este de mais fácil leitura, ajuda-nos a pensar nisto: imaginemos Adão, criado em pleno vigor do seu entendimento e sem experiência. Inferiria ele que E (o efeito) se segue a C (a causa), por exemplo no caso acima? A resposta é não. Então, de acordo com David Hume, quando nós próprios inferimos nessas circunstâncias e esperamos coisas determinadas, é o hábito (*custom*) e não a razão que guia o nosso pensamento. Só por costume nós supomos tais coisas, a razão só por si nunca seria capaz de dar este passo: todos os raciocínios que dizem respeito a causas e efeitos estão fundados na experiência, e todos os raciocínios fundados na experiência estão fundados na crença de que o curso da natureza continuará uniformemente o mesmo. Ora, já vimos qual era a origem desta última crença.

Esta forma de olhar para o nosso ‘instinto indutivo’ é subversiva, inclusivamente enquanto teoria da racionalidade, e é, como tal, um marco na história da epistemologia. Ela marcará formas futuras de pensar na

natureza da causação. Nós vemos constantemente causação entre objectos e eventos quando pensamos sobre questões de facto, nós esperamo-la subjectivamente; no entanto o Adão do *Abstract* não “espera” como nós esperamos. Isto é assim porque não há aí possibilidade alguma de demonstração, não há qualquer contradição em imaginar que as coisas pudessem ser de outro modo. A pura razão não vê nada no evento-Causa que torne possível dele deduzir o evento-Efeito. O efeito é totalmente diferente da causa e nunca será nela descoberto. O nosso conhecimento aqui deriva inteiramente da experiência e todas as conclusões experimentais promanam da suposição de que o futuro será idêntico ao passado. Tentar a prova desta suposição com argumentos relativos à existência é, segundo Hume, circular. E se o curso da natureza mudasse, o passado deixaria de ser regra para o futuro e toda a experiência se tornaria inútil.

É importante não pensar que este escrutínio do pensamento baseado na experiência de alguma forma aniquila a nossa possibilidade de conhecer: há uma solução céptica para as dúvidas cépticas aqui encontradas. É certo que é o costume que guia o nosso pensamento dos *matters of fact* e não a razão. É certo que em todos os raciocínios a partir da experiência, a mente dá passos que não são apoiados ‘racionalmente’. É certo que a conexão ligando os fenómenos reside em nós: é certo que isso a que se chama pensar é sempre transcender o dado e que nós não podemos conhecer princípios últimos. Mas o hábito não é uma propensão irregular. Conviria mais caracterizá-lo como uma ‘sabedoria da natureza’, essencial à nossa sobrevivência, um instinto que a natureza implantou em nós e que impele o pensamento numa marcha correspondente à que existe nos objectos externos. O hábito é um princípio da natureza humana, reconhecido universalmente e muito conhecido pelos seus efeitos, dirá Hume na *Investigação sobre o Entendimento Humano*.

Este optimismo marca, por estranho que pareça, o cepticismo de Hume. Passemos então à natureza geral do cepticismo humeano, para além da teoria da indução. Na Conclusão do Livro I do *Tratado da Natureza Humana*<sup>17</sup> Hume caracteriza a escolha humana (quando se trata de racionalidade) como sendo uma escolha entre ‘uma razão falsa e nenhuma razão’. Esta observação serviu em cursos de epistemologia ao longo dos anos para marcar o contraste com o cepticismo cartesiano e a forma pacificadora como este conduz ao cogito e ao fundacionalismo e a última palavra de Hume no *Tratado*. A última palavra de Hume é que “Só descuido e desatenção resolverão as nossas dúvidas cépticas” – i.e. não há

<sup>17</sup> Conclusão da Parte IV (O sistema céptico e outros sistemas de filosofia) do Livro I (Do Entendimento) do *Tratado*.

lugar para qualquer pacificação cartesiana. O que se dá então? Restará, em alternativa, a abdicação céptica clássica de qualquer possibilidade de conhecer? Não, as coisas para Hume não são tão-pouco assim: temos a natureza e a forma como ela 'é' em nós e nos guia nas nossas tarefas cognitivas. Olhemos para os casos emblemáticos do cepticismo acerca do mundo real exterior e acerca do eu tal como estes são expostos no *Tratado*<sup>18</sup>. O que é que nos conduz a acreditar na existência de um mundo independente e continuado exterior ao nosso pensamento? Não podem ser os sentidos, que não nos dão senão seres partidos, interrompidos. Se a crença no mundo real exterior resultasse, em alternativa, da pura razão, muito poucos fariam tal inferência, como muito poucos raciocinam acerca de tal ponto. Ora nós não fingimos apenas que há existência exterior independente e continuada de um mundo: nós de facto acreditamos na existência de um mundo exterior ao pensamento independente e continuado. Hume pensa que o que nos conduz a acreditar na existência continuada e independente de um mundo externo é a imaginação, a natureza operando em nós, não a nossa razão.

E o que é que nos conduz, por outro lado, a acreditar num eu simples e continuamente existente, num eu que é o que eu sou realmente, fixo para além de tudo o que sinto e penso, se desse eu não há qualquer impressão? Alguns filósofos pensam que estão continuamente, intimamente, conscientes disso a que chamam eu, continua Hume, mas ele não vê nada disso em si. Tudo o que lhe parece haver é um 'feixe ou colecção de diferentes percepções que se sucedem umas às outras com uma rapidez inconcebível e que estão em perpétuo fluxo e movimento'. Quando penetro intimamente em mim próprio, diz Hume de forma célebre, só apanho percepções, nunca um eu. Como é que podemos pensar que as percepções pertencem a um eu ou *self*? Muito simplesmente, para Hume não existe uma qualquer substância a que se chame o *self*, e que compareça à introspecção, mas apenas o feixe das impressões. Ao contrário desses metafísicos que pensam que temos a todo o momento consciência do eu, o 'resto dos homens', diz Hume, não são senão esses 'feixes ou colecções de diferentes percepções que se sucedem umas às outras com uma rapidez inconcebível e que estão em perpétuo fluxo e movimento'. Mas nós acreditamos que somos um eu – isso Hume não nega, o que nega é que disso tenhamos impressão ou demonstração. Não é isso que o eu é: como o mundo real exterior, também ele é efeito da natureza actuando em nós,

<sup>18</sup> Secções II [Do cepticismo em relação aos sentidos] e VI [Da identidade Pessoal] da Parte IV [O sistema céptico e outros sistemas de filosofia] do Livro I do *Tratado*.

pela imaginação. Noutras palavras, quando acreditamos que há mundo exterior, quando acreditamos que somos um eu, o que temos são crenças naturais fundamentais, possivelmente crenças que se encontram em todos os homens, mas que não são baseadas nem em impressões nem na razão: são imaginárias.

Como ficamos? Vale a pena ler cuidadosamente a Conclusão do Livro I do *Tratado*. Lá Hume fala da visão do perigo, que o ‘atinge de melancolia’, e pergunta-se: serei algum “monstro, incapaz de estar em sociedade, expulso de todo o comércio humano”? e inicia uma recapitulação sobre o que descobriu sobre a nossa razão. E o que foi que descobriu? Descobriu que a experiência, o hábito e a imaginação são o que nos sustenta; descobriu que não podemos conhecer directamente, por introspecção, uma ‘essência’ da mente, nem afirmar desde logo que a mente é por exemplo uma alma imortal; descobriu que é preciso experimentar, esperar para ver o que a mente faz, para perceber o que a mente é e como é. Concluiu que o conhecimento da mente não tem um estatuto diferente do conhecimento do resto da natureza; concluiu que apenas experimentando se descobrirá as leis naturais do funcionamento da mente, nomeadamente os princípios da associação das ideias (que incluem a causalidade). Concluiu que o conhecimento é a produção guiada por um ‘instinto indutivo’ e pela imaginação, a partir da experiência, de crenças adaptadas ao mundo exterior, e que transcendem constantemente os dados, aquilo que está imediatamente presente à mente. Concluiu que era a imaginação (e não a razão) que nos permitia dar esses passos para além da experiência<sup>19</sup>.

O que pode ser, nestas circunstâncias, a ciência? A ciência será o progressivo conhecimento do mundo mediante a experiência e dado o funcionamento, regido por leis naturais, da mente humana. Ela deve ser acompanhada de um ceticismo falibilista – esta é a única atitude razoável dadas as conclusões acerca da mente e acções desta, tendo descoberto a fragilidade desse funcionamento. O funcionamento pode ser frágil e daí o desapontamento: a conexão que liga os fenómenos reside em nós; não é razoável desejar conhecer os princípios últimos que residem no mundo exterior, isso seria contradizermo-nos ou falar sem significação. Tudo isto é perturbador, mas a verdade é que não temos nenhuma escolha: a nossa escolha é entre uma falsa razão e nenhuma razão. E esta razão natural

<sup>19</sup> *As relations of ideas, afirmações demonstráveis exclusivamente pela razão e que não dependem da experiência, são uma excepção a esta explicação naturalista. Esta é uma grande diferença relativamente ao empirismo naturalista de filósofos contemporâneos como Quine, que abolem este ‘resguardo’ da lógica e da matemática numa origem a priori. Para Quine, mesmo as verdades destas áreas comparecem ao tribunal da experiência, e como tal podem ‘cair’. Estão aqui em causa, evidentemente, concepções do a priori. Cf. Célia Teixeira 2004, Como não entender o a priori – analiticidade, revisibilidade, necessidade.*

e opaca de que os humanos dispõem não deixa de ser operacional, não deixa de produzir crenças e ciência. E quanto aos desejos de fundamento? E quanto às dúvidas cépticas mais profundas? Aí não haverá saída ou solução. Surge então a célebre declaração: “Felizmente como a razão parece incapaz de afastar estas nuvens, a natureza ela própria será suficiente para tal propósito, e cura-me desta melancolia filosófica e deste delírio (...) Janto, jogo gamão, converso, estou feliz com os meus amigos; e depois de três ou quatro horas de divertimento, voltando a estas especulações, elas parecem-me tão frias (...) e ridículas que não consigo levá-las mais longe“. Mais: ‘se eu devo ser um louco, que as minhas loucuras sejam naturais e agradáveis’, ‘se isto deve ser assim, recomendo a filosofia acima de qualquer superstição’.

Noutras palavras, o cepticismo humeano não é uma abdicação de conhecer nem é um estado em que vivamos em permanência. É qualquer coisa como a condição recomendável do espírito para os assuntos do intelecto e da ciência mas ao mesmo tempo um buraco negro no nosso espírito, que nos atemoriza, e para o qual não podemos estar constantemente ‘acordados’. No entanto, em termos de conhecimento, nomeadamente conhecimento da natureza, as investigações progredirão como tiverem que progredir, uma vez reconhecidas as condições da mente que produz esse conhecimento. É o que temos, é o que usamos.

A investigação céptica sobre a natureza humana traz ainda outros frutos, nomeadamente uma forma de encarar aquilo que tradicionalmente se chamava a metafísica. Para Hume, a metafísica é, na sua maior parte, um conjunto de afirmações que não são fundadas nem em ‘raciocínios acerca da quantidade e do número’ nem em ‘raciocínios experimentais relativos às questões de facto e à existência’. Assim, os livros que contêm essas afirmações ‘deveriam ser lançados às chamas’ (*Investigação sobre o Entendimento Humano*, Secção XII). Se aprendemos alguma coisa com as investigações sobre o entendimento humano, isso deverá fazer-nos limitar as nossas inquirições a objectos que se ajustem optimamente à ‘estreita capacidade do intelecto humano’ (*Investigação sobre o Entendimento Humano*). É dentro deste espírito que deve ser lido o célebre final da *Investigação sobre o Entendimento Humano*: “Ao passarmos os olhos pelas bibliotecas, persuadidos destes princípios, que devastação devemos fazer? Se pegarmos num volume de teologia ou de metafísica escolástica, por exemplo, perguntemos: Contem ele algum raciocínio acerca da quantidade e do número? Não. Contem ele algum raciocínio experimental relativo a questões de facto e de existência? Não. Lançai-o às chamas, porque só pode conter sofisma e ilusão”.

Ter-se-á notado que um tópico essencial para se tratar a natureza do conhecimento em autores como Descartes ou Leibniz esteve totalmente

ausente até aqui em David Hume: esse tópico é Deus. Desde logo, perguntar ‘O que é Deus?’ (uma pergunta sem a qual não se poderia compreender a teoria da mente e do conhecimento de Descartes ou Leibniz) não é uma boa pergunta, uma pergunta bem formulada, aos olhos de Hume. Com uma teoria da mente e do conhecimento como a sua, Hume nunca poderia permitir-se falar directamente, de forma ‘puramente racional’, sobre Deus e sobre as propriedades de Deus, como Descartes ou Leibniz faziam. O que Hume faz é analisar (e afastar) argumentos que supostamente provam a existência de Deus (como o Argumento do Designio), analisar as paixões ligadas à religião, como a superstição e o entusiasmo, e avaliar os seus benefícios e malefícios morais, analisar conceitos religiosos como os conceitos de ‘milagre’ (que rejeita como ilegítimo) e de ‘imortalidade da alma’ (da qual não há qualquer prova), analisar a condenação, pela moral religiosa, de determinados actos como o suicídio (para discordar dessa condenação)<sup>20</sup>.

Deus não é mais, em Hume, uma peça essencial da teoria da mente e do conhecimento – esse é um passo definitivo. Apesar disto, Hume sempre afirmou não ser ateu, mas apenas céptico. Cepticismo é a nota chave e a palavra última da filosofia humeana. Este cepticismo distingue-se quer do cepticismo clássico – que redundava na abdicação da procura do conhecimento – quer do cepticismo cartesiano, que se mantém apenas até ao momento em que se adquire uma certeza racional e a partir daí desaparece. O cepticismo vem para ficar e não impede as investigações cognitivas. Ele não é para Hume um resultado mas sim uma caracterização do estado ‘saúdavel’ de um espírito. Dúvidas cépticas surgem naturalmente a partir da reflexão profunda e intensa; só ‘descuido e a desatenção’ nos reguardarão delas. Mas esta constatação faz-se acompanhar de um cepticismo regrado, não de qualquer lamentação dramática pela impossibilidade de conhecer a realidade em-si. É esse cepticismo que é a melhor protecção contra a metafísica selvagem.

### **Kant: o transcendental como alternativa ao naturalismo humeano**

Ora, como se posiciona Kant perante este cepticismo regrado e moderado, mas nunca vencido racionalmente, de David Hume? Admirando-o, mas não querendo segui-lo demasiadamente. Nomeadamente Kant não quer seguir Hume no apelo que este faz à ‘natureza’ na sua teoria da mente e do conhecimento, i.e. naquilo a que hoje chamariamos o ‘naturalismo’ humeano. O naturalismo, presente por exemplo na solução céptica humeana das dúvidas cépticas, é totalmente estranho a Kant. A alternativa

<sup>20</sup> Cf. Apêndice 2.

kantiana, a grande criação de Kant em filosofia, é a ideia de transcendental. É como base nessa ideia de transcendental que Kant elaborará uma teoria da mente (enquanto realidade ‘composta’ de sensibilidade, entendimento e razão) que lhe permitirá olhar para as relações do pensamento com o mundo de uma forma nova na história da filosofia. Olhemos para uma das mais conhecidas passagens de Kant (e da história da filosofia), a abertura da Conclusão da *Crítica da Razão Prática*: «Duas coisas enchem a mente com admiração e veneração sempre crescentes, quanto mais frequentemente e com maior seriedade delas se ocupa a reflexão: os céus estrelados acima de mim e a lei moral em mim. Eu não conjecturo acerca de nenhuma delas como se fossem obscuridades veladas ou extravagâncias para além do horizonte da minha visão; vejo-as perante mim e ligo-as imediatamente com a consciência da minha existência. A primeira começa no lugar que eu ocupo no mundo exterior dos sentidos e estende a conexão na qual estou à magnitude sem limites de mundos sobre mundos, sistemas sobre sistemas, assim como aos tempos sem fronteira do seu movimento periódico, ao seu princípio e continuação. A segunda começa com o meu eu invisível, a minha personalidade, e mostra-me um mundo que tem verdadeira infinitude, mas que só pode ser detectado pelo pensamento». Esta passagem, acerca da infinitude do mundo e da infinitude da minha liberdade em mim faz alusão a um abismo que Kant se esforçou sempre, com a sua filosofia, por compreender e fechar: o abismo entre Natureza e Liberdade, o abismo entre o mundo concebido como causalidade determinista fechada e a nossa noção de que podemos ser agentes e determinar a forma como as coisas passam a ser. Kant procurou fechar esse abismo com uma teoria da razão humana, uma teoria que deveria ao mesmo tempo explicar a possibilidade de um conhecimento físico-matemático da natureza (a ciência newtoniana) e a possibilidade de liberdade. Isto não significa que Kant tenha tomado como garantidas a liberdade, a imortalidade da alma ou Deus, como o tinham feito filósofos racionalistas seus antecessores, tais como Descartes e Leibniz. Antes, as ‘leis’ dos dois lados do abismo, i.e. as leis da ciência física e as leis da moralidade, o âmbito da ‘Razão teórica’ e o âmbito da ‘Razão prática’, convergiriam, segundo Kant, num sujeito-autor a que chamou sujeito transcendental. A filosofia de Kant é uma exploração desse sujeito transcendental. Como diz Paul Guyer, «Depois de Kant ter escrito nunca mais ninguém pôde pensar quer na ciência quer na moralidade como questões de pura receptividade de uma verdade ou de uma realidade inteiramente externa»<sup>21</sup>. De um modo ou outro, a forma como Kant posicionou as questões da filosofia determinou o rumo do trabalho filosófico futuro.

<sup>21</sup> P. Guyer 1992: 3. Introduction.

O texto de Kant é difícil de ler. O próprio Kant não se considerava um grande escritor<sup>22</sup> e a verdade é que ele não é certamente um Platão ou um Quine, conhecidos pela forma, pela finura, pela beleza ou pela ironia dos seus escritos. Analisar a *Crítica da Razão Pura* nas aulas de epistemologia intimidou muitas levas de estudantes. O que se segue provém dos guiões de leitura *Crítica da Razão Pura* que eram fornecidos procurando obviar a essa reacção e procurando delinear de uma forma tanto quanto possível directa o projecto kantiano de uma filosofia transcendental. Da uma leitura da *Crítica* podem sair respostas a quase todas as questões nucleares tratadas numa introdução à epistemologia, desde questões relativas ao cepticismo, realismo e idealismo a questões relativas à natureza do conhecimento e à natureza do a priori.

O ponto de partida de um percurso poderia ser a leitura os Prefácios A e B<sup>23</sup> da *Crítica da Razão Pura*. A leitura seria guiada pelas seguintes questões: Porque falharam os metafísicos até agora? Que futuro se pode esperar para a metafísica? Trata-se de compreender antes de mais a forma como Kant vê as coisas aqui: Kant não despreza a metafísica, ele pensa antes que a razão humana está destinada a ser atormentada por questões metafísicas, que não há forma de lhes escapar e que é vão afectar indiferença perante elas. No entanto, pensa que a metafísica se esgota em contravérsias e não apresenta nenhum progresso. Na linguagem de Kant, ela não parece ‘ter entrado no caminho seguro da ciência’, ao contrário do que se passa com outros saberes (como a lógica desde Aristóteles, a matemática desde os Gregos, a física desde Galileu). A razão desta situação é o facto de os metafísicos procurarem obter conhecimento de objectos que não são dados na experiência, objectos como Deus e a Alma (os dois objectos das *Meditações* de Descartes, recorde-se). Ora, para a metafísica entrar no caminho seguro da ciência ela não poderá ir simplesmente em busca de tais objectos: ela deverá, segundo Kant, ocupar-se com os objectos da experiência possível. Essa é obviamente uma conclusão insólita dada a tradicional ambição da metafísica, mas o facto é que Kant pensa que a metafísica dita ‘transcendente’, que procura obter conhecimento racional directo de objectos que não são dados na experiência, deverá simplesmente acabar. Como se verá mais à frente, Kant deixará ainda assim aberta a possibilidade de um ‘conceito racional do transcendente’, mas apenas do ponto de vista *prático* (esta palavra em Kant significa frequentemente ‘moral’; nas palavras do próprio Kant, ‘Prático é tudo aquilo que é possível pela liberdade’). O trabalho que permanece para ser feito pelo filósofo consiste

<sup>22</sup> Cf. Kuehn 2001.

<sup>23</sup> A *Crítica da Razão Pura* teve duas edições, em 1781[A] e 1787[B], e a diferença entre estas é objecto de muito debate entre os leitores e intérpretes de Kant.



em investigações sobre a estrutura da razão ela própria. A finalidade desse trabalho é, para Kant, o inventário sistemático daquilo que possuímos pela razão pura e a demarcação do âmbito de legitimidade do uso daquilo que assim possuímos. Esta é a decisão produtiva de Kant, e é trabalhando nas linhas assim traçadas que a *Crítica da Razão Pura* surgirá.

Mas voltando à metafísica: não apenas a metafísica sempre existiu e existirá (Kant fala de uma *metaphysica naturalis*) como a par dela sempre andou e sempre andarà uma dialéctica da razão pura, que lhe é natural. Ora, importa extirpar a influência nefasta desta dialéctica. Para perceber exactamente de que se trata nesta dita ‘dialéctica’ leituras da Dialéctica Transcendental serão úteis (a Dialéctica Transcendental é a segunda divisão da Lógica Transcendental<sup>24</sup>). A questão directora das leituras seria a seguinte: porque é que a metafísica tradicional se apresenta como apresenta? Porque é que os metafísicos fazem o que fazem pensando que podem legitimamente fazê-lo?

Ora, a tese de Kant é que a maioria das doutrinas da metafísica tradicional são falaciosamente obtidas utilizando conceitos do entendimento sem dispôr de evidência da sensibilidade. Este é, para Kant, um ‘uso ilegítimo dos conceitos’. O entendimento e a sensibilidade são faculdades ou aspectos do espírito ou mente respectivamente de espontaneidade e de passividade de que se falará mais à frente. Por agora, para compreender o teor da discussão na Dialéctica, importa compreender que as falácias não acontecem por acaso. Somos para elas empurrados pela inclinação natural da razão, a inclinação para uma completude que nunca pode ser dada na experiência.

Uma vez analisadas as definições de Razão no sentido especificamente kantiano<sup>25</sup> (a ‘faculdade de unificar as regras do entendimento mediante princípios (...) que nunca se dirige imediatamente à experiência nem a nenhum objecto mas tão-só ao entendimento, para conferir ao diverso dos conhecimento essa faculdade uma unidade a priori, graças a conceitos<sup>26</sup>) e de Ideias Transcendentais da Razão<sup>27</sup>, Kant analisa exemplos

<sup>24</sup> Para além do Prefácio e da Introdução, a *Crítica da Razão Pura* divide-se em duas grandes partes: a Parte I, ‘Teoria Transcendental dos Elementos’, cuja primeira parte é a Estética Transcendental e cuja segunda parte é a Lógica Transcendental, e a Parte II, intitulada ‘Doutrina Transcendental do Método’. A primeira divisão da Lógica Transcendental é a Analítica Transcendental enquanto que a segunda divisão é a Dialéctica Transcendental de que aqui se está a falar.

<sup>25</sup> Sentido em que ‘razão’ contrasta com sensibilidade e entendimento. Sensibilidade, entendimento e Razão podem em conjunto ser ditos darem-nos a imagem kantiana da racionalidade humana.

<sup>26</sup> Kant, *Crítica da Razão Pura* [CRP]: 300.

<sup>27</sup> Conceitos que se referem à unidade sintética incondicionada de todas as condições em geral: ‘Por conseguinte, todas as ideias transcendentais podem reduzir-se a três classes, das quais a primeira contém a unidade absoluta (incondicionada) do sujeito pensante, a segunda a unidade absoluta da série das condições do fenómeno e a terceira a unidade absoluta da condição de todos os objectos do pensamento em geral’ (Kant, CRP: 321). Noutras palavras: alma, mundo, Deus.

dos movimentos problemáticos do pensamento com que os metafísicos se comprometem. Os metafísicos têm por uso, por exemplo, tomar a simplicidade lógica do pensamento do eu por conhecimento de uma alma simples, imaterial e imortal. É disso que Kant trata nos Paralogismos da Razão Pura<sup>28</sup>. Pensar, por outro lado, que a ideia de um ‘fundamento de toda a possibilidade’ é equivalente ao conhecimento da existência necessária de tal fundamento é o lance problemático do pensamento (relativo a Deus, obviamente) de que Kant trata no Ideal da Razão Pura<sup>29</sup>. É aqui que se encontra, na *Crítica da Razão Pura*, a análise kantiana das tradicionais provas da existência de Deus, recusando-as uma a uma.

O caso é um pouco diferente com os paradoxos metafísicos que são as Antinomias da Razão Pura<sup>30</sup>: há uma antitética perfeitamente natural da razão em questões cosmológicas<sup>31</sup>. Há quatro antinomias (‘pares’ tese/antítese) relativas ao mundo como conjunto de todos os fenômenos<sup>32</sup>:

**Primeira Antinomia** O mundo tem um começo no tempo e é também limitado no espaço / O mundo não tem começo nem limites no espaço; é infinito tanto no tempo como no espaço

**Segunda Antinomia** Toda a substância composta, no mundo, é constituída por partes simples e não existe nada mais que o simples ou o composto pelo simples / Nenhuma coisa composta, no mundo, é constituída por partes simples, nem no mundo existe nada que seja simples

**Terceira Antinomia** A causalidade segundo as leis da natureza não é a única de onde podem ser derivados os fenômenos do mundo no seu conjunto. Há ainda uma causalidade pela liberdade que é necessário admitir para os explicar / Não há liberdade, (...) tudo no mundo acontece unicamente em virtude das leis da natureza.

**Quarta Antinomia** Ao mundo pertence qualquer coisa que, seja como sua parte seja como sua causa, é um ser absolutamente necessário / Não há em parte alguma um ser absolutamente necessário, nem no mundo nem fora do mundo, que seja a sua causa.

Kant considera que as duas primeiras antinomias são afastadas reconhecendo o espaço e o tempo como formas da nossa intuição (tratará esta

<sup>28</sup> Kant chama ‘paralogismo’ à ‘falsidade de um raciocínio quanto à forma, seja qual for o seu conteúdo’ (Kant CRP: 327).

<sup>29</sup> Kant, CRP: 485 e segs.

<sup>30</sup> Kant, CRP: 379.

<sup>31</sup> Kant chama conceitos cosmológicos a ‘todas as ideias transcendentais, na medida em que se referem à totalidade absoluta na síntese dos fenômenos’ (Kant, CRP: 380).

<sup>32</sup> Cf. Kant CRP, pp. 392-393, 398-399, 406-407, 412-413.

proposta na Estética Transcendental), não sendo as coisas em si nem infinitas nem finitas temporalmente e espacialmente. As duas restantes antinomias não estão ligadas da mesma forma à estrutura da sensibilidade humana. O que Kant propõe é que a tese e a antítese podem, nestes casos, ser ambas verdadeiras, uma vez aceite a distinção entre fenómeno e númeno, distinção esta avançada na Analítica Transcendental. O mundo fenoménico poderá assim ser concebido como um domínio de existências contingentes governado por leis naturais enquanto que no mundo numérico não apenas existe um ser necessário como também existem ações livres. Supostamente esses mundos são o mesmo mundo: *este mundo*. Isto significa evidentemente que a crítica que Kant faz da metafísica tradicional na *Crítica da Razão Pura* deixa (afinal) abertas as possibilidades de liberdade e de Deus. Aqui se enraizará a ‘metafísica moral’ caracteristicamente kantiana. Mas o outro aspecto (prévio, preparatório) da ‘metafísica’, e o mais importante para a história da epistemologia, é uma teoria do conhecimento e da realidade conhecida cuja exposição ocupa a quase totalidade da obra. E é isso que interessa compreender, visando essencialmente a forma como Kant avança uma proposta acerca da mente e do conhecimento que é uma doutrina acerca da mente como sensibilidade e do entendimento que conjuga receptividade e espontaneidade, intuitividade e discursividade.

No percurso de leituras, se se passar à Introdução (B), encontrar-se-á que aí são fixados os pontos básicos na visão kantiana: todo o conhecimento começa com a experiência (i.e. nada na mente precede a experiência), mas isso não significa que todo o conhecimento derive da experiência. O conhecimento por experiência é um composto do que recebemos e daquilo que a nossa capacidade de conhecer produz por si mesmo (acrescente-se que nós não distinguimos usualmente esse acréscimo da matéria prima). Noutras palavras, existe um a priori, de acordo com Kant, que se distingue do que é empírico, i.e. a posteriori, ou de experiência. Quais serão as marcas da existência desse a priori? Para Kant, a necessidade e a universalidade (inseparáveis uma da outra) são os sinais seguros do conhecimento a priori. De facto, e isso é significativo, e é essa a pista para Kant, nós dispomos já de conhecimentos que saem do campo das experiências possíveis. O problema é saber quanto longe se pode legitimamente ir.

No foco de toda esta análise levada a cabo por Kant está uma teoria do juízo. De acordo com Kant, é possível distinguir juízos analíticos (explicativos, que não alargam o conhecimento que já possuímos) de juízos sintéticos (extensivos, que alargam o conhecimento). Os juízos de experiência são todos sintéticos. Ora, o que é importante para uma teoria do conhecimento científico da natureza, que é parte do que Kant faz na *Crítica da Razão Pura*, é saber se pode haver juízos que possuam ao mesmo

tempo as características da universalidade e necessidade (que assinalam usualmente os juízos analíticos) e a extensividade (característica dos juízos sintéticos). Esses juízos, a existirem, serão os juízos sintéticos a priori. A tese defendida na *Crítica da Razão Pura* é que tais juízos são possíveis, existem, e são aliás eles que tornam possível a ciência. Os juízos matemáticos e físicos são exemplos de juízos sintéticos a priori (por exemplo, um juízo aritmético como  $'7+5=12'$  é para Kant um juízo sintético a priori). Os juízos da metafísica também deveriam ser juízos sintéticos a priori. Temos assim delimitado o problema geral da Razão Pura: ele é expresso pela questão 'Como são possíveis os juízos sintéticos a priori?'

Kant parte dos factos: as ciências constituídas por juízos desse tipo estão dadas e portanto que são possíveis demonstra-o a sua realidade. O problema é saber como é que essas ciências, que existem, são possíveis. Nasce assim a ideia de uma *Crítica da Razão Pura* (uma ciência particular da faculdade que nos fornece os princípios do conhecimento a priori<sup>33</sup>), e o objectivo de um 'conhecimento transcendental' (todo o conhecimento que em geral se ocupa menos dos objectos do que do nosso modo de os conhecer, na medida em que este deve ser possível a priori<sup>34</sup>).

Começa então a exposição da teoria kantiana da mente e do conhecimento. A Estética Transcendental, 'ciência de todos os princípios a priori da sensibilidade' é a parte da *Crítica da Razão Pura* em que se analisa a nossa mente tanto quanto esta é receptiva. Kant chama sensibilidade à capacidade de receber representações (receptividade), graças à maneira como somos afectados pelos objectos. Chama sensação ao efeito de um objecto sobre a capacidade representativa. Chama intuição empírica àquela que se relaciona com o objecto por meio de sensações e chama intuição *tout court* à relação imediata do conhecimento com objectos (considera, de resto, que a intuição é o fim para o qual tende o pensamento). É deste conjunto de definições que 'sai' a noção de fenómeno: fenómeno, diz Kant, é o objecto indeterminado de uma intuição empírica. Matéria é o que no fenómeno corresponde à sensação, forma aquilo que possibilita que o diverso do fenómeno possa ser ordenado segundo determinadas relações. O espaço e o tempo, a que Kant chamará intuições puras são as formas puras da sensibilidade, que se encontra a priori no espírito (nosso, humano)<sup>35</sup>. Espaço e tempo não são 'conceitos empíricos que derivem de uma experiência qualquer', nem conceitos discursivos ou universais, que subsumam instâncias. Eles não 'vêm ao espírito' nem são formas de este 'classificar'

<sup>33</sup> Kant, CRP: 52.

<sup>34</sup> Kant, CRP: 53.

<sup>35</sup> As grandes alternativas aqui seriam as concepções newtoniana e leibniziana de espaço de tempo. No primeiro caso trata-se de referenciais absolutos, no segundo caso de 'relações relativas'.

o que há<sup>36</sup> – eles estão no espírito e tudo o que for humanamente intuível será ‘especializado’ e ‘temporalizado’. Seguem-se, na *Crítica*, as chamadas Exposição Metafísica e Exposição Transcendental do espaço e do tempo para discutir estas teses e outras, e a Estética Transcendental conclui com a distinção entre fenómeno e coisa-em-si<sup>37</sup>. A ideia básica é que toda a nossa intuição não é mais do que a representação do fenómeno, no sentido em que as coisas não são em si mesmas tal como nos aparecem: as coisas aparecem à nossa subjectividade, se fizéssemos abstracção da subjectividade, as relações dos objectos no espaço e no tempo, nomeadamente, desapareceriam. Fenómenos não existem em si mas unicamente em nós; é-nos desconhecido o que são os objectos em si, conhecemos apenas o nosso modo de os perceber mas nada nos diz sequer que esse seja o modo de perceber de todos os seres, de todas as mente possíveis.

O que nos interessa em seguida para compreender a teoria kantiana da mente e do conhecimento é a Analítica Transcendental, que Kant define como uma ‘decomposição de todo o nosso conhecimento a priori nos elementos do conhecimento puro do entendimento’. Para se chegar à Analítica, que é, como já foi dito, uma parte da Lógica Transcendental, é preciso compreender a distinção entre Lógica geral e Lógica Transcendental<sup>38</sup>, uma distinção de resto interessante para pensarmos, mesmo hoje, nas fronteiras e nas relações entre lógica e epistemologia. A lógica geral, segundo Kant, abstrai totalmente do conteúdo do conhecimento, é pura, não tem princípios empíricos, nada vai buscar à psicologia, considera apenas ‘a forma do pensamento em geral’. A lógica transcendental, por contraste, ocupa-se da origem dos nossos conhecimentos (‘na medida em que tal origem não pode ser atribuída aos objectos’, diz Kant). É essa que interessa ao epistemólogo, por contraste com o lógico. Na lógica transcendental isola-se o entendimento (a faculdade não sensível do conhecimento, uma faculdade de conhecer por conceitos, e não por intuição). O conhecimento produzido pelo entendimento humano é discursivo e mediato segundo Kant, é conhecimento por conceitos e juízos. A mediação é feita por conceitos, que assentam em funções, diz Kant, por contraste com as intuições que assentam em afecções. Conceitos nunca se referem imediatamente a objectos, mas apenas a outras representações, sejam estas intuições, sejam já outros conceitos. É esse o estatuto dos juízos: eles são conhecimento mediato de objectos, representação de uma representação do objecto<sup>39</sup>.

<sup>36</sup> Isto é o que faz um conceito empírico como ‘cão’.

<sup>37</sup> Kant, *CRP*: 78.

<sup>38</sup> Kant, *CRP*: 88-92.

<sup>39</sup> Kant, *CRP*: 102.

A ideia fundamental de Kant é que se encontrarão todas as funções do entendimento se se puder expôr totalmente as funções de unidade nos juízos. As categorias (Kant usa conscientemente o termo aristotélico<sup>40</sup>), ou conceitos puros, são as ‘responsáveis’ pela síntese nos juízos. A síntese pura, diz Kant, ‘dá’-nos o conceito puro do entendimento. Acerca de síntese, uma noção fundamental em todo este quadro, Kant diz ainda: ‘entendo por síntese (...) o acto de juntar umas às outras diversas representações e conceber a sua diversidade num conhecimento’<sup>41</sup>. O entendimento é a capacidade de julgar e o juízo pressupõe síntese. Quer a dedução metafísica das categorias quer a dedução transcendental lidarão com esta questão da ‘síntese’ operada pelo entendimento. As tabelas dos juízos e das categorias podem por isso ser vistas da seguinte forma. A tabela dos juízos dá-nos as sínteses que a mente faz: recolhemos isso ao mapear os procedimentos do espírito nas suas actividades<sup>42</sup>. A tabela das categorias dá-nos o que tem que existir a priori no espírito humano, de acordo com Kant, para que mentes como as nossas possam fazer os juízos que fazem acerca da realidade<sup>43</sup>. Note-se que na tabela das categorias se encontrará formas de pensarmos nas coisas tão básicas e fundamentais como a unidade, ou a totalidade, a relação substância-acidente, a causalidade e as modalidades (possibilidade-impossibilidade; existência-não existência; necessidade-contingência).

Mas voltemos à ideia kantiana de categoria. O que são exactamente as categorias ou conceitos puros do entendimento, sabendo desde já que este é um conceito-chave para compreender a forma como Kant vê a espontaneidade do espírito ou mente? As categorias ou conceitos puros do entendimento são, por um lado, traços gerais de qualquer conceito determinado de um objecto, que permitem a aplicação de um juízo a esse objecto. Por exemplo no caso de um juízo de forma ‘S é P’, alguma coisa no objecto deve ser representada como substância, correspondendo ao conceito do sujeito e alguma outra coisa como a propriedade, correspondendo ao conceito do predicado no juízo. Mas as categorias são também, por outro lado, acções universais do espírito ou mente, acções por

<sup>40</sup> Mas considera ser mais sistemático e exaustivo do que este na listagem das categorias, na medida em que a sua lista de categorias é gerada por um princípio, a origem no entendimento como capacidade de julgar. A lista aristotélica seria, em contraste, ‘rapsódica’. As categorias aristotélicas eram além do mais, como é sabido, categorias do ser. O que estava em causa eram portanto formas de ser das coisas tais como estas são em si mesmas. Não é assim com as categorias kantianas.

<sup>41</sup> Kant, CRP: 109.

<sup>42</sup> Temos assim juízos universais, particulares e singulares [Quantidade], afirmativos, negativos e infinitos [Qualidade], categóricos, hipotéticos e disjuntivos [Relação], problemáticos, assertóricos e apodicticos [Modalidade].

<sup>43</sup> Temos assim unidade, pluralidade, totalidade [Qualidade], realidade, negação, limitação [Qualidade], inexistência e subsistência, causalidade e dependência, comunidade [Relação], possibilidade-impossibilidade, existência-não existência, necessidade-contingência [Modalidade].

meio das quais podemos pensar um objecto em geral, e portanto, segundo Kant, conteúdo transcendental ‘trazido’ às representações. As categorias não apenas tornam possível a aplicação de juízos àquilo que é dado, que só pode ser dado na intuição, mas constroem de forma determinada essa aplicação. I.e., não se trata, para Kant, de qualquer coisa como uma mera adição de informação ao dado da intuição, mas algo de constitutivo, formatador. Assim sendo, é preciso encontrar uma fonte do constrangimento, e explicá-la, expô-la, analisá-la, e mostrar que as categorias se aplicam a todos os objectos e não apenas a alguns. Este é o propósito da Dedução Transcendental das categorias, na qual se discute a relação das categorias com a auto-consciência (dita ‘Unidade Transcendental da Apercepção’).

Há muito a dizer acerca desse ‘coração da filosofia kantiana’ que é a Dedução Transcendental das Categorias na Analítica Transcendental (em A e B). No Prefácio à primeira edição da *Crítica* (A, 1781), Kant afirma que ‘não conhece investigações mais importantes para chegar ao fundo da faculdade a que chamamos entendimento (...) do que aquelas que empreendeu na Segunda Parte da Analítica Transcendental sob o título de Dedução dos Conceitos Puros do Entendimento’. Reconhecerá no entanto que ‘essa parte da *Crítica* que deveria ter sido a mais clara é a mais obscura, ou mesmo move-se em círculo’ (afirma isto nos *Fundamentos Metafísicos da Ciência Natural*, 1786). Na segunda edição da *Crítica*, em 1787, o texto da Dedução Transcendental será objecto de profundas modificações<sup>44</sup>.

É importante tentar compreender qual é exactamente a tese que Kant tenta provar na Dedução Transcendental, quer em A quer em B, e também compreender se, ao contrário do que o próprio declarou, as alterações não foram apenas de apresentação<sup>45</sup>. Mas como deve isso fazer-se? Antes de mais parece haver uma intenção razoavelmente clara que persiste de A para B: trata-se de procurar provar a *conexão inseparável entre a auto-consciência e o conhecimento de objectos*, impedindo assim qualquer cartesianismo (entendido como afirmação da possibilidade de conhecimento de estados internos independentemente do conhecimento do mundo exterior). Procura-se ao mesmo tempo provar a *validade objec-*

<sup>44</sup> As edições actuais da *Crítica* mantêm os dois textos.

<sup>45</sup> Muitos foram os leitores da *Crítica* que consideraram que as alterações não são de todo indiferentes. Evidentemente joga-se aqui também a forma como o propósito da *Crítica* é visto: entre os neo-kantianos (por exemplo H. Cohen), para quem esse propósito é explicar as pressuposições da ciência da natureza, e Heidegger, para quem o propósito da obra é questionar a natureza e possibilidade da metafísica, expondo o conhecimento ontológico (conhecimento do ser) pressuposto pelo conhecimento óntico (conhecimento de seres particulares), as diferenças serão abissais. Heidegger (cf. *Interpretação fenomenológica da Crítica da Razão Pura*, publicada em 1977 e *Kant e o problema da metafísica*, 1929) interessou-se particularmente pela ‘raiz comum’ da unidade das intuições do espaço e do tempo e da unidade dos conceitos nos juízos. Essa raiz comum seria a síntese da imaginação, pela qual uma visão dos humanos e outras entidades como sendo fundamentalmente temporais é produzida. Devido a este foco na imaginação e no tempo, o interesse de Heidegger pela *Crítica* concentra-se na dedução transcendental e no esquematismo dos conceitos puros do entendimento.

*tiva dos conceitos puros do entendimento.* Na medida em que uma dedução procura estabelecer uma legitimidade, um *quid juris*, a legitimidade que está aqui em causa é a da aplicação das categorias, condições subjectivas do pensamento, a priori a todos os objectos, tendo assim validade objectiva ('objectos' devem aqui ser entendidos por contraste com representações meramente subjectivas, modificações do espírito). A proposta de Kant passa pela Unidade Transcendental da Apercepção, ou auto-consciência, o 'eu penso que acompanha todas as minhas representações'. Kant pensa que condições subjectivas do pensamento (as categorias) podem ter validade objectiva sendo condições necessárias da própria apercepção, i.e. da auto-consciência, e portanto, *a fortiori*, condições para a representação de quaisquer objectos no *medium* de estados subjectivos dos quais somos conscientes.

Noutras palavras: Kant vai tentar argumentar que o conhecimento de objectos (que existe, que nós possuímos) não pode existir sem a a reprodução de uma multiplicidade de aparências dadas na mente ou espírito e que essa reprodução só pode ter o seu fundamento numa consciência da representação de uma identidade numérica necessária (do eu) ao longo das suas várias representações. Este eu 'numericamente idêntico não pode ser pensado por meio de dados empíricos', por isso Kant sugere que existe aí uma 'Unidade Transcendental da Apercepção' (UTA) ou auto-consciência caracterizada por espontaneidade. As condições para a unidade da auto-consciência são condições necessárias para o conhecimento de objectos. Compreender a natureza e o estatuto desta Unidade Transcendental da Apercepção é obviamente um dos problemas que daqui saem. No núcleo das alterações da apresentação da dedução transcendental de A para B estarão questões relativas à Unidade Transcendental da Apercepção, à espontaneidade desta, ao tempo e à imaginação.

Procurar-se-á em seguida esquematicamente caracterizar o que é dito na dedução transcendental na primeira e na segunda edições da *Crítica*, deixando em aberto a questão de saber se houve ou não profundas alterações.

Um resumo de alguns pontos centrais da argumentação de Kant em A seria algo como o seguinte: (1) Todas as minhas representações são por mim necessariamente reconhecidas como minhas. Tenho pois conhecimento *a priori* de que pertencem a um eu numericamente idêntico (mas o que é isso de as minhas representações serem minhas e de pertencerem a um eu numericamente idêntico?); (2) A Unidade Transcendental da Apercepção (UTA), a auto-consciência, é não apenas uma unidade analítica mas uma unidade sintética (i.e. as minhas representações não têm nenhuma



marca comum pela qual seriam analiticamente reconhecidas como minhas; elas são antes reconhecidas como minhas com base numa conexão entre elas – e é essa a marca comum) (o que será essa conexão?); (3) Se a UTA implica uma conexão sintética das representações independente do seu conteúdo empírico tem que haver uma síntese *a priori*, com as suas regras *a priori*; (4) A fonte da conexão é a faculdade do entendimento; as regras *a priori* requeridas para a síntese *a priori* da UTA são as categorias; (5) As condições necessárias para a síntese de todas as representações na UTA são as condições necessárias para a representação dos objectos quaisquer.

Um aspecto de tudo isto é crucial na primeira edição e encontra-se logo no início do texto de Kant: todas as minhas representações, de outros objectos ou de estados internos, pertencem ao sentido interno, logo estão ‘submetidas’ ao tempo como forma. A única maneira de representar fenómenos é representar uma unidade na multiplicidade, ora para representar uma diversidade é preciso distinguir o tempo na série das representações, i.e. representar as representações como ocorrendo em momentos sucessivos a ainda assim de alguma forma ‘manter a pista’. Para isso há, segundo Kant, uma ‘tripla síntese’ cujo princípio é a espontaneidade (este assunto é tratado em Da Síntese da Apreensão na Intuição; Da Síntese da Reprodução na Imaginação; Da Síntese da Reconhecimento no Conceito). Isto significa que itens numa multiplicidade temporalmente sucessiva têm que ser sucessivamente *apreendidos*. Itens previamente apreendidos têm que ser *reproduzidos* conjuntamente com os posteriores para as conexões entre eles serem *reconhecidas*. Ora, tem que haver algum conceito em virtude do qual a conexão de várias representações sucessivas seja reconhecida como representação-do-mesmo-objecto (‘sem a consciência de que aquilo que nós pensamos é precisamente o mesmo que pensávamos no instante anterior, seria vã toda a reprodução na série de representações’, diz Kant). Tem que haver um princípio da unidade da consciência na síntese do diverso, e é a isso que Kant chamará a espontaneidade da Unidade Transcendental da Apercepção (por contraste com a apercepção empírica, sempre mutável, como Hume viu), cuja ligação com a imaginação e o tempo Kant enfatizava na primeira versão.

E quanto à edição de 1787? O que mudou? O que é que Kant considerou que tinha que aperfeiçoar? Uma hipótese é que Kant trabalha com uma concepção agora mais sofisticada de juízo<sup>46</sup>. Um possível resumo da argumentação em B seria algo como o seguinte. Começa-se (§15) com a ideia de que a ligação ou conjunção do diverso é um acto de espontaneidade ou actividade do sujeito: Kant sugere várias vezes que a validade objectiva

<sup>46</sup> Guyer 1992, Introduction: 49.

das categorias na síntese poderia ser derivada desta simples premissa. Mas Kant quer argumentar mais: quer argumentar especificamente que as combinações pressupõem a forma de síntese da Unidade Transcendental da Apercepção. Assim Kant reitera (§ 16) o que em A fora dito sobre a apercepção (i.e. que todas as representações têm uma relação com o *eu penso*). Afirma que a conexão das impressões no eu não pode ser meramente analítica, pois não há nenhuma impressão única do eu em todas as possíveis representações (a consciência empírica é em si diversa, como Hume bem vira). A unidade analítica da auto-consciência pressupõe uma unidade sintética. Tem que haver pois uma unidade sintética da apercepção, já que tenho conhecimento a priori genuíno da conexão necessária de todas as representações como minhas. Em seguida (§ 17-19), procura mostrar a conexão entre a apercepção e o conhecimento de objectos, chamando objecto ‘aquilo em cujo conceito está reunido o diverso de uma intuição dada’. Toda a reunião exige a unidade da consciência. Fala como se a cognição de objectos fosse uma condição necessária da unidade da apercepção, e como se as condições a priori do unidade da apercepção pudessem ser derivadas das condições para o conhecimento de objectos (e não vice versa) (§18). Seguidamente (§19), afirma-se insatisfeito com a explicação que os lógicos dão dos juízos, considerando-os como representações de relações de conceitos sem explicar em que consiste a relação. Para se pensar na natureza dos juízos é necessário, segundo Kant, considerar que eles trazem ou conduzem algo à unidade objectiva da apercepção. Em seguida contrasta os juízos como asserções de unidade necessária com as relações de representações que são válidas de forma meramente subjectiva, e afirma que tem que haver fundamentos a priori no entendimento para tal unidade necessária. Afirma então (§20) que a diversidade dada na intuição sensível é sujeita à Unidade Sintética da Apercepção e o que acto do entendimento pelo qual a multiplicidade é trazida para debaixo da apercepção é a função lógica do juízo. Portanto, toda a multiplicidade é determinada em relação às funções lógicas do juízo. Ora, as categorias não são senão as funções lógicas do juízo. Toda a multiplicidade é assim determinada em relação às categorias.

A seguir (§21) afirma diz que o argumento precedente foi apenas um começo da dedução das categorias. Apenas se mostrou que as categorias são condição necessária da consciência empírica de uma dada multiplicidade de uma intuição, não se mostrou ainda a sua validade objectiva. O argumento completo irá envolver (§26) o reconhecimento de que a unidade do espaço e do tempo requerem elas próprias uma síntese do entendimento. A unidade do espaço e do tempo, que na Estética Transcendental era tida como dada, vai portanto ser mostrada como resultado da actividade

do entendimento. E como nada nos pode ser dado a não ser espacializado ou temporalizado ou ambos, nada nos pode ser dado que não seja sujeito à actividade do entendimento e portanto às categorias. Portanto, toda a síntese só é possível sob as categorias.

As dificuldades desta argumentação (se é que devemos chamá-la assim) são muitas. Enumero algumas: por exemplo, como é que Kant justifica a pretensão de que a Unidade Transcendental da Apercepção é uma certeza *a priori* da identidade numérica do eu, uma certeza que requer a síntese de representações de acordo com regras *a priori*? Como é que ‘superou’ Hume, se acerca do eu empírico faz afinal afirmações semelhantes às que Hume fizera? Como e porquê são estas regras identificadas com as categorias? De onde vem o número de categorias? Porque é que as categorias são doze? Porquê estas? Qual é a diferença, por exemplo, entre a ‘realidade’ como categoria da Qualidade e a ‘existência’ como categoria da Modalidade? Não serão antes sinónimos? O próprio Kant por vezes enumera apenas cinco, quatro, ou só três categorias. Quando reduz a lista a três ‘categorias de síntese’ aquelas que ficam são as três categorias da relação (substância, causalidade, comunidade). Apenas estas formas, em virtude das quais empregamos juízos categóricos, hipotéticos e disjuntivos, seriam então realmente distintas. Elas correspondem a tipos de juízos diferentes e não apenas a diferentes valores de um tipo de juízo, e seriam portanto absolutamente essenciais ao que Kant quer demonstrar. O mesmo não se passa no entanto com todas as categorias.

As dúvidas podem continuar: exactamente qual é a conexão entre a Unidade Transcendental da Apercepção e os objectos da cognição de modo a que estes sejam distintos da nossa representação (isto é, não colapsem nela)? Será que Kant realmente prova que a Unidade Transcendental da Apercepção requer conhecimento de objectos distintos do eu? Haverá algum espaço lógico nesta teoria para a distinção entre meras representações e verdadeiro conhecimento de objectos?

O próprio Kant vê e procura evitar alguns perigos que ameaçam a sua posição, por exemplo nas Analogias da Experiência (Analítica dos Princípios), onde mostra que os juízos acerca de relações temporais de estados de objectos só podem ser feitos contrastando-os com relações temporais de estados meramente subjectivos pelo uso de categorias como substância e causação. Também na Refutação do Idealismo (Analítica Transcendental, Analítica dos Princípios) Kant pretende mostrar que mesmo os juízos acerca de relações temporais em estados meramente subjectivos requerem a sua correlação mas não identificação com estados objectivos subsumidos às categorias (i.e., quer mostrar que a auto-consciência requer o conhecimento de objectos sem por isso cair na identificação com o conhecimento

de objectos). Em suma, Kant vê os perigos como perigos, e a ser assim, de facto, formalmente, a Dedução Transcendental falha no estabelecimento da sua tese; a tese é retomada em várias outras partes da *Crítica da Razão Pura*, que devem ser consideradas em conjunto na avaliação do que está em causa.

Com todos os problemas que possam permanecer, estas são as grandes linhas da teoria kantiana da mente e do conhecimento. De acordo com elas, aquilo a que chamamos mente ou espírito é um composto de receptividade, espontaneidade e totalizações heurísticas (as ideias da razão), todas podendo, de resto, ser concebidas como ‘sínteses’. Ora como se afirmou no início deste trajecto pela *Crítica*, o propósito central de Kant não era fazer esta teoria da mente ou da razão por si mas sim ‘salvar a metafísica’. Como ficamos quanto a esse propósito? Algo já foi avançado quanto à ‘metafísica moral’ de Kant. Com a intenção de avançar um pouco por esse caminho, que excede a epistemologia, poder-se-á tomar as seguintes questões directoras: em quantos mundos vivemos afinal? Como é que a teoria apresentada nos deixa, a nós próprios, relativamente ao conhecimento do mundo como fenómeno de que somos capazes? Elas permitir-nos-ão falar da visão kantiana da liberdade: afinal aquele mundo cujo conhecimento científico, através de juízos sintéticos a priori, é possível, é o mesmo mundo em que eu existo como agente livre. Mas como é isso possível?

A questão da liberdade é tratada em vários pontos da *Crítica*, nomeadamente, como já foi indicado atrás, na Dialéctica Transcendental (Antinomias da Razão Pura: Terceiro Conflito e Solução das ideias cosmológicas que dizem respeito à totalidade da derivação dos acontecimentos do mundo a partir das suas causas (Nona secção, III)) e também na Doutrina Transcendental do Método (Capítulo II, O cânone da razão pura, Segunda secção, Do Ideal do Sumo Bem como fundamento determinante do fim último da razão pura). A ideia fundamental de Kant é que se os fenómenos fossem coisas em si não seria possível salvar a liberdade. Mas os fenómenos não são coisas em si, e Kant procurará salvar a liberdade enquanto causalidade pela razão. O ponto de partida é o seguinte: ‘Por um lado, ele mesmo (o homem) é, sem dúvida fenómeno, mas do ponto de vista de certas faculdades (razão, entendimento) é também um objecto meramente inteligível, porque a sua acção não pode atribuir-se à receptividade da sensibilidade (...) E que a razão possua uma causalidade é o que ressalta dos imperativos que impomos como regras em toda a ordem prática, às faculdades activas” (CRP, Dialéctica Transcendental, A Antinomia da Razão Pura, Nona Secção, III, Esclarecimento da Ideia Cosmológica de uma Liberdade em União com a Necessidade Universal da Natureza).

Na Dialéctica Transcendental Kant não pretende expôr a realidade da liberdade, nem demonstrar a possibilidade da liberdade. Pretende apenas tratar a liberdade como uma Ideia Transcendental. Que ideia é essa? É a ideia pela qual ‘a razão pensa iniciar absolutamente, pelo incondicionado do ponto de vista sensível, a série das condições no fenómeno, enredando-se assim numa antinomia com as próprias leis que prescreve ao uso do entendimento’ (Esclarecimento da Ideia Cosmológica de uma liberdade em união com a necessidade universal da natureza). Nesse ponto da *Crítica*, Kant pensa ter mostrado que essa antinomia assenta em mera aparência e que a natureza não está em conflito com a causalidade pela liberdade.

O ponto de partida de Kant para a construção teórica substancial em torno da liberdade será a seguinte constatação: o dever, i.e. o dever-ser, é qualquer coisa que não ocorre em qualquer outra parte da natureza que não no humano. O dever é qualquer coisa de diferente daquilo que o entendimento pode conhecer. O entendimento só pode pensar acerca da natureza o que esta é, foi, ou será. O dever, em contrapartida, ao contrário de uma mera acção da natureza cujo fundamento terá que ser sempre um fenómeno, ‘exprime uma acção possível cujo fundamento é um simples conceito’<sup>47</sup>. Sendo a a Liberdade uma Ideia da razão especulativa, ela permaneceria puramente problemática, pensa Kant, se o facto da lei moral não nos ensinasse que somos livres. Mas esse facto impõe-se-nos. Kant considera então que a razão é a condição de todas as acções voluntárias pelas quais o homem se manifesta. A razão pura age livremente na cadeia das causas naturais. A liberdade não se deve considerar apenas negativamente, como independência das condições empíricas, mas também positivamente, como faculdade de iniciar por si própria uma série de acontecimentos. A razão pura é, para Kant, a ‘condição incondicionada de toda a acção voluntária’.

Tudo isto conduz, como se sabe, mas isso excede um curso de epistemologia, à importância do uso prático da Razão. Kant defende que ao relativo insucesso da filosofia da razão pura no seu uso especulativo se pode opôr o sucesso num outro caminho: o caminho do uso prático da razão pura. É a este que diz respeito o Cânone da Razão Pura (o Capítulo II da Doutrina Transcendental do Método). Kant expõe a sua concepção do ‘propósito final da especulação da razão’. Este diz respeito a três objectos: a liberdade da vontade, a imortalidade da alma e a existência de Deus. Não há ‘uso imanente’ para estes três objectos: pensá-los seria um esforço ocioso se não fosse o nosso fortíssimo interesse prático. Em suma, ‘A razão, no seu uso especulativo, conduziu-nos através do campo da experiência e,

<sup>47</sup> Kant, *CRP*: 472.

como neste nunca pode encontrar satisfação completa, levou-nos daí às ideias especulativas que, por sua vez, nos trouxeram de novo à experiência e assim cumpriram a sua intenção, de uma maneira útil, é certo, mas nada de acordo com a nossa expectativa<sup>48</sup>.

Na parte da Crítica intitulada ‘Do Ideal do Sumo Bem como um fundamento determinante do fim último da razão pura’ Kant expõe aquilo que pensa encontrar neste caminho. Procura fazê-lo respondendo à questão “O que me é permitido esperar?”. Dar resposta a esta questão impõe que se relacione a liberdade com a moralidade e a felicidade. A relação é feita através de Deus: o Ideal do Sumo Bem é a ideia de uma inteligência que comanda o mundo de acordo com leis morais e que ‘reúne’ assim a ‘esperança de ser feliz’ com o ‘esforço incessante de se tornar digno da felicidade’. O mundo inteligível, que é um mundo moral, do qual devemos pensar fazer parte enquanto seres morais e livres embora os sentidos só nos apresentem um mundo fenoménico, só pode ser um mundo futuro para nós. Deus e uma vida futura são, portanto, pressupostos, segundo Kant, da obrigação que o uso de uma razão livre nos impõe. Se nada disto fôr o caso, as leis morais serão vãs quimeras. Kant chama a este seu discurso ‘teologia moral’.

No núcleo do pensamento moral de Kant está então a afirmação da autonomia: os seres humanos são autónomos, são eles próprios a origem da lei pela qual se regem e isto é assim mesmo se eles são também, seres físicos num mundo determinista. A vontade dos seres humanos, enquanto razão prática, não é (já) desejo. Isto significa que Kant pensa que a forma superior do querer é descontínua em relação ao desejo sensível e que é o ‘lugar de uma legislação universal’. Na sua formulação, a lei moral ordena-nos que pensemos a máxima que rege a nossa vontade como ‘princípio de uma legislação universal’. Uma vontade determinada pela pura forma da lei, independentemente da condição sensível, é aquilo a que Kant chama uma vontade livre. A legislação da qual os seres humanos são autores incide sobre a causalidade de seres livres (aquilo pelo qual seres livres são causa de alguma coisa). A liberdade, por oposição à causalidade natural que rege os fenómenos, define-se então como o ‘poder de começar por si mesmo um estado cuja causalidade não se situa por seu turno (como na lei natural) sob outra causa que a determina quanto ao tempo’ (Dialéctica Transcendental, Solução das ideias cosmológicas que dizem respeito à totalidade da derivação dos acontecimentos do mundo a partir das suas causas<sup>49</sup>).

<sup>48</sup> Kant, CRP: 639 [Do Ideal do Sumo Bem].

<sup>49</sup> Kant, CRP: 462 e seguintes.

A liberdade é a ideia, criada pela razão, de uma espontaneidade que poderia começar a agir por si mesma, sem que uma outra causa a devesse ter precedido. Esta é a ‘liberdade transcendental’, sobre a qual assenta a ‘liberdade prática’; a liberdade no sentido prático é a independência do arbítrio face à coacção dos impulsos da sensibilidade. A ideia de liberdade goza de um privilégio em relação às outras ideias da razão (imortalidade, Deus) pois o facto de nós nos podermos determinar praticamente ‘dá às coisas em si o sentido ou a garantia de um facto’, ‘faz(nos) efectivamente penetrar no mundo inteligível’ - é o que afirma Kant no Parágrafo 91 da *Crítica da Faculdade de Julgar*.

Pode-se ler a *Crítica da Razão Pura* durante anos, e na verdade foi isso que muitos filósofos posteriores a Kant fizeram. Esta foi apenas uma sugestão de trajecto. No contexto de uma iniciação à epistemologia a grande importância de Kant reside na alternativa que o transcendentalismo constitui à proposta quineana de uma epistemologia naturalizada. Tracei acima as linhas principais desse transcendentalismo. Restam obviamente as questões: terá de facto Kant respondido à acusação - que foi a acusação de alguns dos seus contemporâneos - de ser um Berkeley sob disfarce, i.e. um idealista? Kant procurou empenhadamente rebater esta acusação. Mas se Kant não é um idealista, qual é a relação do sujeito transcendental, de toda a ‘maquinaria’ da racionalidade, com o mundo da natureza?

## Conclusão

A epistemologia é, conjuntamente com a metafísica, uma das disciplinas centrais da filosofia, uma disciplina sem a qual o propósito de fazer filosofia não se entende. Compreender a mente e o conhecimento é um propósito subjacente a tudo o mais que se pretenda fazer em filosofia. Disponibilizar um trajecto possível para a exploração de questões relativas à mente e ao conhecimento, correspondente àquilo que tem sido a minha prática na FLUP desde 1996, tendo por um lado a teoria da mente e do conhecimento tal como ela se apresenta contemporaneamente e por outro lado a história da filosofia do conhecimento, foi a ideia subjacente à reunião dos textos que se seguem neste volume. Esses textos tratam, na sua maioria, de temas e autores contemporâneos. Pretendi com a presente Introdução exemplificar a comunicação que penso existir entre esses tópicos e autores com a história da filosofia e em particular com Descartes, Hume e Kant. Creio que tal comunicação deve ser clara no ensino da epistemologia. Espero ter assim contribuído para mostrar a variedade dos problemas encontrados na tentativa de compreender a mente e o conhecimento, bem como a variedade de fontes de respostas com que os filósofos hoje se deparam.

Ronda - Porto, Verão de 2009

## Referências principais

Descartes, R., 1985, *Meditações sobre a Filosofia Primeira*, introdução, tradução e notas de Gustavo de Fraga, Coimbra, Livraria Almedina. (Texto original das *Meditações* em latim em <http://www.wright.edu/cola/descartes/medl.html>)

Hume, D., 2002, *Tratados Filosóficos – Investigação sobre o Entendimento Humano*, tradução e introdução de João Paulo Monteiro, Lisboa, INCM.

Hume, D., 2001, *Tratado da Natureza Humana*, tradução de S. Silva Fontes, prefácio e revisão técnica da tradução de João Paulo Monteiro, Lisboa, Fundação Calouste Gulbenkian.

Kant, I., 1985, *Crítica da Razão Pura*, tradução de Manuela Pinto dos Santos e Alexandre F. Morujão, introdução e notas de Alexandre F. Morujão, Lisboa, Fundação Calouste Gulbenkian.

Leibniz, G., 1987, *Princípios de Filosofia ou Monadologia*, tradução, introdução e notas de Luís Martins, Lisboa, INCM.

## Outras referências

Ameriks, Karl, “The critique of metaphysics: the structure and fate of Kant’s dialectic”, 2006, *The Cambridge Companion to Kant and Modern Philosophy*, Cambridge, Cambridge University Press.

Cottingham, John, 1992, *The Cambridge Companion to Descartes*, Cambridge, Cambridge University Press.

Deleuze, Gilles, 1973, *Empirisme et subjectivité*, Paris, PUF.

Dennett, Daniel, 1996, *Kinds of Minds*, New York, Basic Books.

Fate Norton, David, 1993, *The Cambridge Companion to Hume*, Cambridge, Cambridge University Press.

Guyer, Paul, 1992, *The Cambridge Companion to Kant*, Cambridge, Cambridge University Press.

Guyer, Paul, 2006, *The Cambridge Companion to Kant and Modern Philosophy*, Cambridge, Cambridge University Press.

Guyer, Paul, 1992, “Transcendental Deduction of the categories”, *The Cambridge Companion to Kant*, Cambridge, Cambridge University Press.

Hanna, Robert, “Kant’s theory of judgment”, *The Stanford Encyclopedia of Philosophy* (E. Zalta ed.) <http://plato.stanford.edu/entries/kant-judgment/>



Hatfield, Gary, “René Descartes”, *The Stanford Encyclopedia of Philosophy* (E. Zalta ed.) <http://plato.stanford.edu/entries/descartes/>

Hume, D., 2002, *Ensaíais Morais, Políticos e Literários*, Lisboa, INCM. Traduções de João Paulo Monteiro, Sara Albieiri e Pedro Galvão.

Hume, D., 2005, *Tratados Filosóficos II – Dissertação sobre as Paixões e Investigação sobre os Princípios da Moral*, traduções de João Paulo Monteiro e Pedro Galvão, Lisboa, INCM.

Hume, D., 2005, *Obras sobre Religião*, Lisboa, Fundação Calouste Gulbenkian. Traduções de Pedro Galvão e Francisco Marreiros. Introdução e Revisão Técnica de João Paulo Monteiro.

Jolley, Nicholas, *The Cambridge Companion to Leibniz*, Cambridge, Cambridge University Press.

Kant, Immanuel, 1994, *Crítica da Razão Prática*, Lisboa, Edições 70.

Kitcher, Patricia, 2006, “Kant’s philosophy of the cognitive mind”, *The Cambridge Companion to Kant and Modern Philosophy*, Cambridge, Cambridge University Press.

Kuehn, Manfred, 2001, *Kant – a biography*, Cambridge, Cambridge University Press.

Longuenesse, Béatrice, 1998, *Kant and the Capacity to Judge*, Princeton, NJ, Princeton University Press (versão inglesa revista de *Kant et le pouvoir de juger: sensibilité et discursivité dans l’analytique transcendentale de la critique de la raison pure*, Paris, PUF, 1993).

Longuenesse, Béatrice, 2006, “Kant on a priori concepts: the metaphysical deduction of the categories”, *The Cambridge Companion to Kant and Modern Philosophy*, Cambridge, Cambridge University Press.

Look, Brandon C., “Gottfried Wilhelm Leibniz”, *The Stanford Encyclopedia of Philosophy* (E. Zalta ed.) <http://plato.stanford.edu/entries/leibniz/>

Look, Brandon C., “Leibniz’s Modal Metaphysics”, *The Stanford Encyclopedia of Philosophy* (E. Zalta ed.) <http://plato.stanford.edu/entries/leibniz-modal/>

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

Miguens, S. & Mauro, C. E. 2006, *Perspectives on Rationality*, Porto, FLUP.

Morris, William Edward, “David Hume”, *The Stanford Encyclopedia of Philosophy* (E. Zalta ed.) <http://plato.stanford.edu/entries/hume/>

Newman, Lex, “Descartes’ Epistemology”, *The Stanford Encyclope-*

*dia of Philosophy* (E. Zalta ed.) <http://plato.stanford.edu/entries/descartes-epistemology/>

Pessin, Andrew, “Descartes’ Theory of ideas”, *The Stanford Encyclopedia of Philosophy* (E. Zalta ed.) <http://plato.stanford.edu/entries/descartes-ideas/>

Quine, W. V., 1969, *Ontological Relativity*, in *Ontological Relativity and Other Essays*, New York, Columbia University Press.

Teixeira, Célia, Como não entender o a priori – analiticidade, revisibilidade, necessidade, *Intelectu* 10, Novembro de 2004, [www.intelectu.com](http://www.intelectu.com)

## APÊNDICE 1 – LEIBNIZ E OS MUNDOS POSSÍVEIS

Descartes e Leibniz são, na história da filosofia, casos por excelência em que com apoio numa concepção de Deus se caracteriza o espírito humano e o conhecimento humano. E caracteriza-se de formas muito diferentes entre si. A noção leibniziana de ‘mundos possíveis’ é uma pista para compreender este contraste. G. W. Leibniz (1646-1716) foi um dos últimos espíritos universais da história do pensamento, um homem que não apenas dominou o conhecimento da sua época, como foi inovador em diversos campos, desde a matemática<sup>50</sup> à geologia e ao estudo da linguagem – no entanto ele é, por várias razões (uma delas sendo certamente a forma de falar de Deus) pouco acessível intuitivamente aos leitores contemporâneos. Talvez valha a pena recordar, assim sendo, aquilo que Bertrand Russell, numa famosa obra sobre o pensamento de Leibniz, *A Critical Exposition of the Philosophy of Leibniz* (1900) afirmou sobre Leibniz, mais especificamente sobre a *Monadologia*<sup>51</sup>, que tão estranha pode parecer aos leitores

<sup>50</sup> Leibniz foi um dos criadores do cálculo infinitesimal. Houve aliás uma disputa ácida acerca da prioridade de Leibniz ou de Newton na criação do cálculo e Leibniz foi maltratado pela Royal Society inglesa. Pensa-se hoje que Leibniz e Newton teriam criado o cálculo independentemente.

<sup>51</sup> Obra escrita por Leibniz em 1714, em que ele expõe o seu sistema metafísico a partir da noção de mónada. As mónadas são substâncias simples e imateriais, criadas, que ‘não nascem nem morrem’. Este sistema metafísico imaterialista foi em parte despoletado pela crença de Leibniz na inexistência de átomos – i.e. simples – físicos. Tudo aquilo que há são mónadas. As mónadas têm percepções e apetições e cada uma delas exprime todo o universo. No nosso caso, a mónada-alma tem também apercepção ou consciência. A percepção é o estado interno da mónada representando coisas exteriores, a apercepção é a consciência ou conhecimento reflexivo deste estado interno. A apercepção não existe em todas as mónadas. Isto quer dizer que as percepções que caracterizam todas as mónadas podem ou não ser objecto de apercepção, i.e. podem ser ou não ser conscientes. Leibniz dirá que nem todas as mónadas [seres perceptivos e apetitivos] são espíritos – elas serão espíritos conforme nelas exista ou não a percepção das percepções e de si enquanto percebendo. Por isso Leibniz fala em *mónadas nuas* [correspondendo à matéria inanimada], *mónadas memorantes* [correspondendo por exemplo a animais] e *mónadas reflexivas* [correspondendo aos humanos]. Definitivamente, a mente e a representação não são para Leibniz exclusivas dos humanos, embora o conhecimento das verdades necessárias e dos princípios (i.e. a ‘razão’) lhes esteja reservada. Mas aquilo que no

contemporâneos: trata-se de ‘um conto de fadas metafísico, talvez coerente mas totalmente arbitrário’, apenas compreensível fazendo apelo à teoria lógica que lhe subjaz. Suponhamos então que é para a teoria lógica que um leitor contemporâneo deverá olhar primeiro, se desejar evitar a estranheza. O que é a lógica, para Leibniz? A lógica não é para Leibniz aquilo que é para os lógicos contemporâneos, i.e. uma ciência formal; Leibniz chama ‘lógica’ à sua teoria da natureza da proposição e da natureza da verdade.

De acordo com Leibniz, todas as proposições têm, expressa ou implicitamente, a forma ‘S é P’. A predicação é a afirmação da inclusão de P (predicado) em S (sujeito), com fundamento na natureza das coisas. Numa proposição verdadeira o conceito do predicado está contido no conceito do sujeito. Leibniz pensa ainda que uma proposição verdadeira é em última análise redutível a uma afirmação de identidade. Ora, é a partir daqui que Leibniz começa a aproximar-se do interesse do epistemólogo contemporâneo, nomeadamente quando faz uma proposta acerca de tipos de verdades existentes. Segundo Leibniz existem verdades de razão (verdades necessárias, cujo ‘oposto é impossível’, por exemplo as verdades matemáticas e lógicas) e verdades de facto (verdades contingentes ou de facto, cujo ‘oposto é possível’, tais como ‘Júlio César atravessou o Rubicão’ ou ‘Adão pecou’). A esta proposta acerca da existência de diferentes tipos de verdades acresce uma ideia acerca dos princípios que fundam os raciocínios. Estes são o Princípio da Contradição (‘em virtude do qual julgamos falso o que implica contradição’) e o Princípio da Razão Suficiente (‘em virtude do qual consideramos que nenhum facto pode ser verdadeiro ou existente e nenhuma enunciação verídica sem que haja uma razão suficiente para que isso assim seja e não de outro modo’)<sup>52</sup>. Tudo isto será esclarecido em termos de mundos possíveis.

Em contraponto com o interesse de Leibniz pela lógica concebida desta maneira, que pode servir de entrada a todo o seu sistema metafísico, diga-se desde já que Leibniz não teve qualquer interesse pelo projecto cartesiano de encontrar fundamentos para o conhecimento: preferiu os

*universo parece inanimado ou morto é um atordoamento das mónadas imortais. A noção de mónada traduz ainda a actividade ou espontaneidade que para Leibniz caracteriza o espírito. Para Leibniz não há percepção sem ‘auto-realização’ e mudança (ao contrário do que se passa com Descartes, para quem a percepção era um modo do pensamento e o movimento uma determinação da extensão). Na medida em que a matéria é em toda a parte a mesma, Leibniz pensa que se só houvesse matéria no mundo não haveria actividade, diversidade ou harmonia. Como o mundo tem essas características, deve haver mentes. Especificamente, mentes particulares existem porque o ser supremo julga harmonioso que deva existir ‘aquilo que compreende’, que é como um espelho intelectual ou réplica do mundo. É ainda às mónadas que é devida a unidade da representação, inexplicável por razões mecânicas.*

<sup>52</sup> Segundo Leibniz, o Princípio da Contradição é suficiente por si só para a demonstração da totalidade das matemáticas, mas o Princípio da Razão Suficiente é necessário para passar da matemática à física, que lida com um reino de entidades contingentes.

princípios metafísicos, muito particularmente o acima referido Princípio da Razão Suficiente (segundo o qual *Nihil est sine ratione*) e a procura de isomorfismos entre áreas teóricas tais como a matemática, a metafísica e a teoria da linguagem. Nos seus escritos não há um tema que sirva como pedra angular única, como era o caso do cogito cartesiano: como o próprio admite, no seu sistema há várias entradas, que vão desde Deus até à mónada, à teoria da noção completa<sup>53</sup>, à ideia de expressão, à ideia do melhor dos mundos possíveis.

Mas voltemos à lógica e aos mundos possíveis. Leibniz foi satirizado no *Candide* (1759) de Voltaire (1694-1778) como o filósofo que (absurdamente) considerou este mundo como o ‘melhor mundo possível’ (qualquer pessoa razoável deveria perguntar como seriam os outros...). Para saber se isto é assim tão absurdo (Leibniz de facto defendeu tal tese) interessa compreender o que se entende por ‘mundos possíveis’, relacioná-los com a concepção leibniziana de Deus e compreender como é que os mundos possíveis e Deus moldam uma concepção do conhecimento humano.

Um mundo é um modo completo (em todos os seus detalhes passados, presentes e futuros) de as coisas serem. Possível é aquilo que não envolve contradição. Cada mundo é uma colecção de possíveis compostíveis. Os possíveis A e B são compostíveis se são não apenas possíveis isoladamente mas também consistentes entre si. Mundos possíveis serão portanto modos completos de possíveis compostíveis, i.e. modos completos de as coisas serem, incluindo o modo completo como as coisas são. Os mundos possíveis – exceptuando o mundo em que estamos, i.e. o mundo possível actual – têm para Leibniz uma existência apenas conceptual, nas ‘ideias de Deus’. Por alguma razão, que Leibniz explicará, apenas um único mundo, de entre os mundos possíveis, existe. Antes de prosseguirmos e para compreender que este problema excede qualquer âmbito teológico, vejamos de um ponto de vista prático, contemporâneo, por que será útil falar de mundos

<sup>53</sup> *Tal como a mónada, esta é uma noção fulcral em Leibniz e bastante estranha à compreensão contemporânea. Uma noção completa é a série total dos predicados contidos na noção do sujeito, correspondendo à total explicitação de uma substância individual (mónada). A noção supõe algo como ‘o ponto de vista de Deus’ [que tal ponto de vista se possa ou não evocar faz uma grande diferença em filosofia]. Leibniz pensa que existem conceitos individuais [por exemplo o conceito de ‘Alexandre o Grande’] e que a série total dos predicados que compõem esses conceitos individuais é acessível do ponto de vista de Deus. Temos que imaginar um conceito de pessoa – eu, por exemplo – e um ponto de vista a partir do qual tudo o que fui, sou, serei, é acessível. Este ponto de vista é obviamente Deus, a que Leibniz chama a ‘mónada das mónadas’, e que é um ser necessário, fonte das existências e das essências. Deus é livre e cria de acordo com a razão. Aliás, só agindo de acordo com a razão é que ele é livre. Um deus arbitrário como o deus cartesiano [i.e. que cria as próprias ‘verdades eternas’] não seria, de acordo com Leibniz, livre. Para Leibniz o bem não depende da vontade de Deus: o bem é antes a Razão Suficiente daquilo que Deus quis fazer. Em geral, apenas as verdades contingentes dependem da vontade de Deus, as verdade necessárias são o objecto interno do seu entendimento.*

possíveis. Desde logo é útil falar de mundos possíveis porque constantemente assumimos nas nossas crenças e raciocínios que certas coisas são possíveis, outras impossíveis, outras muito pouco susceptíveis de serem senão como são ('necessárias', dir-se-á em vocabulário mais sofisticado). Por exemplo: perdi o avião – mas poderia perfeitamente tê-lo apanhado. Estou na aula de Filosofia do Conhecimento – mas poderia ter ficado em casa a jogar jogos no computador. Aristóteles nasceu em Estagira. Mas podia ter nascido em Atenas e ter sido ainda assim o autor da *Poética*, da *Retórica*, da *Metafísica* e o aluno de Platão. Estou a ver esta árvore – esta árvore é verde. Mas poderia ser vermelha ou castanha. Dois mais dois são quatro – será que podiam ser cinco? Água é  $H_2O$  – será que não podia ser outra coisa qualquer? Mesmo com estas variações incipientes noto que algumas coisas que eu penso parecem mais 'flexíveis' que outras a poderem ser de outra maneira – porque será que isso é assim? De qualquer forma é algo que levou filósofos clássicos como Leibniz, a diferenciar tipos de conhecimento (por exemplo conhecimento lógico de conhecimento histórico) em termos de mundos possíveis. Nos nossos raciocínios correntes usamos ainda condicionais contrafactuais. Condicionais contrafactuais são condicionais com a forma '*se tal e tal tivesse sido o caso, então tal e tal teria sido o caso*' que dizem o que aconteceria em situações não actuais (Se Obama não tivesse ganho as eleições, John McCain teria sido Presidente). É evidentemente problemático saber o que pensar quanto à natureza de tais 'situações não actuais', como o mundo em que Obama perdeu as eleições e não é Presidente – serão elas cenários pensados ou pensáveis mas que não têm nada a ver com a forma de as coisas serem? Bom, mas são pelo menos pensáveis – o que é que isso significa?<sup>54</sup>

Voltando a Leibniz: embora Leibniz tenha 'introduzido' os mundos possíveis, para ele só existia um mundo, um mundo seleccionado por Deus por entre os mundos possíveis. Os outros mundos possíveis – exceptuando o mundo em que estamos, i.e. o mundo possível actual – têm para Leibniz uma existência apenas conceptual. Ora, a razão para este mundo em que estamos existir é o facto de ser o 'melhor dos mundos possíveis'. Este é o clímax da teodiceia leibniziana, um frequente objecto de escárnio e uma tese muito importante da filosofia de Leibniz, um aspecto que no limite poderia ser trabalhado sem evocar qualquer aspecto religioso acerca de

<sup>54</sup> A questão dos condicionais contrafactuais e da sua natureza é importante e incontornável, nomeadamente na filosofia da ciência, quando se considera a natureza das leis científicas – uma lei científica distinguir-se-ia de uma generalização acidentalmente verdadeira precisamente pela sua capacidade de resistir a condicionais contrafactuais.

Deus e pensando apenas em implicações metafísicas e epistemológicas. No contexto de uma disciplina de epistemologia é aliás obviamente menos importante a questão de Deus do que a diferença entre as verdades necessárias e as verdades contingentes, tal como estas aparecem do ponto de vista do intelecto divino e do ponto de vista do intelecto humano. É aí que os mundos possíveis importam. Vejamos como aparece a natureza dos diferentes tipos de verdades em termos de mundos possíveis. As verdades contingentes exprimem aquilo que é mas não teria que ter sido. Existem portanto razões para aquilo que elas afirmam ser verdadeiro (essas razões baseiam-se naquilo que parece o melhor entre vários possíveis). ‘Contingente’ poderia ser definido como aquilo que é verdadeiro em pelo menos um mundo possível e falso em pelo menos um mundo possível. Verdades necessárias, em contraste, são aquelas que seriam verdadeiras em qualquer mundo possível. Exemplos das primeiras poderiam ser ‘Júlio César atravessou o Rubicão’, ‘Adão pecou’, ou ‘Joana está na aula de Filosofia do Conhecimento’ e das segundas ‘ $2+2=4$ ’. A ideia de Leibniz é que as verdades contingentes (e apenas elas) se baseiam na vontade (de Deus ou das criaturas), a sua escolha resulta de razões (já que *Nihil est sine ratione*).

Ao diferenciar desta forma verdades contingentes e verdades necessárias do próprio ponto de vista de Deus uma das coisas que Leibniz está a fazer é opôr-se à doutrina cartesiana da livre criação divina das ‘verdades necessárias’. Para Descartes não fazia sentido pensar que o intelecto de um ser onipotente fosse de alguma forma constrangido pelas ‘verdades necessárias’ – assim, mesmo estas (por exemplo  $2+2=4$ ) teriam sido livremente criadas por Deus, sendo portanto de alguma forma contingentes, dada a sua natureza de ‘criaturas’. Mas Leibniz considerou que essa doutrina mostrava que Descartes pensava em Deus como se este não tivesse entendimento...

Mas, quanto à forma como os humanos as pensam, qual é a diferença entre as verdades necessárias e as verdades contingentes? De um ponto de vista humano, uma verdade contingente (‘A Joana está na aula de Filosofia do Conhecimento’) é um caso especial de inclusão de um predicado num sujeito, um caso em que essa inclusão não é óbvia (parece-nos que a Joana poderia ser a Joana e não estar agora na aula de Filosofia do Conhecimento, estar por exemplo em casa, ou na praia). A análise do conceito (‘Joana’)—que proporcionaria a afirmação da inclusão—é assim infinita, não exequível pelo intelecto humano. Pelo contrário no caso das verdades necessárias, a análise é finita e a inclusão é óbvia. Os humanos podem demonstrar por meio de um número finito de passos a inclusão do conceito do predicado no conceito do sujeito (‘Um triângulo tem três

lados' seria um exemplo). As verdades necessárias baseiam-se directamente na 'possibilidade ou impossibilidade das próprias essências'; aquilo que elas exprimem teria que ter sido como é em qualquer mundo, pois o seu oposto envolve contradição – um triângulo não poderia ser um triângulo e não ter três lados.

Acontece que para Leibniz esta diferença, do ponto de vista humano, entre verdades contingentes e verdades necessárias é de alguma forma rebatida; as verdades contingentes são elas próprias um certo tipo de necessidade, que do ponto de vista de Deus aparece como tal, embora seja inacessível ao pensamento humano. Noutras palavras, neste mundo, porque nada é sem razão, as coisas expressas por verdades contingente teriam que ter sido exactamente assim: a Joana não seria a Joana sem estar agora na aula de Filosofia do Conhecimento. A escolha das formas de ser expressas por estas verdades têm razões, embora estas razões, diz Leibniz, inclinem sem necessitar.

Está-se aqui, como é óbvio, num terreno teológico e moral muito complicado. Pensemos em cada um de nós, e nas coisas que somos e fazemos. De acordo com Leibniz não se pode pensar num mesmo indivíduo como 'membro ou parte' de vários mundos possíveis: i.e. eu não poderia ser eu e 'pertencer a outro mundo'. O conceito completo de um indivíduo está ligado a um mundo possível. Pense-se por exemplo nos conceitos individuais "Júlio César" e "Adão" e nas afirmações "*Julio César atravessou o Rubicão*" e "*Adão pecou*". Em qualquer conceito individual, considerado do ponto de vista de Deus, está incluída a série completa dos predicados, logo as noções "Júlio César" e "Adão" podem ser compreendidas pelo desdobramento dos seus predicados. A compreensão da noção completa de substâncias individuais é para nós, actualmente, infinita, ela é, no entanto, segundo Leibniz, possível por princípio, nas ideias de Deus. Deus 'tem a chave' de aquilo que existe ser como é. Por exemplo: Adão pecou, no entanto um Adão não pecador não é impossível ou contraditório em si. O que houve, para Leibniz, foi qualquer coisa como uma "exclusão original": o Adão-não-Pecador não é impossível, no entanto a sua existência envolveria todo um outro mundo. Ora apenas existe um mundo, este mundo. Para a passagem à existência dos possíveis é necessário que os possíveis sejam compossíveis (por exemplo, Adão-pecador e Adão-não-pecador, embora sejam possíveis cada um por si, não são compossíveis). Deus não decidiu se Adão pecaria ou não pecaria; decidiu pela série de coisas na qual há um Adão cujo conceito completo envolve pecar pelo facto de esta ser preferível a outras. A leitura da noção de melhor mundo possível é então a seguinte: este mundo, o único mundo que existe, é apenas o melhor possível, i.e. o melhor que se poderia fazer. Imaginando o acto divino de criar

mundos e os mundos que poderiam ter sido, temos que existe este mundo simplesmente porque é o ‘melhor possível’.

Repare-se também que nem tudo o que é possível existe: o universo não é a colecção de todos os possíveis. Nem todos os possíveis são compostíveis e apenas podem existir possíveis compostíveis. Há portanto possíveis não actualizados. Quanto à relação, no mundo possível actual, entre o actual e o necessário, temos o seguinte: as coisas neste mundo são como são, mas elas não tinham que ter sido como são. No entanto, tudo o que existe teria que ter sido tal como é neste mundo, e há razão para este mundo, e não outro, existir (ele é o melhor mundo possível). Em suma, nem tudo o que é actual é metafisicamente necessário. Mas o que é actual pode não ser metafisicamente necessário e ser ainda assim uma necessidade de um certo tipo. Para Leibniz há necessidades hipotéticas: são as verdades contingentes, as verdades morais e físicas. Elas são verdadeiras neste mundo, por contraste com as verdades matemáticas e lógicas que teriam sido verdadeiras em todos os mundos possíveis (i.e. são necessidades metafísicas).

Tudo isto parece uma metafísica necessitarista. No entanto, de acordo com Leibniz, não é esse o caso, porque se admite a acima explicada contingência objectiva. Para Leibniz existem verdades objectivamente contingentes. Assim, adscrever contingência a verdades não é apenas uma marca da ignorância humana (ao contrário do que pensou por exemplo Espinosa). De um ponto de vista necessitarista, o entendimento da forma como as coisas realmente são mostra que o que quer que exista, existe necessariamente, e o que não existe, não pode existir. Para Leibniz, pelo contrário, é absurdo supôr que o oposto daquilo que as verdades contingentes exprimem seja impossível, ‘desde a eternidade’. Outras coisas poderiam existir em vez das coisas que existem, outras formas de as coisas serem poderiam ter sido. É porque existem possibilidades não actualizadas que existem verdades objectivamente contingentes.

Pensando na aplicação destas ideias metafísicas à questão da liberdade humana, o que se obtém? Isto parece um fundamento muito ténue para a liberdade humana, dependente como esta está da possibilidade de as coisas serem de uma ou outra maneira. Mas Leibniz pensa que os humanos são de facto livres. O presente estado do mundo, ou qualquer outro estado do mundo, não é absolutamente necessário mas apenas hipoteticamente necessário. Seres livres fazem escolhas. Ora, não é que as escolhas não sejam ‘necessitadas’, mas aquilo que delas resulta não é absolutamente necessário, no sentido em que tem todo o sentido pensar que o curso dos eventos pudesse ter sido outro. Uma verdade que expressa uma escolha livre não é, assim, absolutamente necessária. A escolha livre não é por-



tanto necessária em absoluto. No entanto, a verdade é que ela só poderia ter sido essa, não poderia ter havido outro curso de acção. Pode-se razoavelmente duvidar que sem possibilidade actual a liberdade seja realmente possível<sup>55</sup>...

## APÊNDICE 2 – HUME E A RELIGIÃO

Existe, em David Hume, uma conexão clara entre Filosofia do Conhecimento e Filosofia da Religião. Ela pode ser explorada da seguinte forma. Um primeiro passo seria procurar compreender as posições defendidas por Hume nas secções X (Dos Milagres) e XI (De uma Providência Particular e de um Estado Futuro) da *Investigação sobre o Entendimento Humano*, relacionando-as com as questões epistemológicas gerais da crença, evidência e experiência. Em seguida passar-se-ia aos escritos sobre religião, nomeadamente *The Natural History of Religion* (1757) e os *Dialogues Concerning Natural Religion* (1779, póstumo). O tema do primeiro é a religião como fenómeno psicológico; a obra foi denunciada como a obra de um ateu mesmo antes de ser publicada. Aí se explica a forma como a crença religiosa se origina nas partes mais obscuras da natureza humana, no medo e na ignorância. Argumenta-se que o monoteísmo, embora pareça uma posição religiosa sofisticada, é moralmente retrógrado pois encoraja o zelo e a intolerância, constituindo um perigo social por ser a causa de actos violentos e imorais contra os heterodoxos. Os *Dialogues Concerning Natural Religion*, o segundo dos escritos de Hume sobre religião, são uma obra extraordinária. A. Schopenhauer considerou estes Diálogos entre Filon, Cleanto e Demea como uma das grandes obras primas literárias e filosóficas da história. O tema central é a crítica ao argumento do desígnio (*design argument*), um argumento a favor da existência de Deus de acordo com o qual o universo, sendo como é, só pode ser o efeito de uma causa

<sup>55</sup> Para uma continuação desta temática leibniziana na lógica e metafísica contemporâneas, cf. João Alberto Pinto, *Materialismo, Superveniência e Experiência*, Porto, Campo das Letras, 2007. Note-se ainda que além de uma visão da liberdade humana baseada nos mundos possíveis, a concepção leibniziana do humano inclui uma explicação da *causação mental e da identidade pessoal*. Para Leibniz todas as almas (mónadas) estão sempre conjugadas com corpos, e nunca completamente separadas deles. No entanto, de acordo com o paralelismo não existe interacção entre o físico e o mental, apenas co-ocorrência do mental e do físico de acordo com a harmonia pré-estabelecida (i.e. quer as almas quer os corpos seguem as suas próprias leis, estas leis encontram-se em virtude da harmonia pré-estabelecida). A espontaneidade está 'do lado' das mónadas (para Leibniz a actividade é a essências das substâncias) e é em virtude dessa espontaneidade que os eventos se desenrolam e que a organização dos corpos acontece. É por meio da mónada-alma que há uma verdadeira unidade em cada humano. Como todas as mónadas, cada mónada alma nada recebe de fora, antes foi criada com perfeita espontaneidade relativamente a si própria e perfeita conformidade relativamente ao exterior. Assim, cada mónada representa todo o universo à sua maneira, como se a cada momento apenas existissem Deus e essa substância, e no entanto nela se exprimem as coisas exteriores. É esta comunicação regulada à partida que produz a união alma-corpo. A identidade pessoal é a noção completa do indivíduo, correspondente à mónada-alma.

supremamente inteligente. Ora, basicamente, Hume pensa que esta conclusão vai muito além das provas disponíveis. As personagens dos *Diálogos*, Fílon, Cleanto e Demea, são respectivamente um céptico, um cientista religioso, defensor de um teísmo experimental, e um religioso ortodoxo. É Cleanto quem enuncia o Argumento do Desígnio. Ele diz: «Olhai em volta, para o mundo. Contemplai o todo e cada parte do todo: descobrireis que ele não é senão uma grande máquina, subdividida num número infinito de máquinas menores, que por sua vez admitem subdivisões até um grau que está para além daquilo que os sentidos e as faculdades humanas podem seguir e explicar. Todas estas variadas máquinas, mesmo nas suas mínimas partes, estão ajustadas umas às outras com uma precisão que conduz à admiração todos os homens que alguma vez as contemplaram. A curiosa adaptação de meios e fins, em toda a natureza, assemelha-se exactamente, embora a exceda muitíssimo, às produções do engenho humano; do desígnio humano, pensamento, sabedoria e inteligência. Como portanto os efeitos se assemelham, somos conduzidos a inferir, por todas as regras da analogia, que as causas também se assemelham, e que o Autor da Natureza é de alguma forma semelhante à mente do homem, embora possuindo faculdades muito superiores, proporcionais à grandeza da obra que executou. Por este argumento a posteriori, e unicamente mediante este argumento, provamos ao mesmo tempo a existência de uma Divindade e a sua semelhança com a mente e a inteligência humanas». Interessa, para compreender o alcance das posições de Hume quanto a esta argumentação, recordar a forma como um empirista anterior, John Locke, via as coisas (*Ensaio sobre o Entendimento Humano*, 1690 (IV, x,10)): “Se deve haver algo de eterno, vejamos que tipo de ser será. E é muito óbvio à Razão que deve ser necessariamente um ser Cogitativo. Pois é tão impossível conceber que alguma vez a pura Matéria não cogitativa produzisse um ser pensante inteligente como que o nada por si próprio produzisse a Matéria (...) Suponhamos a Matéria eterna; ela não é capaz de produzir nada, nem sequer o Movimento, que deve ser desde toda a eternidade acrescentado à Matéria por algum ser mais poderoso do que a Matéria (...) Mas suponhamos a Matéria não Cogitativa e o Movimento eternos: quaisquer mudanças que pudessem produzir de figura e dimensão, nunca poderiam produzir o Pensamento. (...) O Pensamento nunca poderia começar a ser”. O Argumento do Desígnio na religião natural afirma o mesmo que Locke afirma aqui, simplesmente não é a priori, não visa o que pode ou não ser concebido, faz antes apelo à experiência da natureza, das formas dos seres, e ao curioso ajustamento entre meios e fins com que aí nos deparamos. Não poderia haver essa ordem, esse ajustamento, sem uma razão inteligente que os produzisse. No núcleo da questão está portanto saber se o pensamento pode ser concebido como derivado ou se tem mesmo que ser

a origem. Para Locke, o pensamento tem que ser a origem. Mas as coisas mudaram muito entre Locke e Hume. Convém recordar que os *Diálogos* de Hume tiveram uma dupla posteridade, agnóstica e fideísta (o fideísmo é uma visão de fé mas uma visão pessimista acerca das capacidades da razão no que respeita a questões religiosas). Esta dupla posteridade não é injustificada. Embora na maior parte do livro se procure mostrar que o Argumento do Desígnio não é um bom argumento, a personagem mais facilmente identificável como representando as posições de Hume, Filon, acaba por 'recuar' para uma confissão de religiosidade, aparentando uma conversão às teses teístas. São no entanto legítimas as dúvidas acerca da sinceridade da personagem Filon quando na secção final dos *Diálogos* professa a sua religiosidade.

Ao longo dos *Diálogos* o Argumento do Desígnio é 'acusado' de muita coisa. É acusado de ser antropomorfista: qual seria o privilégio desta pequena agitação do cérebro (Parte II) que justificasse o facto de fazermos dela o modelo do universo? porque não a aranha da cosmogonia brâmane (Parte VII) em vez do pensamento e do cérebro do homem? porque nos parece mais ridículo? É acusado de inferir mais poder, inteligência e bondade do que aqueles que seriam suficientes para produzir aquilo que existe, já que este mundo, com os seus erros e incoerências, bem poderia ser resultado de tentativas abandonadas de um deus já idoso e morto, ou o trabalho de diversos artífices não inteligentes, ou apenas um por entre vários universos desperdiçados, ou obra de uma divindade inferior objecto de riso para os deuses mais altos, ou resultado de um melhoramento cego e inintencionado em idades infinitas de 'world-making'. Tudo isso também se poderia inferir e não seria mais incoerente do que a ideia de uma única divindade onnipotente e sábia (Parte V). O argumento é acusado, finalmente, de, sendo um raciocínio de experiência, partir de um único caso (o universo) para inferir a partir de um efeito a sua causa, quando no raciocínio a partir da experiência e pela causalidade precisamos de vários casos (Parte II). Nos próprios *Diálogos*, as alternativas ao Argumento do Desígnio são hipóteses materialistas pré-darwinistas acerca da origem da ordem na existência de combinações e tentativas e erros. O darwinismo fez muito para mudar os dados com que pensamos nesta questão; no entanto, já antes de Darwin, Hume, nos *Diálogos*, desbravou o terreno.

### APÊNDICE 3 – POR TRÁS DA CONCEPÇÃO KANTIANA DE ESPAÇO E TEMPO

A ideia kantiana segundo a qual o espaço e o tempo são 'formas da intuição' e não características da realidade em si tem um lugar específico

na história da ciência. Por trás dela está nomeadamente a célebre controvérsia que teve lugar em 1715-1716 entre Leibniz e Samuel Clarke, um filósofo newtoniano inglês, uma controvérsia na qual se discutiu acerca de Deus e da sua glória, elucidando ao mesmo tempo pontos essenciais da ciência física moderna e tendo sido esboçada uma concepção de espaço e tempo como realidades relativas e não absolutas. O fim do debate foi totalmente exterior a este – ninguém o venceu: “Mr Leibnitz was prevented by Death from returning my answer”, constatou Clarke. O contexto do debate foi o seguinte. Para os newtonianos o espaço é absoluto, real e independente dos corpos que pode ou não conter. A matéria é a parte menos importante do universo, que é sobretudo constituído por espaço vazio. Este espaço é um só e é resultado da existência de um ser infinito e eterno (Deus). Newton chega a chamar-lhe *sensorium dei*. Para Newton, a filosofia natural transcende-se a si própria e conduz a Deus: o estudo científico do universo revela a acção criadora óbvia de Deus e a inteligência do seu desígnio. Deus não está acima do espaço e do tempo, ele constitui a duração (a sua eternidade é duração eterna) e o espaço (a sua compressão é ubíqua). É deste modo que Deus exerce a sua soberania. Como se vê, o Deus de Newton não é um deus ‘filosófico e ausente’ como o Deus de Descartes e Leibniz – pelo contrário ele está em toda a parte de um modo muito pouco metafórico. Ora Leibniz ironizou acerca das concepções religiosas dos newtonianos, nomeadamente acerca da noção de *sensorium dei* (o espaço seria o órgão de que Deus se serve para sentir as coisas?...) e da noção de *manus emendatrix* (Deus necessita de tempos a tempos de dar corda ao seu relógio senão este não continuaria a funcionar?...). Tendo falado da ‘impiedade que grassa em Inglaterra’, devido às ideias acerca de Deus dos newtonianos, Leibniz será, por seu lado, acusado de ser ‘materiaalista’, ‘deicida’ e ‘fatalista’.

Interessa olhar para alguns passos deste debate, particularmente para uma suposição avançada por Samuel Clarke. Disse Clarke: se Newton tem razão e o espaço é absoluto, pode existir um outro universo idêntico a este, transposto no espaço e será outro, e não o mesmo. Pelo contrário Leibniz teria que defender que um sistema de coisas físicas ordenadas ‘transportaria consigo’ os seus referenciais, sendo portanto impossível decidir se ele está ‘aqui’ ou ‘ali’, antes ou depois. Esta consequência ‘relativista’ parece decorrer da teoria leibniziana e Clarke considera-a absurda.

Vejamus um pouco mais especificamente a forma como as coisas se punham para Leibniz quanto ao mundo físico: para Leibniz (1) a realidade actual são as mónadas; a discernibilidade e a individuação daquilo que existe deve-se a estas substâncias simples e imateriais; (2) a corporeidade não é auto-explicativa; ela não deve ser no entanto eliminada, mas reduzida ao que é, um fenómeno resultante das existências simples, as mónadas,

que têm realidade absoluta; (3) o espaço e o tempo são idealidades, que pressupõem a existência de coisas para serem ordenadas na sucessão e na coexistência. Assim sendo, sem “criaturas” (2), o espaço e o tempo não seriam actuais. A física e a matemática devem segundo Leibniz ser pensadas no contexto da sua filosofia e da forma como nesta a relação entre matéria e espírito é pensada. É daí que Clarke retira a sua suposição.

Na verdade Leibniz pensa que a situação imaginada por Clarke nem sequer poderia chegar a ser, pois não há Razão Suficiente para ela, e *Nihil est sine Ratione*; Leibniz não admite nenhuma réstia de ‘não-razão’ no universo. Clarke afirma não compreender os conceitos filosóficos de que Leibniz se serve na discussão (nem as mónadas, nem os princípios metafísicos (o Princípio da Razão Suficiente, o Princípio da Identidade dos Indiscerníveis)). E na verdade Leibniz utiliza o Princípio da Razão Suficiente para evitar admitir o que seria uma consequência da sua teoria relacional da espacialidade. Muito mais haveria de interessante a analisar nesta controvérsia<sup>56</sup> - aqui convém reter que quando Kant defende que o espaço e o tempo são ‘formas da intuição’ ele está a dizer que o espaço e o tempo não são nem realidades absolutas newtonianas nem entidades relacionais leibnizianas.

<sup>56</sup> Para uma mais profunda análise acerca deste debate, cf. F. Gil, *Provas, A suposição de Samuel Clarke*, Lisboa, INCM, 1986.

**CONHECIMENTO**



## 1. Compreender os problemas filosóficos - será que a história da filosofia é essencial? A filosofia, o subjectivo e o objectivo.

No seu livro *O que quer dizer tudo isto*<sup>58</sup>, que é uma introdução aos problemas da filosofia, Thomas Nagel<sup>59</sup> parte do princípio de que é impossível compreender os textos dos filósofos contemporâneos, usualmente escritos numa linguagem técnica e especializada, sem se compreender os próprios problemas com que estes se debatem. Nagel defende que para uma tal compreensão a história da filosofia não é essencial: o que importa são os próprios problemas e não a erudição. Se um problema filosófico é um verdadeiro problema, ele será um problema para nós, no presente, e deverá poder ser analisado sem qualquer referência histórica erudita.

<sup>57</sup> Este texto tem constituído material de apoio da cadeira de Filosofia do Conhecimento da Faculdade de Letras da Universidade do Porto (1º Ano da Licenciatura em Filosofia). Ao longo de mais de uma década de ensino introdutório de epistemologia os objectivos gerais da disciplina mantiveram-se: trata-se de conhecer e utilizar o vocabulário técnico da epistemologia contemporânea e de caracterizar a natureza da mente e do conhecimento. Quanto aos objectivos específicos, eles têm sido os seguintes: Propôr definições de 'epistemologia'; Caracterizar o mental a partir das noções de crenças e qualia; Compreender a definição tripartida de conhecimento (Platão, Teeteto); Compreender que a definição tripartida mostra que a epistemologia lida com questões psicológico-cognitivas, normativas e lógico-metafísicas; Definir crença (a partir de intencionalidade (aboutness) e pretensão de verdade (truth-claim)); Definir 'justificação'; Definir 'verdade' (a partir de Aristóteles, Metafísica); Identificar teorias da verdade; Identificar fontes de crenças; Dar exemplos de crenças provenientes das diversas fontes de crenças; Identificar problemas específicos relativos a cada fonte de crenças; Explicar a ideia de 'teoria da justificação epistémica'; Definir 'fundacionalismo'; Definir 'coerentismo'; Definir 'fiabilismo'; Caracterizar os Problemas de Gettier; Relacionar os problemas de Gettier com a definição tripartida de conhecimento; Criar exemplos de Problemas de Gettier; Definir 'cepticismo'; Identificar e caracterizar diferentes tipos de cepticismo (clássico, cartesiano, humeano-falibilista); Compreender a ideia quineana de 'epistemologia naturalizada'; Distinguir o solipsismo do cepticismo quanto ao mundo real exterior; Caracterizar o solipsismo; Definir 'realismo' e 'idealismo'; Defender uma posição quanto ao problema céptico acerca do mundo real exterior; Expor o problema das outras mentes; Explicar a relação entre o problema das outras mentes e os casos de qualia invertidos e zombies; Defender uma posição perante o problema das outras mentes; Expor o problema mente/corpo; Identificar posições dualistas e fiscalistas perante o problema mente/corpo; Defender uma posição perante o problema mente/corpo. Quando a introdução geral à teoria da mente e do conhecimento regida por estes objectivos é concluída, segue-se uma introdução histórica, com o objectivo de fornecer aos estudantes uma perspectiva cronológica dos grandes marcos da história da epistemologia

<sup>58</sup> O livro referido constitui leitura obrigatória da cadeira de Filosofia do Conhecimento I.

<sup>59</sup> Thomas Nagel (n.1937) é professor de Filosofia e de Direito na Universidade de Nova Iorque (NYU) e autor de *The Possibility of Altruism* (1970), *Mortal Questions* (1979), *The View from Nowhere* (1986), *Equality and Partiality* (1991) e *Other Minds* (1996). Os seus artigos mais recentes, bem como os seus cursos, estão habitualmente disponíveis on-line nas páginas do Departamento de Filosofia da Universidade de Nova Iorque ([www.nyu.edu/gsas/dept/phil/](http://www.nyu.edu/gsas/dept/phil/)).



Nagel faz questão, em *O que quer dizer tudo isto?* de não citar um único filósofo<sup>60</sup>. No entanto, a verdade é que nas entrelinhas do texto qualquer conhecedor de filosofia reconhece ideias de Kant, Hume, Wittgenstein e vários outros autores.

A caracterização geral que Nagel faz dos problemas da filosofia em *O que quer dizer tudo isto?* é a seguinte. A filosofia analisa ideias e práticas que tomamos usualmente como garantidas quando pensamos e agimos (ideias como por exemplo ‘tempo’, ‘número’, ‘certo’ e ‘errado’, e práticas tão gerais e básicas como o conhecimento e a linguagem). Para a teoria da mente e do conhecimento, interessa-nos particularmente a análise que Nagel faz das ideias e práticas relativas ao conhecimento, à realidade, à mente e à linguagem. Esses tópicos são para Nagel casos particulares da questão mais geral que é a questão da relação entre pensamento e mundo físico, ou mais precisamente, a questão da relação entre objectividade<sup>61</sup> e subjectividade. Numa realidade sem pensamento, num mundo que não incluísse uma cisão entre pensante e pensado, não haveria lugar para questões acerca do verdadeiro, do bem, do belo, do justo, etc. A ideia de Nagel é portanto que as questões da filosofia surgem não tanto de uma tradição textual determinada como de um hiato entre o objectivo e o subjectivo que é característico do nosso mundo. Esse hiato é aliás insusceptível de ser eliminado, já que, de acordo com Nagel, nenhum tipo de conhecimento ou pensamento constituirá alguma vez uma ‘visão sem perspectiva’, ou a partir de lugar nenhum (uma ‘*view from nowhere*’), como se fosse de um ponto de vista de Deus (*‘God’s eyeview’*).

## 2. O problema epistemológico. As crenças.

O problema que nos interessa de entre os problemas da filosofia que Nagel descreve em *O que quer dizer tudo isto?* é o problema epistemológico. Este pode ser resumido através de questões como estas: Em que é que eu hei-de acreditar? e Como posso justificar as minhas crenças?

<sup>60</sup> Ao contrário do que se passa na filosofia a que nos países anglo-saxónicos se chama ‘continental’, na filosofia analítica não é em geral concedida demasiada importância à história da filosofia, i.e. a filosofia não é à partida vista como uma disciplina que parte da história e que se baseia em exegese e comentário textual. Isto não significa obviamente que a história da filosofia não é praticada mas apenas que ela não aparece como um fim em si, sem necessidade de justificação.

<sup>61</sup> A objectividade é, para Nagel, uma estratégia cognitiva, algo que está ‘do lado’ do pensamento, não sendo simplesmente identificável com ‘realidade’ ou ‘mundo físico’. A diferença entre subjectividade e objectividade mede-se em termos do maior ou menor preendimento ao ponto de vista local com que todo o pensamento começa e do maior ou menor avanço na prossecução de uma ‘*View From Nowhere*’, uma visão a partir de lado nenhum que seria o ideal (nunca de facto alcançado) da objectividade.

A partir de agora considerarei que o problema epistemológico é o problema da justificação das crenças<sup>62</sup>. Para compreender uma tal descrição do problema epistemológico será necessário saber o que é crença e o que é justificação. Numa primeira descrição, podemos dizer que crenças são estados mentais com conteúdo, estados dirigidos (sobretudo, mas não apenas) para fora da mente. Esta característica das crenças (o facto de elas serem assim ‘dirigidas-a’) é a sua intencionalidade. A Intencionalidade (=aboutness) pode ser considerada como uma marca que distingue o mental por entre as coisas que existem. Repare-se que sem crenças não poderíamos viver ou agir, pois são as crenças que guiam as nossas acções. Se eu não acreditar que tenho um corpo, que o meu corpo é este, e que existem objectos exteriores a mim, em relação aos quais tenho desejos variados, não terei razão para cumprir os mais elementares movimentos no mundo, nem sequer aqueles que asseguram a sobrevivência desse corpo. No entanto, as nossas crenças não são usualmente - nem têm que ser - explicitadas ou explícitas para nós próprios. Não é necessário que as crenças sejam explicitadas ou justificadas para funcionarem como crenças, permitindo-nos viver e agir. Ter uma crença não é portanto a mesma coisa que saber que se tem uma crença, nem é a mesma coisa que poder dizer que se tem exactamente essa crença e que se sabe como justificá-la (justificar uma crença é dar razões, enumerar provas). Embora as minhas crenças sejam de certa maneira pretensões de verdade (se eu acredito alguma coisa acredito

<sup>62</sup> Este cruza-se com questões relativas a fontes de crenças. Para uma primeira classificação de fontes de crenças, cf. o quadro abaixo.

<b>Crenças de percepção</b>	Aquela mesa é castanha Este rebuçado é doce Um pássaro passa, a voar
<b>Crenças de memória</b>	Aquela mesa era castanha Aquele rebuçado era doce Há dez anos atrás eu ia às aulas de Filosofia do Conhecimento na FLUP Estive em Veneza em 1999
<b>Crenças de raciocínio (a priori e indutivo)</b>	A mesa é mais alta do que a cadeira; logo, a cadeira é mais baixa do que a mesa Vejo a Joana passar por baixo da minha janela às 15 horas (na 2ª, 3ª, 4ª, 5ª, 6ª, sábado, domingo); logo, todos os dias a Joana passará debaixo da minha janela às 15 horas Alguém escreve no quadro: 110011011001101100110 e pergunta 'Que algarismos se seguem a estes?' Alguém responde: 11001101100....
<b>Crenças obtidas por introspecção</b>	Eu penso que 2+2 são quatro Eu sinto-me feliz Eu sei que estou a ver uma mesa castanha Eu sei que estou a saborear um rebuçado doce Estou a imaginar uma praia nas Caraíbas com água verde e calma Estou a tentar resolver esta equação
<b>Crenças obtidas por testemunho</b>	D. Afonso Henriques foi o 1º Rei de Portugal Houve grandes inundações em New Orleans 'Vi o arguido empurrar aquele homem' (dito em tribunal por alguém; segue-se a formação da crença 'o arguido empurrou aquele homem' em várias pessoas na assistência)

que o que acredito é verdadeiro) elas não são pretensões de verdade explícitas. Isto significa que nós não podemos dar razões claras, ou provas, em relação à maior parte das coisas em que acreditamos, quer dizer, não estamos prontos para justificar racionalmente desde logo as nossas crenças de forma auto-consciente, imediata e sem problemas. Experimente-se por exemplo perguntar: Porque é que eu acredito que existe um mundo exterior ao meu pensamento? Porque é que eu acredito que as outras pessoas são seres conscientes? Porque é que eu acredito que esta mão com que escrevo é minha? Porque é que eu acredito que quando decido mover a mão para escrever a palavra 'palavra' ela me obedece? Porque é que eu acredito que 'mão' significa mão? Porque é que eu acredito que eu sou eu? (Estou a supôr que todas as pessoas que lêem estas palavras têm de facto as crenças mencionadas).

A justificação destas crenças não está pronta a ser dada a quem no-la pedir. O facto de termos uma crença não significa só por si que sejamos capazes de explicitar as razões por que a temos, i.e. que sejamos capazes de a justificar. É por isso que o problema epistemológico, o problema da justificação das crenças, é um problema posterior ao estado comum a todos os humanos e (talvez) a muitos animais<sup>63</sup> que é o estado de ter crenças. A questão (psicológica, cognitiva) das crenças, do conteúdo destas e da relação das crenças com a realidade física dos sistemas cognitivos é em grande parte tratada no âmbito da filosofia da mente. Pertencendo ou não em primeiro lugar à teoria da mente, a noção de crença é básica em teoria do conhecimento, e vai ser a partir de agora considerada como estando minimamente definida como estado mental intencional, proposição tomada como verdadeira, ou atitude proposicional.

### 3. Conhecimento.

Se posteriormente pretendermos definir o conhecimento a partir da crença, poderemos afirmar que o conhecimento (se é que o conhecimento é possível: um céptico radical diria que não<sup>64</sup>) é constituído por crenças especiais por entre as crenças. De facto, as crenças que constituem conhecimento são crenças verdadeiras e especialmente bem justificadas. A esta ideia segundo a qual a análise daquilo em que consiste o conhecimento nos conduz às crenças justificadas e verdadeiras chama-se 'definição tripartida

<sup>63</sup> Saber quais são esses animais e onde pára a aplicação da ideia de crença é um problema muito discutido na filosofia da mente contemporânea. Não parece muito problemático afirmarmos acerca de um cão ou de um gato que eles têm crenças, mas será que diríamos o mesmo de uma aranha ou de uma amiba?

<sup>64</sup> O cepticismo vai ser aqui considerado como a contestação, local (por exemplo, no campo ético) ou global, do direito às crenças.

do conhecimento'. De acordo com essa definição tradicional são condições necessárias e em conjunto suficientes para saber que p que (1) a pessoa A acredite que p, que (2) p seja verdadeira<sup>65</sup>, que (3) p seja sustentada por provas<sup>66</sup>.

A análise da justificação das crenças mostra que o facto de as crenças não serem perfeitamente explícitas nem facilmente individuáveis não é o seu único problema. Outro problema que surge mal comçamos a pensar no processo de justificação das crenças é o facto de a maioria das crenças se apoiar noutras crenças, o que faz temer uma regressão infinita da justificação. Perante a apresentação de cada crença poder-se-ia perguntar 'Porque acreditas nisso?' 'Como é que sabes?' continuando sempre a obter novas crenças que funcionam como razões para as crenças anteriores. Aparentemente, a única saída seria encontrar uma crença que não se apoiasse noutras crenças, uma crença básica, ou um tipo básico de crenças, que fossem auto-garantidas. Este é o ideal do fundacionalista. O fundacionalista olha para o conhecimento da seguinte maneira. Para ele existem crenças básicas (que são auto-garantidas) e crenças não básicas (i.e. crenças mediadas, garantidas ou justificadas por outras crenças). O empirismo clássico<sup>67</sup>, por exemplo, é um tipo de fundacionalismo<sup>68</sup>. Para o empirista clássico as crenças relativas a afecções sensoriais são básicas e auto-garantidas. A afirmação cartesiana do cogito corresponde a um outro tipo de fundacionalismo: o cogito é precisamente uma crença básica e auto-garantida.

#### 4. Mundo Real Exterior.

Para avançarmos na análise do problema epistemológico convém

<sup>65</sup> Este apelo à noção de verdade faz com que, obviamente, a discussão da definição de conhecimento se alargue à metafísica e à lógica filosófica. Para uma orientação quanto às teorias da verdade aí discutidas, cf. Paul Horwich, *Teorias da verdade*, in Branquinho & Murcho, *Enciclopédia de Termos Lógico-Filosóficos*, Lisboa, Gradiva, 2001.

<sup>66</sup> A definição tripartida é posta em causa pelos chamados 'problemas de Gettier', que são casos de crenças verdadeiras e justificadas que não considerariamos de boa vontade conhecimento. Considere-se o seguinte exemplo: Henry está a ver televisão numa tarde de Junho. Assiste à final masculina de Wimbledon e, na televisão, McEnroe vence Connors; o resultado é dois a zero e match point para McEnroe no terceiro set. McEnroe ganha o ponto. Henry crê justificadamente que: (1) acabou de ver McEnroe ganhar a final de Wimbledon deste ano, e daí infere sensatamente que (2) McEnroe é o campeão de Wimbledon deste ano. No entanto, as câmaras que estavam em Wimbledon deixaram de funcionar e a televisão está a passar uma gravação da competição do ano passado. Mas, enquanto isto acontece, McEnroe está prestes a repetir a retumbante vitória do ano passado. Portanto a crença (2) de Henry é verdadeira e ele tem decerto justificação para crer nela. No entanto dificilmente aceitaríamos que Henry sabe que (2) McEnroe é o campeão de Wimbledon deste ano.

<sup>67</sup> Cf. autores como John Locke (1632-1704), George Berkeley (1685-1753), David Hume (1711-1776).

<sup>68</sup> O fundacionalismo é um dos principais tipos de teorias da justificação epistémica. De acordo com o fundacionalismo, uma crença é justificada se e só se é ou uma crença básica (caso em que é auto-garantida) ou uma crença mediada que em última análise se reporta a uma crença básica. A ideia é que tem que haver um ponto de paragem da regressão da justificação de crença para crença. A 'imagem' do corpo das crenças correspondente ao fundacionalismo é a imagem da pirâmide. Segundo o coerentismo uma crença é justificada se e só se a minha teoria acerca do mundo é mais coerente com ela do que sem ela. A ideia é que não existem crenças básicas – mesmo crenças perceptivas são já conceptuais. A 'imagem' correspondente ao coerentismo é a imagem da jangada. De acordo com o fiabilismo uma crença é justificada se é produzida por processos fiáveis do sistema cognitivo. A ideia é que o conceito de justificação é em parte causal; um factor externalista tem que ser considerado.

considerar algo que não é já do nível da crença comum a todos os homens e a alguns animais. Acontece que embora as nossas crenças sejam na sua maioria acerca do mundo exterior, se pensarmos nisso podemos chegar à conclusão de que realmente nós só temos acesso directo ao interior da nossa própria mente. Quer dizer, mesmo que o meu pensamento seja acerca desta mesa na minha frente, ele é o meu pensamento, portanto eu tenho acesso à mesa 'exterior' através do meu pensamento, e não sendo a mesa ou possuindo a mesa de alguma forma em mim. O pensamento-de-mesa é acerca da mesa, não é idêntico à mesa (e, já agora, não é também acerca do cérebro nem é idêntico ao cérebro).

A primeira coisa que o facto de notar que eu só tenho acesso directo à minha própria mente (e que quando falo do que sei só posso falar em primeira pessoa) me permite, é constatar a existência do pensamento e a minha natureza mental. As coisas afiguram-se-me do seguinte modo: existe pensamento e esse pensamento é o meu pensamento (Wittgenstein<sup>69</sup> poderia dizer 'O mundo é o MEU mundo mas - estranhamente - o sujeito não pertence ao mundo', i.e. eu, enquanto sujeito pensador do mundo, não estou lá, nesse mundo).

É claro que esta não é uma coisa em que eu penso constantemente quando creio sem problemas que há uma mesa na minha frente, ou quando coloco coisas em cima dela. Para isso tenho que acreditar que a mesa está aqui e que não desaparece, que eu sou capaz de manipular estes objectos, que a mesa é capaz de os sustentar, etc. Eu penso na mesa e nos objectos, não penso que sou eu que penso na mesa e nos objectos, e que só tenho acesso à mesa a aos objectos através dos meus estados mentais. Outra das características curiosas dessa mente à qual eu tenho acesso directo (por oposição ao acesso indirecto que eu tenho à mesa ou a outros 'objectos' dos meus estados intencionais) é que por mais que ela seja difícil de capturar conceptualmente (pois afinal o que é isso de eu ser eu? O que é isso de eu ser esta mente e não outra mente ou outra coisa?), ela parece ser sempre a mesma, i.e. eu, seja eu o que fôr, não passo do estado de ser-um-eu-1 para o estado de ser-um-eu-2, diferente do primeiro.

Em suma, de tudo isto interessa sobretudo notar que embora me pareça ter consciência directa daquilo que é pensado e experienciado, de facto parece ser apenas através de estados mentais que eu penso e experienço. Por isso, eu só tenho acesso directo à mente e a nada mais. Esse acesso é além do mais privilegiado, na medida em que a mente a que eu tenho

<sup>69</sup> Cf. Wittgenstein, *Tractatus Logico-Philosophicus*. Este é o início do solipsismo, a ideia segundo a qual o meu pensamento é a única coisa que existe.

acesso directo é minha e só minha. I.e., só eu tenho esse acesso directo a essa mente (aliás nem sei se mais alguém existe...<sup>70</sup>).

No entanto, mesmo se eu apenas penso em alguma coisa através de estados mentais, eu não penso nesse estados mentais directamente (estados mentais acerca de estados mentais são estados mentais especiais, chamados estados introspectivos). Eu apenas sou consciente dos seus objectos intencionais, do seu conteúdo. Por isso não sei se esse acesso directo à mente, às coisas pensadas, é um acesso incorrigível ao modo como são verdadeiramente as coisas a cujas aparições mentais acedo. Eu teria um acesso incorrigível se fosse totalmente impossível enganar-me acerca da existência ou da natureza das coisas pelo facto de ter acesso directo aos conteúdos mentais que são acerca dessas coisas. Mas, no melhor dos casos, parece que a incorrigibilidade do acesso à mente é incorrigibilidade do acesso ao modo como as coisas aparecem à mente, e apenas isso.

Assim sendo, há aqui um outro problema: se eu só tenho acesso à minha mente, que pode ser chamada provisoriamente um 'interior mental' (o que não é idêntico a ser um interior físico, como o interior de uma caixa), como é que eu sei que as coisas pensadas - ou outras às quais a sua aparição mental corresponda - existem como coisas independentes, exteriores ao pensamento dessas coisas por mim? Como é que eu sei, em geral, que existe um mundo real exterior, 'independente e continuado'<sup>71</sup>, além da mente à qual tenho acesso directo?

É este o ponto em que o problema epistemológico conduz ao problema ontológico e metafísico, que é o problema de saber o que é que existe, o que é que há e se o que existe existe em si, de forma completamente independente do facto de ser pensado. Se o que existe existe de forma completamente independente do facto de ser pensado, nesse caso a realidade será definível como existência em si. Se de alguma maneira aquilo que existe se correlaciona com o ser pensado, a realidade será melhor definida como algum tipo de verificabilidade.

Repare-se no entanto no seguinte: é muito mais simples afirmar que esta mesa existe em si, independentemente de ser pensada, do que dizer que o passado e o futuro, ou os números, ou as possibilidades alternativas dos cursos das coisas, existem independentemente de serem pensados. Noutras palavras, a própria definição de realidade é problemática, e

<sup>70</sup> A expressão é utilizada por David Hume, no *Treatise of Human Nature* [1739].

<sup>71</sup> Classicamente é mais comum falar de 'idealismo'. São filósofos idealistas por exemplo G. Berkeley e G. W. F. Hegel [1770-1831]. Berkeley [*A Treatise Concerning the Principles of Human Knowledge*, 1710] apoiou a sua afirmação segundo a qual 'esse est percipi' [ser é ser percebido] pedindo que se tentasse formar a ideia de objectos não percebidos. Isto revela-se impossível: quando procuramos pensar uma árvore não percebida, pensamos numa imagem de árvore, e o ser da imagem supõe o perceber. Quanto a filósofos contemporâneos, um exemplo de anti-realista é o filósofo britânico Michael Dummett.

problemática de forma diferente de objecto para objecto. É desta situação que surgem as polémicas entre os realistas, que afirmam que aquilo que existe existe em si, de forma independente da mente (humana ou de quaisquer seres pensantes), e os anti-realistas<sup>72</sup>, que afirmam que o que existe depende de certa forma do pensamento. Um realista tenderá a afirmar que a realidade se alarga para além do alcance do pensável e que podem existir coisas que são inconcebíveis por nós. O anti-realista ligará a definição de realidade aos métodos de pensamento de que dispomos para abordá-la. O que há está de certa forma ligado àquilo que nós, ou os seres humanos futuros, ou outros seres, podemos conceber. Pensar em qualquer coisa que não se consegue conceber é uma ideia que não faz sentido, do ponto de vista do anti-realista<sup>73</sup>. É por isso que por exemplo Nagel afirma (1986: 91) que o anti-realismo envolve um ‘teste epistemológico de realidade’<sup>74</sup>.

A ideia de um ‘teste epistemológico de realidade’ está muito especialmente ligada ao estatuto concedido ao conhecimento científico do mundo. Nagel, que é um realista, pretende afastar o teste epistemológico de realidade e resistir à “natural tendência para identificar a ideia de mundo tal como este realmente é com a ideia do que pode ser revelado, no limite, por um crescimento indefinido na objectividade de ponto de vista” (Nagel 1986: 91).

Note-se finalmente que se, perante a questão da dependência ou independência da realidade em relação ao pensamento, argumentarmos que “Se as coisas não existissem não seriam pensadas ou percebidas”, isto não nos exime de problemas. De facto, com esse argumento estaremos a pressupor a confiança na relação entre a mente e o mundo real exterior, que é precisamente o que está em causa.

## 5. Mundo real exterior como sonho ou ilusão.

*‘We are such stuff as dreams are made on / Our little life surrounded with a sleep’,*

*Shakespeare, The Tempest, Acto IV*

Perguntar o que é a ‘realidade real’ e se alguma coisa me pareceria diferente se tudo existisse apenas na minha mente constitui uma tentação perene. Se fôr possível argumentar plausivelmente que nada me pareceria diferente se tudo existisse apenas na minha mente, então estará dado o

<sup>72</sup> A entrada em cena do ponto de vista de uma mente infinita (Deus) em muitos sistemas idealistas clássicos faz evidentemente toda a diferença aqui, permitindo congregar e i-limitar a realidade definida como correlativa do ser-pensada.

<sup>73</sup> Nagel não é um anti-realista; o que afirma acerca do anti-realismo tem uma intenção crítica.

<sup>74</sup> Esta é uma via de ligação entre o verificacionismo (frequentemente associado ao pensamento científico) e o idealismo.

passo essencial para defender que a realidade é toda ela ilusória, toda ela um gigantesco sonho bem urdido.

A ideia do carácter ilusório da realidade acompanha desde tempos imemoriais o pensamento filosófico e religioso. Platão (c. 429-347) distinguiu o Mundo Sensível (supostamente ilusório) do Mundo Inteligível (realmente real). Descartes (1596-1650) experimentou a ideia de 'totalidade da realidade como engano' através da figura imaginária do Génio Maligno. Kant (1724-1804) distinguiu o Fenómeno, ou âmbito do cognoscível, de um Númeno, que pelo menos na sua acepção positiva é um em-si pensável para além do fenómeno (o qual se torna assim aparência ou pelo menos aparição de). No budismo, a ideia de carácter ilusório do mundo é nuclear, sendo por exemplo o sofrimento humano explicado precisamente pelo prendimento dos homens a essa ilusão. A libertação (nirvana) supõe o desprendimento da ilusão do mundo, inclusive o desprendimento em relação a um eu distinto do resto do mundo, já que a separação do eu faz parte faz parte da ilusão do mundo. O par Vontade / Representação com o qual A. Schopenhauer (1788-1860) pensa a natureza do mundo está imbuído da imagética budista do carácter ilusório do mundo (Schopenhauer chega inclusivamente a considerar que o dom filosófico possuído por alguém se revela precisamente na capacidade de imaginar por vezes que os homens e as coisas não são mais do que fantasmas ou imagens de sonhos). O par conceptual que rege a interpretação estética da existência proposta por F. Nietzsche (1844-1900) - o par Apolíneo / Dionísio - é por sua vez inspirado em Schopenhauer, e portanto indirectamente inspirado pelo budismo e pela ideia de ilusoriedade da realidade (é certo que Nietzsche subverte totalmente as recomendações morais budistas perante o carácter ilusório da realidade). Os exemplos poderiam ser multiplicados, pois possivelmente não existiu um filósofo que não tivesse sido tentado de alguma forma pela distinção aparência/realidade.

## **6. Aparência e realidade.**

Quanto a esta distinção, é preciso notar o seguinte: para podermos fazer sentido da distinção aparência / realidade temos que poder comparar 'o-que-corresponde' com 'o-que-não-corresponde' à 'verdadeira realidade'. Ora, é impossível comparar coisas se não existem duas coisas para ser comparadas. E a própria ideia de contraste ou comparação supõe a possibilidade de estarmos simultaneamente, ou pelo menos alternadamente, dentro e fora da nossa mente, para termos acesso à aparência e à verdadeira realidade, a qual justificará o estatuto de aparência da aparência. Ora, isso não parece possível.



A ideia segundo a qual não podemos alcançar cognitivamente a verdadeira realidade pelo facto de estarmos cognitivamente limitados a aparências contraditórias, entre as quais é impossível decidir, é o motivo céptico clássico. O que importa compreender no motivo céptico é o seu inapelável compromisso com uma determinada ideia de realidade (no caso clássico, com a possibilidade de distinção entre realidade e aparência). Nesse compromisso decide-se a prioridade e poder que se concede ao pensamento na concepção geral de realidade.

O cepticismo clássico<sup>75</sup>, grego, constitui-se em torno da procura de um critério de verdade. Os cépticos chegam à conclusão de que esse critério não existe e que portanto não é possível sustentar crenças racionais acerca da verdadeira realidade. Como outras escolas filosóficas helenísticas, o cepticismo clássico era mais propriamente uma filosofia de vida, uma ética, do que uma investigação cognitiva sistemática. Precisamente, o cepticismo clássico conduzia à abdicação da investigação cognitiva, por tomar como prioridade a procura da felicidade (que os cépticos definem, de modo semelhante aos estoicos e epicuristas, como 'apatia' - ausência de paixões - ou 'ataraxia' - imperturbabilidade). De facto, dada a inexistência de critério de verdade, a investigação cognitiva constitui um obstáculo para essa felicidade, ao ser fonte de perturbação.

Nos *Esboços Pirronistas*, o céptico Sexto Empírico define o cepticismo como uma habilidade mental (e não uma doutrina) pela qual devido à equivalência das razões opostas numa controvérsia se atinge um estado de suspensão mental (epochê), que é um estado de não pronunciamento, e em seguida a um estado de imperturbabilidade ou quietude (apatia ou ataraxia) identificável com a felicidade (eudaimonia). Este cepticismo é assim uma técnica para um ideal de vida. Supostamente o (actual) céptico lançara-se, como outros, na procura da verdade, identificada com 'a realidade por detrás das aparências', nada encontrando. A proposta céptica era então que isso que se procura só se encontra quando se deixa de procurar (esta ideia era ilustrada pela história do pintor Apeles que queria pintar a espuma na boca do cavalo e que, desesperado por não conseguir fazê-lo, atira com a esponja, conseguindo assim finalmente o efeito pretendido).

O que importa aqui sublinhar é o facto de o cepticismo clássico estar comprometido com uma distinção aparência/realidade que é ela própria problemática. A estratégia céptica clássica era justificada pelo facto de as

<sup>75</sup> O cepticismo é uma escola filosófica helenística, cujo fundador teria sido Pirro de Elis [c. 365 - 275 a.c.]. Certos textos cépticos, nomeadamente através das compilações de Sexto Empírico (sec-II/III d.c), tiveram um grande impacto na filosofia moderna [os *Esboços Pirronistas* de Sexto Empírico foram traduzidos em 1592].

aparências serem relativas, e pelo facto de ser impossível decidir fundamentadamente, entre aparências contraditórias, qual delas corresponderia à verdadeira realidade. Daqui os cépticos concluíam que não se conhece. No entanto, poder-se-ia igualmente questionar o sentido da ideia de ‘verdadeira realidade’ por detrás das aparências. Sem esta ideia, o movimento céptico de abdicação do inquérito deixa de parecer correcto.

## 7. Solipsismo e Cepticismos.

Se eu achar que posso ser céptico ao ponto duvidar de tudo o que está fora da minha mente, retendo agora apenas as aparências para mim, obtidas pelo meu acesso directo ao interior da minha mente, e deixando de lado o aspecto dialéctico da oposição de aparências para diferentes sujeitos característico do cepticismo clássico, chegarei à posição solipsista. O solipsismo corresponde à crença segundo a qual eu sou o único sujeito de experiências. Em tudo o que existe, só há um lugar de pensamento (‘aqui’) e só há um pensamento, o meu. Pode até ser que nada exista de exterior ao meu pensamento. Assim, reduzindo os meus compromissos cognitivos àquilo de que posso estar absolutamente seguro, verifico que fico apenas com o que acontece no meu pensamento no momento presente.

O solipsismo envolve assim o cepticismo acerca do mundo externo. Este consiste na contestação da possibilidade de justificação de crenças acerca do que quer que seja de exterior ao meu pensamento. Percepção e ciência - os dois principais sentidos em que se fala de “conhecimento do mundo real externo ao meu pensamento” - são igualmente vulneráveis a este tipo de cepticismo, pois em ambos os casos estão envolvidas afirmações acerca do mundo exterior.

O solipsismo envolve ainda o cepticismo acerca de tudo o que não seja o momento presente do pensamento, e portanto, se for consequente, será um cepticismo acerca do passado. O filósofo inglês B. Russell (1872-1970) imaginou o seguinte caso para exemplificar esse tipo de cepticismo: o mundo poderia ter começado a existir há 5 minutos, com pessoas cheias de memórias ilusórias acerca do seu próprio passado e da história da humanidade, pessoas para quem o passado ‘existe ilusoriamente’. Como poderíamos justificar a implausibilidade dessa ideia? Do ponto de vista do solipsista, nunca poderíamos fazê-lo.

Outra consequência da restrição dos meus compromissos cognitivos ao momento presente do acontecer do meu pensamento é o cepticismo acerca das outras mentes. Podemos duvidar da existência de outras mentes: afinal só temos acesso directo à nossa própria mente, e será sempre assim, durante toda a nossa vida. Uma vez iniciado o caminho céptico,

parece difícil alguma vez dele sair. A justificação de tudo aquilo em que cremos parece frágil.

Deve haver alguma coisa de errado com o cepticismo nas suas várias formas. Talvez possamos encarar a questão da seguinte forma: o que está errado no cepticismo tem possivelmente mais a ver com os compromissos ontológicos e epistemológicos que uma posição céptica assume do que propriamente com o imperativo metodológico do cepticismo (o filósofo empirista escocês David Hume afirmou que o cepticismo como método é o único estado possível de uma vida racional saudável). É conveniente portanto distinguir o cepticismo como assunção do carácter falível e provisório das nossas crenças (se as crenças não fossem falíveis, provisórias e substituíveis não faria por exemplo sentido a intenção de conhecer mais e melhor) do cepticismo como dúvida acerca da realidade e do conhecimento do mundo exterior, do passado e das outras mentes.

Tomando o caso do cepticismo acerca da realidade do mundo exterior, é possível argumentar que ele é sintomático de uma megalomania da razão e de uma confiança exagerada no nosso próprio domínio racional sobre as nossas crenças. Ora, as nossas crenças não estão tão submetidas à nossa vontade como o céptico acerca do mundo exterior pretende fazer crer. Talvez seja mesmo impossível acreditar genuinamente que o mundo real exterior não existe, exactamente pela razão referida no início deste texto: as nossas crenças não são absolutamente explícitas, racionais e justificadas (nem, acrescente-se, voluntárias: não podemos decidir crer que p e passar a crer que p!). Talvez se esteja a conceder um poder desmesurado ao pensamento racional ao atribuir-lhe como tarefa a fundamentação (i.e. a sustentação através de razões e provas, controladas e explícitas, a qual pode não ser conseguida) da crença na existência do mundo real exterior. Hume defendeu uma ideia semelhante: pode dar-se o caso de a nossa crença no mundo real exterior ser função da nossa imaginação, do funcionamento cognitivo normal da nossa mente, mas não fundamentada racionalmente nem submetida à nossa vontade de um modo tal que possamos suspendê-la.

O céptico está também a afirmar demais quando assume simplesmente a possibilidade de comparar uma perspectiva verdadeira com uma perspectiva ilusória sobre a realidade. Como foi dito, essa comparação envolve uma alternância entre 'estar dentro da mente a observar aparências' e 'estar fora da mente a observar a verdadeira realidade'. Esta alternância não parece possível (pelo menos sem deixarmos de existir...). É claro que negar totalmente a possibilidade de comparação redundaria em algum tipo de verificacionismo, com os respectivos riscos idealistas.

## 8. Outras mentes.

As razões que conduzem ao cepticismo acerca das outras mentes são importantes para a filosofia da mente em geral. Se eu perguntar a mim próprio ‘O que é que eu sei da outra pessoa e como é que eu sei aquilo que sei?’ verificarei que só tenho acesso (acesso indirecto, de acordo com o que foi dito atrás) a um corpo e a um comportamento, inclusivamente verbal - o qual é, de resto, especialmente importante em todo o processo - desse corpo. Posso ter acesso ao interior físico do corpo do outro, abri-lo de várias maneiras, mas nunca à sua interioridade mental, supostamente idêntica à minha. Mesmo numa operação ao cérebro de outra pessoa não se encontram ‘seres mentais’ da sua interioridade, tais como pensamentos e sentimentos, os quais no entanto eu lhe atribuo, pelo menos se lido com ela como pessoa. Entidades desse tipo só as encontro em mim, pelo meu acesso directo à minha mente.

Os filósofos chamam qualia às qualidades da experiência subjectiva, às qualidades fenoménicas tal como são experienciadas, ao facto de a vida mental ser sentida. Ora, parte do que estou a pretender fazer notar é que só existem qualia em primeira pessoa. Isso coloca vários problemas. Por exemplo, duas pessoas podem ter o mesmo vocabulário para cores e discriminar cores da mesma maneira (i.e. elas chamarão ‘vermelhas’ e ‘verdes’ às mesmas coisas nas mesmas ocasiões) e no entanto desde sempre terem visto cores invertidas, por exemplo vermelho por verde e verde por vermelho. Nesse caso, nunca se saberia.

Por outro lado, se uma pessoa tivesse um comportamento absolutamente indistinguível do de todas as outras pessoas, ela poderia ser totalmente desprovida de qualia sem que isso alguma vez se descobrisse. O primeiro caso é conhecido como o problema do espectro invertido, o segundo é o problema do zombie. A ideia de zombie é muito utilizada na filosofia da mente contemporânea, sobretudo para lidar com a questão da natureza da consciência, mas ela não é nova: por exemplo a teoria cartesiana do animal-máquina supõe precisamente que podem existir seres capazes de comportamento complexo que são totalmente desprovidos de consciência e sensibilidade. Descartes pensou assim acerca dos animais não humanos por ter ligado estreitamente todo o pensamento (mesmo o ‘sentir’) à consciência, e a consciência à alma. E Descartes não queria atribuir alma aos animais.

A ideia básica do problema das outras mentes é, como se vê, que apenas num caso (o meu) eu tenho acesso directo à correlação entre corpo, comportamento e qualia. A crença na interioridade mental de outros seres resulta sempre de uma inferência, que pode ser inválida, não fundamenta-

da. Não resulta nunca de um acesso directo. Torna-se então problemático saber onde mais existe consciência, além da minha. Será que por exemplo, ao contrário do que pensou Descartes, ela existe em animais? Mas em quais? Se há casos em que parece simples admitir que existe consciência, como nos casos de cães ou macacos, há casos em que admitir a existência de consciência parece estranho, como os casos de aranhas, moscas, ou seres unicelulares. Entre uns e outros, em que ponto ‘entrará’ em cena a consciência? E ela entrará em cena subitamente ou gradualmente? Por outro lado, porque se há-de restringir a atribuição de consciência a seres biológicos? Se a consciência é, pelo menos em nós, física e biologicamente baseada (como facilmente se comprova pondo alguém inconsciente dando-lhe uma enorme pancada na cabeça) não será possível a existência de consciência e do correspondente sentimento de si com outras bases físicas, por exemplo em máquinas computacionais complexas?

A atribuição de consciência pode pecar por excesso e por defeito e muitos debates em filosofia da mente e em ética dependem precisamente da diferença de opiniões entre as pessoas quanto ao que constitui excesso e defeito na atribuição da consciência. Pensar que a inteligência consciente é um apanágio exclusivamente humano é uma espécie de chauvinismo (a espécie de chauvinismo a que alguns chamam especismo (*speciesism*)...), e podemos ter perante essa posição a mesma indignação que hoje muitas pessoas sentem em relação à diferença estabelecida por Descartes entre humanos e animais, quando afirmou que só há consciência onde há alma, e só há alma onde há humanos. No extremo oposto desta restrição da consciência ao humano está o pampsiquismo, a posição segundo a qual todas as partes da matéria implicam ou podem implicar consciência. O pampsiquismo não tem que ser uma ideia imediatamente absurda: em versões contemporâneas ele é identificável com a defesa da ideia segundo a qual os constituintes físicos básicos do universo têm propriedades mentais, sejam ou não parte de organismos vivos. Um pampsiquismo assim é um materialismo, não reducionista, realista, correspondente à não aceitação da ideia de emergência. É uma posição que pode ajudar a conceber o aparecimento da consciência no universo, mas que é evidentemente muito perturbadora no que respeita à identidade pessoal, que é supostamente uma e una. O filósofo G. Leibniz (1646-1716), na *Monadologia*, apresentou também uma hipótese pampsiquista, embora não materialista.

## 9. O problema mente-corpo.

Não é apenas a distribuição de consciência por outras partes do mundo que não o eu e por outras mentes que não a minha que é complexa.

Mesmo em mim essa ‘distribuição’ é difícil de conceber. Como foi dito, é à minha mente que eu tenho acesso directo, e não aos fenómenos físicos que a possibilitam, nomeadamente aos eventos neuronais no meu cérebro. Eu posso saber a partir de fora, do ponto de vista da objectividade científica, o ponto de vista de terceira pessoa, que esses eventos físicos são responsáveis pelo meu pensamento, por exemplo estudando neuroanatomia e neurofisiologia, mas não tenho acesso directo a eles, nem tenho deles qualquer conhecimento incorrigível. Esta situação assimétrica constitui o problema mente-corpo (por vezes também chamado problema mente-cérebro, uma vez que o cérebro é a parte do corpo mais directamente envolvida, pelo menos nos humanos, com a existência de consciência e pensamento).

O dualismo cartesiano é um exemplo de posição perante este problema, afirmando a existência de duas substâncias de natureza irredutível, a consciência e a matéria. O monismo espinosista é outra possibilidade: Espinosa (1632-1677) defendeu que o pensamento e a matéria são dois modos dos modos infinitos da uma mesma substância única, Deus.

Contemporaneamente, pode-se considerar que a discussão do problema mente-corpo parte do seguinte ponto comum: o que acontece na consciência depende do que acontece no corpo e nomeadamente no cérebro. Isso é tomado como certo. O problema é saber se a mente consciente é o cérebro ou é diferente dele, pois o facto é que nós temos consciência do que acontece na mente, e não temos consciência do cérebro ou do que nele acontece (aliás, nunca sequer o vimos).

A conclusão mínima possível é que a interioridade mental e a interioridade física são diferentes, e que precisamos de conceber a interioridade mental para pensarmos o pensamento. Podemos fazê-lo pensando que existe uma alma ‘ligada’ ao corpo e que portanto nós somos duas coisas, ou pensando que os estados mentais são estados cerebrais. Neste último caso, seremos physicalistas, e este é o ponto de partida da maioria dos filósofos da mente contemporâneos. No entanto, é possível argumentar, como faz por exemplo T. Nagel, que o physicalismo é uma concepção incompleta de realidade, na medida em que considera a realidade como se esta fosse puro exterior, o que não é o caso. Por isso Nagel, que não admite a existência de uma alma, defende uma teoria do aspecto dual, i.e. defende que o cérebro não é apenas um sistema físico mas um sistema com aspectos físicos e mentais, nomeadamente a consciência. Este duplo aspecto da realidade não é, segundo Nagel, uma característica local, mas presumivelmente uma característica inerente aos constituintes gerais do universo e às leis que os governam. Assim, segundo Nagel, não é possível uma teoria puramente física da consciência. No entanto, a falsidade do physicalismo não requer a admissão de substâncias não físicas, por exemplo almas cartesianas, mas

apenas a ideia de que aquilo que é verdadeiro acerca de seres conscientes não é redutível a termos físicos.

### Referências

Hume, David, 1978, *A Treatise of Human Nature*, Oxford, Oxford University Press.

Nagel, Thomas, 1995, *O que quer dizer tudo isto?*, Lisboa, Gradiva.

Nagel, Thomas, 1986, *The View From Nowhere*, Oxford, Oxford University Press.

Wittgenstein, Ludwig, 1987, *Tratado Lógico-Filosófico*, Lisboa, Fundação Calouste Gulbenkian, tradução M.S.Lourenço.

## INTRODUÇÃO À TEORIA DA MENTE E DO CONHECIMENTO – PARTE II (ASPECTOS HISTÓRICOS)<sup>76</sup>

Resumo: A história esquemática da teoria da mente e do conhecimento que a seguir se apresenta - necessariamente parcial na escolha de marcos temporais fulcrais - é orientada pelas seguintes questões:

1. Que sentido(s) tem a ideia de mente e como é que esse(s) sentido(s) determinam a teoria do conhecimento?
2. Como pôde surgir historicamente a ideia de 'mente' como 'esfera interna de observação de ideias' por oposição a um mundo 'externo'?
3. Como pôde surgir a ideia de uma 'Teoria do Conhecimento' como disciplina separada e independente do próprio conhecimento (nomeadamente do conhecimento científico)?
4. Como pôde surgir a ideia segundo a qual o núcleo da filosofia seria a epistemologia (i.e. a teoria do conhecimento)<sup>77</sup>?
5. Qual foi a importância de Kant na promoção da ideia de filosofia como teoria do conhecimento? Em que consiste o programa transcendental kantiano em teoria do conhecimento?
6. Qual foi o contexto histórico da separação entre filosofia e ciências?
7. Como se deu o afastamento entre filosofia analítica e filosofia continental? Quais são as origens da filosofia analítica?
8. Sendo o empirismo clássico (dos séculos XVII/ XVIII) uma das origens da filosofia analítica, como veio a acontecer a crítica aos 'dogmas do empirismo' dentro da própria filosofia analítica?

<sup>76</sup> Os objectivos específicos da introdução à epistemologia que se procurou ao longo dos anos prosseguir a partir deste texto foram os seguintes: Conhecer as noções platónicas de mente e de conhecimento; Conhecer as noções aristotélicas de mente e de conhecimento; Conhecer os argumentos do cepticismo clássico acerca da impossibilidade de conhecer; Conhecer as noções cartesianas de mente e de conhecimento; Conhecer as noções kantianas de mente e de conhecimento; Conhecer as noções quineanas de mente e de conhecimento; Expor e discutir a noção quineana de epistemologia naturalizada em contraste com a proposta kantiana de uma 'filosofia transcendental'.

<sup>77</sup> Para traduzir esta ideia adoptar-se-á a expressão 'filosofia epistemologicamente centrada'. Cf. Rorty, *A Filosofia e o Espelho da Natureza*, Lisboa, Dom Quixote, 1988. Muitas das hipóteses históricas apresentadas no presente texto provêm desta obra de Rorty, na qual se procura mostrar que a epistemologia e os seus problemas não têm uma forma perene, antes resultam de uma constelação específica e historicamente contingente de ideias. Nomeadamente, Rorty pensa que a intuição dualista [a ideia de mente como espelho imaterial da natureza], que aparentemente todos temos, resulta de propostas intelectuais historicamente datadas, tais como a Hipostasiação dos Universais (movimento pelo qual estes são pensados como particulares imateriais e não como abstrações de particularidades) e a proposta cartesiana da Indubitabilidade como marca do mental.



9. Qual foi a importância de Quine na crítica ao empirismo?
10. O que justifica o programa quineano de naturalização da epistemologia?
11. Que importância tem o argumento wittgensteiniano da linguagem privada na teoria da mente?
12. Que argumentos apresentam respectivamente Quine e Wittgenstein a favor do holismo?

## Introdução

Ao longo da história do pensamento várias teorias da mente, espírito ou alma foram apresentadas como parte integrante de teorias do conhecimento. Frequentemente, na origem da procura do conhecimento pelo espírito foi apontada uma motivação ética. Falar-se-á, nesses casos, de ‘motores éticos’ do conhecimento. Serão apresentados neste texto apenas alguns exemplos de teorias da mente, espírito ou alma associadas a teorias do conhecimento.

### 1. Os gregos e a *theoria*.

De acordo com a teoria platónica da reminiscência, a alma faz parte do Mundo Inteligível e o seu ‘ser de outro mundo’ revela-se precisamente pela capacidade de conhecer. O conhecimento é considerado como a via para a purificação (*catharsis*), para a separação da alma (*psychê*) em relação ao corpo (*soma*), que é o túmulo ou prisão da alma neste mundo. O motor ético do conhecimento é portanto, segundo Platão (c. 429 - 347 a.C.), a purificação.

Para Aristóteles (384 - 322 a.C.), conhecer alguma coisa é conhecer a sua razão ou causa e poder assim inseri-la na ordem da expressão lógica (silogística), concebida como explicitação das razões reais. “Todos os homens têm por natureza o desejo de conhecer” - assim se inicia o Livro I da *Metafísica*. No entanto, segundo Aristóteles, aqueles por entre os homens que procuram o ‘conhecer pelo conhecer’, i.e. que procuram o conhecimento como um fim em si e não como um meio para qualquer outra finalidade, procurarão a filosofia, que é a ciência das causas primeiras e dos princípios das coisas. Aristóteles considera que o tipo de conhecimento a que chama *episteme*, um saber teórico que demonstra as coisas a partir dos seus princípios ou causas, é superior à *empeiria* (experiência) e à *technê* (a arte, o saber fazer). Esse tipo de saber, o saber teórico, a que poderíamos chamar ciência, só pôde existir, segundo Aristóteles, quando já existia quase tudo o que é indispensável ao bem estar e à comodidade. Para Aristóteles é por isso evidente que o conhecimento teórico não é procurado por qualquer interesse extrínseco: aquilo que leva os homens a

procurar conhecimento teórico não pode ser da ordem das necessidades. O que se passa é antes o seguinte: “assim como chamamos homem livre ao que pertence a si mesmo e não tem dono, da mesma maneira este conhecimento é o único entre todos os conhecimentos que merece o nome de livre” (Metafísica, Livro I). O conhecimento teórico, conhecimento das causas, seria portanto um fim em si, e nesse sentido ‘livre’, manifestação da liberdade de quem o procura. A liberdade seria portanto o motor ético do conhecimento (de tipo superior) segundo Aristóteles.

Considerando os esforços que filósofos gregos como Platão e Aristóteles fizeram no sentido de produzir uma teoria do conhecimento, R. Rorty defende no entanto, em *A Filosofia e o Espelho da Natureza*, que o problema do conhecimento que eles enfrentavam não era ‘uma forma universal e intemporal do problema do conhecimento’, mas sim um problema muito específico. O problema do conhecimento para os gregos era o Problema da Razão, um problema gerado pelo facto de aparentemente só os humanos manterem relações com o invisível, no sentido em que apenas os humanos parecem poder conceber o imutável para além da mudança, conhecer as causas das coisas e o universal para além do particular. Ora, segundo Rorty, “Nunca se teria pensado na existência de um problema da natureza da razão se a nossa raça se tivesse limitado a apontar estados particulares de coisas - prevenções de abismos escarpados e da chuva, celebração dos nascimentos e mortes individuais. Mas a poesia fala do Homem, do nascimento e da morte como tais e a matemática orgulha-se de deixar passar os detalhes individuais. Quando a poesia e a matemática alcançaram a auto-consciência (...) havia chegado a altura de dizer algo em geral acerca do conhecimento dos universais. A filosofia empreendeu o exame da diferença entre saber-se da existência de fileiras paralelas de montanhas para ocidente e saber que linhas paralelas estendidas até ao infinito nunca se encontram, a diferença entre saber-se que Sócrates era bom e saber-se o que era a bondade” (Rorty 1988: 40). Segundo Rorty, teria sido para dar resposta a estas questões que nasceu a ideia de *theoria* como visão do invisível, sendo o ‘invisível’ os objectos do intelecto. Estes estariam para a *theoria* como os objectos visíveis para os olhos físicos. A palavra *theoria* é aliás uma metáfora visual para conhecimento, uma ideia surgida por analogia com a capacidade física de ver. A ideia de *theoria* justifica a capacidade humana de pensar o geral e o universal para além do particular, de pensar o eterno para além da mudança. Esta capacidade seria, por metáfora, uma visão do invisível com os olhos do espírito.

A ideia filosófica de *theoria* como visão do invisível veio ligar-se a considerações acerca da imortalidade da alma, que eram, já muito antes, vulgares nas religiões. Essas considerações religiosas respondiam segundo Rorty a um outro problema relativo à mente ou alma, o Problema da Personalidade (e não ao muito diferente e muito mais intelectual e restrito 'Problema da Razão'). O problema da personalidade consiste em questões como: Será que eu sou mais do que carne? Será que eu sou mais do que este corpo? Será por isso que os humanos são mais dignos do que quaisquer outros seres?. Assim, a "alma, enquanto imaterial-porque-capaz-de-contemplar-universais tornou-se a resposta do filósofo ocidental à pergunta 'Porque é que o homem é único?' durante dois mil anos" (Rorty 1988:43).

É importante notar (nomeadamente para efeitos de comparação com a ideia cartesiana de mente apresentada em seguida) que nas teorias do conhecimento de Platão e de Aristóteles a percepção sensorial fica do lado daquilo que é feito pelo corpo, não envolvendo qualquer visão do invisível, ao contrário do conhecimento dos universais. A junção das percepções e das sensações ao conhecimento dos universais na mente considerada como esfera interior única dependerá da particular contribuição de Descartes para aquilo a que Rorty chama 'A invenção da mente', cujo primeiro passo, grego, teria sido a invenção da *theoria*. O núcleo da contribuição cartesiana para a ideia de mente é a ideia segundo a qual nada é melhor conhecido pela mente do que ela própria. Para Rorty, este será o ponto de junção do Problema da Consciência aos dois problemas anteriores. A consciência é qualquer coisa difícil de inserir nas teorias antigas e medievais do conhecimento, não se identificando nem com a visão dos universais, nem com a alma imortal que perdura para além do corpo.

## 2. Descartes: a consciência e as ideias. O idealismo e a modernidade.

A mente de que Descartes (1596-1650) fala é uma mente des-naturalizada. Descartes definiu a mente como consciência, imediatidade do pensamento de si a si. Pôde assim incluir entre os objectos mentais desde a "dor" até "deus" (que parentesco encontraria Descartes entre Deus, uma sensação visual, uma dor e a ideia de número, a não ser que todos são seres para a consciência subjectiva?<sup>78</sup>). É a mente enquanto consciência que constituirá a esfera das 'ideias'. Em Descartes, uma 'Ideia' é por definição 'aquilo que está em nós de tal modo que dele temos imediatamente consciência'. A ideia moderna de 'ideia' (por oposição à ideia clássica, por exemplo platónica, de Ideia como ser do Mundo Inteligível) está assim

<sup>78</sup> É Rorty quem coloca a questão deste modo.

dependente da definição da mente como consciência. As ideias existem na consciência.

O problema da definição cartesiana da mente como consciência é que a ‘consciência’, além de ser uma maneira de pensar sobre o pensamento como imediatidade de si a si e como indubitabilidade, parece ser uma característica biológica, que aproxima os homens dos animais, ao contrário do que sucedia com a ‘razão’ dos gregos, definida como capacidade de conhecer os universais e as causas das coisas, exclusiva dos humanos. Os animais não são capazes de razão no sentido grego, mas eles parecem ser dotados de consciência. Então eles terão ou não terão mente? Esta ambiguidade será a fonte de muitos dos problemas herdados da noção cartesiana de mente.

Uma vez definida a mente como arena interna e lugar das Ideias, consideradas estas como seres subjectivos, seres para a consciência, está criado o espaço para uma teoria do conhecimento centrada na procura da ‘certeza’. O motor ético da busca do conhecimento no quadro cartesiano é o ideal da certeza. Esta traz a acepção de assentimento subjectivo para a noção de verdade. Este assentimento subjectivo não é um componente necessário de qualquer ideia de ‘verdade’. A certeza é a verdade pensada como verdade para um sujeito, a verdade pensada como indubitabilidade. Esta presença do sujeito, que é a marca cartesiana, moderna, incipientemente idealista, na teoria do conhecimento e da realidade, não tem análogo nas análises platónicas e aristotélicas do conhecimento. A indubitabilidade cartesiana das Ideias já não é uma marca de eternidade, como o tinham sido os universais para os filósofos gregos, mas sim a marca de algo para o qual os gregos não tinham nome, a consciência, a presença da subjectividade naquilo que é pensado. Esta presença da subjectividade corresponde, como se pode verificar na leitura das *Meditações* cartesianas, à introdução no próprio texto filosófico de uma abordagem em primeira pessoa, uma primeira pessoa generalizável (i.e. que cada um de nós pode assumir). Este movimento, central nos textos metafísicos básicos, como as *Meditações sobre a Filosofia Primeira* (1641), coloca a subjectividade no centro do pensamento do mundo. Este é o gérmen do que se chama historicamente idealismo (uma doutrina que afirma a natureza de alguma forma subjectiva, mental ou espiritual da realidade), uma concepção que não se encontra nem nos autores gregos nem nos medievais.

Descartes decide portanto, e é essa a marca que historicamente deixa na teoria da mente e do conhecimento, que a consciência ou indubitabilidade é a essência da mente. Do *cogito ergo sum* (penso, logo existo), que dá ao meditador (o eu qualquer) a sua primeira certeza, Descartes infere *res cogitans sum* (eu sou uma essência pensante), naquele que é para muitos o mais famoso *non sequitur* da história do pensamento. Desta situação

decorrem duas consequências importantes. Por um lado, estabelece-se que a essência pensante ou *res cogitans* não é deste mundo (físico). Por outro lado, cai a distinção entre aparência/realidade ao nível do mental, uma vez que a consciência é o seu próprio aparecimento.

Com estas bases metafísicas Descartes propõe um projecto epistemológico. O cogito será a ‘fundação’ segura para todo o edifício do conhecimento, para todas as ciências. O cartesianismo é um exemplo claro de fundacionalismo (o fundacionalismo é a posição em teoria do conhecimento que vê este como uma arquitectura que deve ter uma base totalmente segura). A consciência, colocada no centro do pensamento sobre o mundo, proporciona assim, além de uma arena interna que é o âmbito das ideias, um fundamento no mundo não físico do conhecimento científico do mundo físico. Concebida como consciência, essência não física, a mente cartesiana começa a possibilitar (na medida em que lhe oferece um campo) uma teoria do conhecimento concebida como procura apriorista de fundamentos. A ideia de ideia propicia também um novo domínio para a teoria do conhecimento: o estudo do ‘véu das ideias’, da zona de mediação. As ideias são seres mediadores entre o sujeito e o mundo, seres ‘terceiros’ e a mente é um espaço interno de quase-observação dessas ideias, trazidas perante o ‘*mind’s eye*’.

3. Kant e o transcendentalismo. A Teoria do Conhecimento como análise da forma a priori. A filosofia pós-kantiana e a separação entre filosofia e ciências. A Razão e o Romantismo.

É possível defender que apenas após o período de vida de Kant (1724-1804) uma separação entre filosofia e ciências se instala mais definitivamente na vida intelectual. Homens como Aristóteles, Descartes ou Leibniz eram indistintamente cientistas e filósofos, ocupados com tópicos que hoje consideraríamos claramente científicos, tais como a classificação das espécies de seres vivos, a geometria analítica e o cálculo infinitesimal. Era precisamente o trabalho relativo a tópicos científicos semelhantes que gerava a necessidade de elaboração de uma metafísica, uma teoria geral da natureza da realidade.

A separação entre filosofia e ciências, que se pode constatar nos últimos duzentos anos, mantém como vestígio da anterior indistinção a ideia de que a filosofia é teoria do conhecimento, diferente das ciências mas relacionada com elas porque pensando sobre o seu fundamento, sobre as suas condições de possibilidade. A filosofia neste sentido é em geral concebida como Crítica, exercício de um Tribunal da Razão (ao qual se submete a razão no seu uso não apenas científico, como também moral e estético).

Este é o estatuto kantiano da filosofia. Aquilo que move o teórico do conhecimento é a crença na capacidade que a razão tem de se analisar a si própria, encontrando-se aliás a si mesma nos produtos do conhecimento, à medida que procede a essa análise.

É neste quadro de análise da razão pela razão que se situa o programa kantiano em teoria do conhecimento. Este é um programa transcendental. Chama-se transcendental não ao ‘conhecimento em primeiro grau’ (por exemplo matemático ou físico ou outro) mas à análise das condições de possibilidade do conhecimento já possuído. Essa análise faz-se segundo Kant investigando as condições a priori do sujeito. A ideia de filosofia como análise das condições a priori substitui, com Kant, a ideia de filosofia como metafísica transcendente. Para Kant a metafísica transcendente falhara ao procurar atingir de forma puramente conceptual e apriorista objectos que nunca poderiam ser dados na experiência (objectos como Deus, a Alma, o Mundo).

A teoria das condições de possibilidade do conhecimento é então para Kant uma propedêutica, ou preparação, necessária à metafísica, se é que esta é possível. Para Kant sê-lo-á, nomeadamente a partir da distinção estabelecida na teoria do conhecimento entre fenómeno (o limite para o cognoscível) e nùmeno. Será esta cisão a deixar espaço para teses essenciais à metafísica kantiana tais como ‘existem acções livres’ e ‘existe um ser absolutamente necessário’.

É importante sublinhar que o apriorismo defendido por Kant não tem que ser um inatismo: aquilo que existe a priori no sujeito, a marca do sujeito naquilo que é conhecido, não é apanágio de uma alma com ideias inatas (esta ideia é cartesiana e não tem lugar na Crítica da Razão Pura lida como teoria do conhecimento). Segundo Kant, são as estruturas a priori do espírito que possibilitam a ciência (considerada como conhecimento universal e necessário, constituído por juízos sintéticos a priori). David Hume, o empirista, teve razão, segundo Kant, ao considerar que a universalidade e a necessidade do conhecimento científico não poderiam provir da experiência. No entanto essas características estão presentes no conhecimento. E se elas estão presentes, só podem provir do sujeito. Só as estruturas a priori do sujeito podem justificar o carácter de ciência dos enunciados da ciência.

De acordo com Rorty, em *A Filosofia e o Espelho da Natureza*, é com autores como Descartes e Kant que se configura a ideia de teoria do conhecimento e aquilo a que Rorty chama a centração epistemológica da filosofia. Este acontecimento histórico estaria assim ligado a Descartes, que inventou a mente como uma esfera de interioridade imaterial desligada do mundo físico, e a Kant, que inventou o Tribunal da Razão, a filosofia definida como crítica, por oposição à ‘produção directa’ de uma metafísica.

Esta tese de Rorty é contestável. Em primeiro lugar, a conjugação dos projectos epistemológicos de Descartes e de Kant não é simples (Kant é em inúmeros aspectos um anti-cartesiano: por exemplo, e centralmente, Kant não pensa que a alma seja uma substância simples cognoscível através da auto-consciência). É contestável também na medida em que a centração epistemológica da filosofia, a consideração da ciência como centro e a maior preocupação para a filosofia, centração característica da época moderna, pode não ser apenas um acontecimento da história das ideias, e como tal um efeito vulnerável à luta das interpretações, mas também um resultado da atenção dos filósofos ao mundo moderno. De facto, a modernidade como época é marcada pelo advento da ciência físico-matemática da natureza. A ciência é o modo de abordar o mundo que se tornou civilizacionalmente dominante a partir da época moderna, tanto que foi possível chamar à civilização ocidental a 'Civilização da Ciência e da Técnica'. O advento da ciência foi certamente um acontecimento histórico contingente, i.e. algo que poderia não ter acontecido, por mais que naturalmente atribuamos à ciência, hoje, uma ligação íntima com a verdade acerca do mundo. É de resto simples ver que fazer ciência não é o único modo de abordar o mundo e de o pensar (basta considerar por exemplo o quanto uma definição religiosa de verdade propicia toda uma outra via de procura para o espírito). A ideia de ciência está portanto ligada a uma particular noção de verdade, por entre outras. No entanto, se a Modernidade é de algum modo definível conceptualmente como 'época' é porque ela representa a realização civilizacional dessa (nova) noção de verdade, ligada ao progresso e à razão, através da ciência.

O primeiro grande movimento de oposição à centralidade da razão na concepção de mundo, centralidade defendida em termos ideológicos nomeadamente pelos iluministas, foi o romantismo, que esteve ligado ao idealismo em filosofia<sup>79</sup>. O Romantismo foi um movimento cultural e artístico contra-iluminista que representou uma fortíssima influência na cultura europeia e americana nas últimas décadas do século XVIII e nas primeiras décadas do século XIX. Mais do que exclusivamente um movimento artístico, o romantismo constituiu uma certa manifestação global de rejeição do pensamento racional, cujo expoente seria o pensamento científico considerado como abordagem por excelência da realidade. A racionalidade científica é, a partir desta perspectiva, considerada como prosaica e des-'espiritualizadora' do mundo. Curiosamente, é em parte à filosofia kantiana que os românticos vão buscar a definição de realidade

<sup>79</sup> Para as ligações entre o Romantismo e o Idealismo filosófico, cf. por exemplo N. Hartmann, *A Filosofia do Idealismo Alemão*, Lisboa, Fundação Calouste Gulbenkian, 1983.

como espiritualidade. Daí passam no entanto para uma posição de certa forma pouco kantiana: a ideia segundo a qual a ‘verdadeira’ realidade não pode ser acedida de forma intelectualista mas apenas através da intuição, da emoção, do sentimento.

A base dos traços programáticos do romantismo em arte - a exaltação da natureza, a expressão dos sentimentos, da emoção, da intuição, da imaginação, o culto do exótico - é uma determinada rejeição da civilização enquanto relação racional com o mundo. Para os românticos, apenas acontecimentos de intuição e de emoção e não argumentos racionais nos darão acesso ao ‘eterno enigma do mundo’. Para fundamentar filosoficamente este programa, os românticos alemães foram buscar a Kant a ideia de númeno ou realidade em si como liberdade. Da ideia de liberdade ou vontade como realidade em si passaram à ideia de natureza como espírito, aspirando à realização. O caso mais extremo desta tendência é a filosofia hegeliana do Espírito Absoluto, realizando-se na história.

#### 4. Separação definitiva? O século XIX e a vocação histórica e existencial da filosofia.

A curta referência ao romantismo teve como intenção sublinhar que a verdadeira cisão histórica entre a filosofia e as ciências (quando considerada do lado da filosofia) se deu no período posterior a Kant, nomeadamente com o movimento contra-iluminista. Essa separação não pode ser considerada consumada no próprio Kant, por mais que o transcendentalismo acentue a divisão do trabalho entre os teóricos da forma e os teóricos do conteúdo do conhecimento (i.e. os filósofos e os cientistas). A Crítica da Razão Pura é uma obra de teoria da ciência e de teoria do conhecimento, mesmo se não deve ser lida exclusivamente como tal.

No entanto, com um filósofo como Hegel e com a sua ideia de filosofia<sup>80</sup>, a separação está consumada. Se olharmos para a história da filosofia anterior ao século XIX veremos inúmeros filósofos-cientistas. No século XIX esse ‘tipo’ intelectual começara a ser mais raro. O desaparecimento da figura, e uma certa ideia de filosofia como prática mais ou menos poética e alusiva que entrará em cena, estão na origem da cisão entre as duas grandes tradições filosóficas contemporâneas, a filosofia analítica, tradição dominante no mundo de língua inglesa, e a filosofia continental, com origens sobretudo francesas e alemãs. Sintomático da localização temporal dessa

<sup>80</sup> Para Hegel a filosofia é sistema da Razão, completude, auto-consciência alcançada pelo Espírito. A filosofia como sistema da razão opõe-se precisamente à ‘incompletude’ daquilo de que o entendimento (nomeadamente científico) é capaz.



cisão é o facto de Hegel ser considerado um grande filósofo na tradição continental, enquanto os filósofos analíticos o consideram, maioritariamente, um autor obscuro e ilegível. Kant, pelo contrário, não é objecto de um tal desdém e consitui uma referência para ambas as tradições.

Apenas na filosofia anglo-saxónica a teoria do conhecimento se manteve bastante viva no século XIX (por exemplo com autores como John Stuart Mill e C. S. Peirce). Do lado da filosofia continental, o século XIX consumou o afastamento entre a filosofia e o pensamento científico seu contemporâneo. Visto sob uma outra perspectiva, nomeadamente ética, política e existencial, o século XIX foi extremamente produtivo em filosofia: foi a época de Hegel, e da sua filosofia do Espírito que se manifesta dialécticamente na história, de Marx, o filósofo da dialéctica invertida (i.e. materialista), crítico da alienação e apólogo da revolução, de Nietzsche, com a sua genealogia da moral e a sua crítica radical da nossa civilização de ‘escravos’, de Kierkegaard e de Schopenhauer, filósofos do sentido da existência. Em suma, foi a época de uma filosofia mais ética, política e existencial do que epistemológica: nenhum destes homens foi (nem tinha que ter sido) um filósofo-cientista. Houve no entanto danos laterais do desvio existencial: na prática, a filosofia nascida desta época, na tradição franco-alemã, deu por si divorciada de um outro tipo de pensamento, que continuava a decorrer ao seu lado, e que se ia tornando social e civilizacionalmente cada vez mais omnipresente e incontornável, o pensamento científico.

A filosofia hoje chamada ‘continental’ (se por tal nome entendermos aquela que tem como figuras de referência autores como Nietzsche, Heidegger, Foucault ou Derrida) está longinquamente ligada à rejeição da imagem racionalista do mundo que se observa neste período.

##### 5. Filosofia analítica e filosofia continental: os herdeiros de Frege e Husserl.

Apesar da reportação ao movimento romântico de alguns dos factores que teriam estado na origem da cisão entre filosofia analítica e filosofia continental, de facto a separação de caminhos entre estas deu-se após o início do século XX, com os ‘herdeiros’ de G. Frege (1848-1925) e E. Husserl (1859-1938). O que é curioso é que estes dois autores, cada um deles iniciador de linhagens metodológicas que se mantiveram até hoje na filosofia continental e na filosofia analítica, nomeadamente a fenomenologia e a análise do pensamento feita através da análise da linguagem e da lógica, além de terem sido contemporâneos, ambos filósofos e matemáticos,

não se interessavam por temas muito diferentes<sup>81</sup>. Basicamente, ambos se interessavam pelo estudo dos conteúdos do pensamento, e ambos pensavam que esse estudo não deveria reduzir-se à psicologia enquanto ciência natural empírica. Mas com os seguidores imediatos de Frege e de Husserl, nomeadamente Wittgenstein (pelo menos Wittgenstein enquanto autor do *Tractatus*<sup>82</sup> e Heidegger<sup>83</sup>, o caso já é outro. É difícil ver neles uma ideia comum quanto ao que se entende por filosofia (embora seja possível ver um mínimo denominador comum no interesse de ambos pela linguagem).

Explicitando aquilo de que se fala quando se fala de filosofia analítica e continental, vamos considerar que os ‘pais’ da filosofia analítica contemporânea, i.e. as figuras individuais mais influentes na sua constituição, são, além de G. Frege, B. Russell (1872-1970) e L. Wittgenstein (1889-1951). Para além das figuras individuais, as origens desta prática filosófica podem ser encontradas no empirismo clássico, nos métodos da lógica formal e na ideia de tarefa filosófica desenvolvida no Círculo de Viena<sup>84</sup>, i.e. na filosofia considerada como análise da estrutura das teorias e da linguagem, de modo a conseguir a depuração da metafísica e a unidade da ciência. É no seio desta tradição constituída nas primeiras décadas do século XX que vêm a surgir filósofos como W.O Quine, H. Putnam, N. Goodman, D. Davidson, M. Dummett, J. Searle, J. Rawls, R. Rorty, S. Kripke e os filósofos da mente que serão estudados a seguir neste curso.

Por seu lado, a linhagem da filosofia continental pós-husserliana comporta o trabalho em filosofia contemporânea que se reconhece sob os títulos de fenomenologia, hermenêutica, estruturalismo, desconstrução, abarcando assim filósofos e teóricos tais como M. Heidegger, J.-P. Sartre, M. Merleau-Ponty, E. Lévinas, H. G. Gadamer, P. Ricoeur, M. Foucault, R. Barthes, J. Lacan, J. Derrida, etc.

A relativa artificialidade desta separação é provada pelo facto de ser muito difícil classificar como ‘analíticos’ ou ‘continentais’ filósofos contemporâneos como por exemplo J. Habermas ou R. Rorty, na medida em que vêm as ‘vantagens’ (diferentes) das duas tradições e se servem delas.

Mas tomando Wittgenstein e Heidegger, dois grandes representantes dos dois campos, considerados frequentemente os maiores filósofos do século XX, os seus casos podem servir para exemplificar o quanto o problema das relações entre a filosofia e as ciências está em jogo na distinção entre a filosofia analítica e a filosofia continental. Heidegger afirmou, e essa afirmação ficou célebre, que a ciência não pensa. Embora a afirmação

<sup>81</sup> Como por exemplo M. Dummett fez notar. Cf. Dummett 1993, *Origins of Analytical Philosophy*.

<sup>82</sup> Será referido como Wittgenstein I.

<sup>83</sup> Heidegger foi de facto discípulo de Husserl. Entre Wittgenstein e Frege a ligação é menos directa.

<sup>84</sup> Wittgenstein não era membro do Círculo de Viena mas foi inspirador do movimento.

tenha sido bastante ridicularizada (do lado da filosofia analítica, evidentemente), ela não é imediatamente absurda. O que é certo é que Heidegger não era nem queria ser um epistemólogo, e por isso não abordava a ciência como epistemólogo, i.e., procurando compreender a ciência como conjunto de afirmações sobre aquilo que a realidade é. Se Heidegger se interessa pela ciência é para a pensar como acontecimento civilizacional, um acontecimento civilizacional que ele considera ocultante. Aquilo que interessa Heidegger enquanto filósofo é o sentido do ser, a história do Ocidente, a crítica da modernidade, o destino da civilização. É a partir de este ponto de vista, que visa compreender as maneiras humanas de existir e a sua contingência, que se torna mais clara por exemplo outra célebre (e na altura atacada com indignação) afirmação de Heidegger (da *Introdução à Metafísica*, 1953): “De um ponto de vista metafísico, a Rússia e a América são iguais, o mesmo triste frenesi tecnológico, a mesma regulamentação sem restrições do homem comum”.

O que importa é que para Heidegger, as ciências são apenas ‘um modo de ser da existência’. Embora Heidegger não apreciase o epíteto ‘existencialismo’ o que está aqui em jogo é uma ideia central do existencialismo: o mundo não está aí só para, ou principalmente para, ser conhecido mas para ser ‘lidado’ (é claro que mesmo do ponto de vista da teoria do conhecimento, e não do ponto de vista da avaliação das civilizações, é possível dizer algo de semelhante: é o que fazem todos os pragmatistas, por exemplo<sup>85</sup>).

Ao contrário de Heidegger, que chega à filosofia através da teologia (Heidegger chegou a pensar enveredar por uma vida religiosa, e fez estudos nesse sentido), Wittgenstein começou por estudar engenharia e aeronáutica e chegou à filosofia através da lógica e da matemática. O *Tractatus Logico-Philosophicus* resulta dos estudos de Wittgenstein, feitos com B. Russell, sobre os fundamentos da lógica e da matemática, e a linguagem é o seu tema essencial (o que não impede que na obra se trate também daquilo que é ‘da ordem do místico’, e do lugar das questões éticas e estéticas). No Wittgenstein I<sup>86</sup>, a lógica é o denominador comum das linguagens e forma o quadro de estruturação do conhecimento do mundo. Por isso a análise das propriedades das proposições da lógica será a via para a teoria acerca do mundo.

Wittgenstein não é um representante típico do naturalismo epistemológico assumido hoje por grande parte dos filósofos analíticos, e o seu

<sup>85</sup> Esta semelhança pode começar a explicar porque é que um autor como Rorty gosta de retratar Heidegger como um pragmatista anti-fundacionalista. Cf. R. Rorty, *Contingency, Irony, Solidarity*, Cambridge, CUP, 1990.

<sup>86</sup> O Wittgenstein do *Tractatus* (1921) por oposição ao Wittgenstein das *Investigações Filosóficas* (1953), a que se chamará Wittgenstein II.

interesse em ciência resume-se à lógica, à matemática (e à psicologia, embora entendida de maneira muito própria). Mas é de qualquer modo muito grande a diferença entre o interesse de Wittgenstein pela lógica enquanto quadro do mundo, e o seu método de análise em filosofia desenhado de forma pelo menos semelhante aos métodos de investigação em ciência, e o interesse de Heidegger no ser e na civilização praticado através de um método de alusão que se afasta radicalmente dos métodos do inquérito científico.

## 6. Empirismo e crítica ao empirismo dentro da filosofia analítica.

O empirismo clássico (de J. Locke, G. Berkeley e D. Hume) é uma das origens, e uma origem sempre revisitada, da tradição filosófica de língua inglesa. No entanto o empirismo que mais directamente marcou a filosofia analítica contemporânea foi o empirismo (ou positivismo) lógico do Círculo de Viena.

O Círculo de Viena foi um movimento intelectual dos anos 20-30 do século XX protagonizado por um grupo de filósofos, cientistas e matemáticos, entre os quais se encontravam M. Schlick, R. Carnap, H. Feigl e O. Neurath. As ideias associadas ao Círculo de Viena são uma das raízes da filosofia analítica contemporânea, e constituíram uma influência dominante nos países anglófonos <sup>87</sup>(11) até aos anos 60. O positivismo lógico é 'positivista' porque considera que as ciências são a única via para o conhecimento propriamente dito: qualquer incursão 'para além' dos limites e dos métodos da ciência arrisca-se a ser cognitivamente vã, sem sentido, e o discurso produzido arrisca-se a ser insusceptível de verdade ou falsidade. Para os positivistas lógicos seria esse o caso de grande parte da metafísica tradicional. ('Metafísica' pode aqui ser definida como a tentativa, que ocupou os filósofos durante séculos, de falar directamente da natureza de deus, da alma e do mundo). O positivismo lógico é 'lógico' porque a definição daquilo que segundo este grupo se deve fazer em filosofia dependeu de desenvolvimentos na lógica e na matemática, e nomeadamente de criações e inovações do âmbito da lógica formal. Os instrumentos lógicos seriam utilizados de acordo com a ideia de filosofia dos positivistas lógicos com o fim de atingir o objectivo da empresa intelectual: a unidade da ciência e a depuração da metafísica. Neste contexto, a tarefa da filosofia é a análise da estrutura das teorias e da linguagem.

<sup>87</sup> O movimento inicia-se no mundo 'germânico', mas devido à 2ª Grande Guerra, a maioria dos participantes emigrou para a Grã Bretanha e para os Estados Unidos. Nos Estados Unidos a influência de R. Carnap foi particularmente importante.

Uma doutrina marcante do positivismo lógico foi a doutrina acerca daquilo que é cognitivamente significativo. Interessa por isso compreender o critério de significação dos positivistas lógicos. (Assume-se que aquilo a que chamamos conhecimento toma a forma de ‘frases’, asserções, declarações de que  $p$  é o caso).

Critério de significação: uma frase é cognitivamente significativa SSE é ou analítica ou em princípio empiricamente verificável ou falsificável.

Este critério está sujeito a contestações várias, mas convém notar antes de mais que ele está moldado de modo a incluir como cognitivamente significativas frases aparentemente vazias de conteúdo empírico mas que são importantes para o conhecimento, nomeadamente as frases lógicas. Estas não são verificáveis nem sequer em princípio, pois não têm conteúdo empírico. No entanto não é possível pensar conteúdos empíricos, nomeadamente científicos, sem estruturas desse tipo. Sob a influência do *Tractatus* de Wittgenstein, as verdades analíticas necessárias serão consideradas tautologias. Tautologias não são frases acerca do mundo, não são acerca de relações que se sustentam independentemente do pensamento em qualquer domínio de objectos. A sua verdade é formal, vazia: elas são verdadeiras em virtude da sua significação e não de qualquer conteúdo empírico.

Sob a influência do positivismo lógico, a filosofia torna-se sobretudo teoria do conhecimento, e a teoria do conhecimento envolve a filosofia da ciência e a filosofia da linguagem, uma combinação de princípios empiristas com análise lógica.

No entanto, o critério de significação dos positivistas lógicos leva a crer que podemos separar, nos enunciados da cognição, os momentos em que ‘respondemos’ à linguagem (seria o caso das frases analíticas) dos momentos em que ‘respondemos’ à experiência (seria o caso das frases sintéticas, com conteúdo empírico, verificáveis). É esta forma de conceber a relação entre linguagem e experiência que Quine irá atacar, num célebre artigo intitulado *Two Dogmas of Empiricism* (1953).

7. Quine, a crítica à distinção analítico-sintético e a naturalização da epistemologia.

W.V. Quine (1908-2000) foi um filósofo e lógico americano, e certamente, pelo menos para quem se interessa por epistemologia, um dos mais importantes filósofos do século vinte. Foi aluno de Carnap (um dos membros do Círculo de Viena) e produziu trabalho importante em lógica e na fundamentação da teoria dos conjuntos. Os seus livros, nomeadamente

*From a Logical Point of View* (1953), que inclui o artigo *Two Dogmas of Empiricism*, e *Word and Object* (1960), tiveram uma enorme influência na teoria da mente e do conhecimento de tradição analítica.

O artigo *Two Dogmas of Empiricism* é um texto incontornável na história da teoria do conhecimento. A ideia de separação entre linguagem e experiência, que será discutida em *Two Dogmas*, e que está presente no critério de significação do positivismo lógico, é mais um tipo de separação entre forma e conteúdo, como já tinha sido visto em Kant, simplesmente agora reformulado em termos de linguagem. A ideia de separação forma / conteúdo é portanto, note-se, comum às teorias do conhecimento de Kant e dos positivistas lógicos, e ela é, aliás, extremamente útil quando se trata de enunciar a tarefa da teoria filosófica do conhecimento. De facto, a separação entre forma e conteúdo coloca a teoria do conhecimento como análise da forma, seja definida como estruturas a priori do sujeito, seja como linguagem. A filosofia estaria assim para a ciência como o estudo da forma para o estudo do conteúdo.

A utilidade de uma ideia não é no entanto suficiente para a justificar, e é precisamente sobre a defesa da impossibilidade da separação forma/ conteúdo no conhecimento que se ergue a ideia quineana de naturalização da epistemologia. De acordo com a ideia de epistemologia naturalizada, quando se trata de estudar o conhecimento, a filosofia e a ciência não podem ser separadas: elas só poderiam ser separadas se fosse possível estabelecer uma separação clara entre a forma e o conteúdo. Segundo Quine não é possível estabelecer essa separação e um passo importante para a justificação desta tese foi a oposição de Quine ao critério de significação dos positivistas lógicos, e a defesa da impossibilidade de uma distinção de natureza absoluta entre enunciados analíticos (verdadeiros em virtude da sua significação) e enunciados sintéticos (verdadeiros com base nos factos). Esta distinção é o primeiro dogma identificado e rejeitado em *Two Dogmas*.

A crítica de Quine em *Two Dogmas* pretende em grande medida mostrar que a teoria da significação dos positivistas lógicos não se sustenta, pois supõe uma visão errada das relações linguagem/experiência. A argumentação de Quine em *Two Dogmas* é, muito resumidamente, a seguinte.

O empirismo (lógico e clássico) assenta sobre 2 dogmas: (1) a separação dos enunciados que exprimem conhecimento em analíticos e sintéticos e (2) o reducionismo (i.e. a ideia segundo a qual existem relações 'atómicas', um a um, entre frases e mundo, e portanto a convicção de que toda a enunciação significativa seria traduzível numa enunciação acerca da experiência imediata). Quine afirma que a distinção analítico/sintético -

- que vem de longe, tendo sido defendida por Hume, Leibniz e Kant, por exemplo, e não apenas pelos positivistas lógicos - se sustenta se e só se o 'reducionismo radical' se sustentar. Quine pensa que ele não se sustenta. Grande parte do artigo é dedicado à proposta de uma visão das relações entre crenças e experiência que se erga como alternativa ao reducionismo. Essa visão é uma visão holista, expressa pela seguinte imagem: a totalidade do nosso conhecimento é como um tecido feito pelo homem que toca na experiência apenas nos bordos; por isso, nenhuma crença comparece sozinha ao tribunal da experiência. Assim, "há muita latitude de escolha quanto a qual crença re-avaliar à luz de uma única experiência contraditória"<sup>88</sup>. Quine pensa até que "qualquer asserção pode ser mantida como verdadeira aconteça o que acontecer se fizermos ajustes suficientemente drásticos noutra parte do sistema". Deste modo, "nenhuma asserção é imune à revisão" (Quine 1953:43). Esta última afirmação abre a possibilidade de mesmo revisões em lógica serem possíveis para manter a harmonia da relação entre crenças e experiência. Ora, por mais que os empiristas anteriores acentuassem a visibilidade das crenças em virtude da experiência, a lógica sempre fora 'intocável'.

Resumindo: Quine pensa que os dois dogmas têm a mesma raiz, que é a crença na possibilidade de separação do componente linguístico (daí a ideia de análise de significados) e do componente factual (daí a ideia de verificação) nas frases que expressam conhecimento. Com esses dois dogmas, tem-se uma teoria empirista do conhecimento e da racionalidade, segundo a qual o que acontece quando pensamos é que ou nos movemos só em pensamento, ou apelamos à experiência. Efectuamos revisões de crenças quando a experiência contradiz as crenças anteriores, enquanto que do lado do analítico nada de novo pode provir. Mas se os dogmas são dogmas e como tal devem ser abandonados, esta é uma má teoria da racionalidade, e é isso que Quine defende. A situação segundo Quine é então a seguinte: se no conhecimento não é separável claramente o componente factual do componente linguístico, na medida em que as crenças são duplamente dependentes da linguagem e da experiência, a maneira correcta de conceber o conhecimento e a racionalidade é holista e pragmatista. Não se trata de rejeitar o empirismo (Quine é um empirista) e nem sequer o verificacionismo mas sim de adaptar ao empirismo a ideia segundo a qual não existe (1) imediatidade com significados nem (2) relação directa entre frases isoladas e estados de coisas.

*Two Dogmas* termina com a apresentação de um quadro ainda empirista, embora diferente do empirismo de Hume ou dos positivistas lógicos,

<sup>88</sup> Quine, *Two Dogmas of Empiricism*, in *From a Logical Point of View*, Cambridge MA, Harvard UP, 1953, pp. 42-43.

da maneira como as nossas crenças se relacionam com a experiência: num quadro empirista holista e pragmatista, nenhuma crença é verificável isoladamente e nenhuma crença é imune à revisão. Esta posição quanto ao conhecimento é holista porque dá prioridade ao todo (a totalidade da ciência) em relação às partes (as crenças isoladas). Em termos de conhecimento, o holismo redundante na afirmação segundo a qual a experiência serve de teste a corpos de teoria e não a crenças isoladas. Uma hipótese *h* não pode pois ser confirmada ou infirmada isolada do corpo da teoria. A posição que Quine defende é pragmatista porque supõe que a crença em verdades tem uma relação inabdicável com o sucesso da acção de entidades no mundo: o conhecimento é um instrumento para a previsão prática e não um armazenamento numa interioridade mental protegida e desligada de interesses corpóreos, mundanos.

Abandonar os dogmas do empirismo terá ainda segundo Quine o efeito de esbater as supostas fronteiras entre metafísica especulativa e ciência natural.

#### 8. Wittgenstein, o interior e a linguagem privada

A posição de Quine em *Two Dogmas of Empiricism* constitui um ataque à ideia de dado, à ideia de imediatidade quer com significados quer com factos isolados. Além de Quine, também Wittgenstein desenvolveu ideias importantes para a oposição ao Mito do Dado (a expressão é de W. Sellars), nomeadamente com os seus argumentos acerca da impossibilidade de uma linguagem privada nas Investigações Filosóficas.

O problema de Wittgenstein diz respeito à maneira como usamos linguagens para exprimir, transportar, partilhar, publicar, conhecimento e coisas mentais. A ideia, muito comum, quanto ao que fazemos quando pensamos e falamos e que Wittgenstein pretende rejeitar é a seguinte:

Ideia: Existe primeiro o meu mundo mental, cujos eventos são directamente conhecidos por mim. Depois, escolho voluntariamente as palavras para exprimir esses eventos mentais. Depois transmito através dessas palavras os eventos às outras pessoas.

A concepção de linguagem descrita no exemplo assim não se sustenta porque:

1. Não há significados intrínsecos em coisas de linguagem
2. O significado é o uso
3. O uso é uma prática pública



4. Assim sendo, para explicar o significado de uma palavra temos que olhar para além do indivíduo isolado
5. Em questões de significação temos que olhar para a comunidade
6. Sem comunidade não existiria linguagem
7. O solipsista não poderia ter uma linguagem privada
8. O solipsista não poderia descrever os acontecimentos da sua vida mental com uma linguagem privada

Daqui se segue para Wittgenstein que:

9. Não pode existir o solipsista
10. Todas as linguagens possíveis são públicas
11. Mesmo os acontecimentos das nossas vidas mentais (ex: falar acerca de dores ou outras sensações) são “vertidos” em linguagem pública

Wittgenstein critica portanto a ideia de que o fundamento da significação, do conhecimento, reside na experiência privada. Isso não é verdade porque não há possibilidade de fazer sentido das noções de ‘dados puros da experiência subjectiva anteriores à linguagem’, ou de ‘linguagem privada’. Linguagem privada seria aquela que eu compreendo e mais ninguém compreende. A ideia de linguagem privada corresponde à figura do solipsista. O solipsista adquiriria a sua linguagem a partir do seu próprio caso: por exemplo, “apontaria” interiormente para a sua sensação de dor, chamar-lhe-ia “dor” e a partir daí a dor seria chamada “dor”.

Não é assim, no entanto, que as palavras significam: o facto de as palavras significarem envolve trabalho preparatório. Este trabalho é da comunidade, e não é um trabalho intelectual, é uma mistura de significações com acções. I.e, o funcionamento de uma linguagem como linguagem com significados determinados não é separável de formas de vida. O solipsista nunca poderia por isso desenvolver uma linguagem privada. É isto que Wittgenstein procura mostrar nas *Investigações Filosóficas* (§ 243 - § 315): que não são possíveis linguagens privadas, só conhecidas do falante e usadas para dar forma exterior a ideias previamente definidas na sua mente. As coisas de linguagem são primeiro públicas, estão primeiro fora. Para percebermos a nossa compreensão dos sentidos de palavras temos que perceber que as palavras e as coisas estão misturadas quando conhecemos e agimos. E isto é assim mesmo quando se trata da nossa experiência interior. Retomando o exemplo de “dor”, “dor” é um nome para a minha dor, a tua dor, a dor de hoje, a dor de ontem, a dor de cabeça, a dor de dentes.

Há muito pouco de privado e íntimo neste funcionamento de “dor”. E sem uso de “dor”, eu não nomeio ou aponto a minha dor.

O solipsista (se ele pudesse existir...) “pensaria” que há um grande hiato entre o nosso comportamento, incluindo o uso de linguagem, e a vida mental interior, privada. Wittgenstein pensa que não é assim: a análise do funcionamento da linguagem mostra a razão por que não podemos ser cartesianos, interioristas obcecados com a certeza subjectiva.

## 9. Nota final

As ideias acerca da mente e do conhecimento analisadas neste texto, especialmente as referentes à história recente da filosofia analítica, constituem o pano de fundo para a emergência da filosofia da mente, que será tratada em seguida. A história tal como pode ser contada do lado da filosofia não é no entanto a única preparação do campo da filosofia da mente. No campo da filosofia da mente vão convergir não apenas desenvolvimentos filosóficos, mas também ideias surgidas nos campos da lógica e da psicologia, assim como nas ciências cognitivas, uma área multidisciplinar que começa a tomar forma nos anos 70 e que congrega as neurociências, a Inteligência Artificial, a psicologia cognitiva, e em geral todo o inquérito racional sobre a cognição<sup>89</sup>.

## Referências

Dummett, Michael, 1993, *Origins of Analytical Philosophy*, Cambridge MA, Harvard University Press.

Kant, I., 1989, *Crítica da Razão Pura*, Lisboa, Fundação Calouste Gulbenkian, trad. A. Morujão e M. P. Santos.

Kim, Jaegwon, 2000, *What is ‘naturalized epistemology’?* in Bernecker & Dretske 2000.

<sup>89</sup> Na progressão dos programas de epistemologia a que o presente artigo serviu de apoio seguiu-se uma unidade de História da Ciência Cognitiva e da Filosofia da Mente (1950-1990), regida pelos seguintes objectivos: Ciência Cognitiva: Identificar as disciplinas da ciência cognitiva; Explicar a metáfora fundadora do paradigma cognitivista; Caracterizar o paradigma cognitivista como funcionalista, representacionista e computacionista; Identificar os pontos da história da lógica que foram fundamentais para a concepção de uma máquina cognitiva; Caracterizar ‘sistema formal’; Definir ‘algoritmo’; Explicar os conceitos de Máquina de Turing e Máquina de Turing Universal; Identificar marcos da história da Inteligência Artificial; Explicar a importância da linguística chomkyana na história da ciência cognitiva; Identificar o problema do texto de Turing *Computing Machinery and Intelligence*; Conhecer a argumentação do texto de Turing *Computing Machinery and Intelligence*; Descrever e interpretar o Teste de Turing. Filosofia da Mente: Identificar a filosofia da mente como ‘estudo de caso’ para a questão da epistemologia naturalizada; Explicar o funcionalismo e a realizabilidade múltipla a partir de H. Putnam; Explicar a Hipótese da Linguagem do pensamento de J. Fodor; Conhecer e avaliar os argumentos de J. Searle contra o cognitivismo; Conhecer e avaliar a argumentação de T. Nagel no texto *What is it like to be a bat?*; Explicar a crítica de D. Dennett ao Teatro Cartesiano. Nos artigos reunidos neste volume sob o título ‘Mente’ encontrar-se-á material de partida para o tratamento destas questões.

Quine, W.O., “Epistemologia naturalizada”, in Carrilho, M.M.1991 (org), *Epistemologia:posições e críticas*, Lisboa, Fundação Calouste Gulbenkian.

Quine, W. O, 1953, “Two Dogmas of Empiricism”, in *From a Logical Point of View*, Cambridge MA, Harvard University Press.

Rorty, Richard, 1988, *A Filosofia e o Espelho da Natureza*, Lisboa, Dom Quixote.

Rorty, Richard, 1990, *Contingency, Irony, Solidarity*, Cambridge, Cambridge University Press.

Wittgenstein, Ludwig, 1987, *Tractatus Logico-Philosophicus / Investigações Filosóficas*, Lisboa, Fundação Calouste Gulbenkian, trad. M.S. Lourenço.

## A FILOSOFIA E O PENSAMENTO COMO 'MEDIDA DO MUNDO'<sup>90</sup>

**Resumo:** Este artigo não visa o pensamento filosófico de Abel Salazar (embora sejam conhecidas as ligações de Abel Salazar a um particular movimento filosófico do século XX, o Positivismo Lógico do Círculo de Viena) mas antes directamente o tema do Ciclo – o pensamento como medida do mundo. Tomando como pretexto o ‘desejo filosófico’ que animou o pensamento de Abel Salazar, e entendendo o desejo filosófico como a ambição do pensamento humano de abordar o mundo a partir de vários ângulos (nomeadamente científicos, morais e artísticos), vou procurar analisar a relação da filosofia, como disciplina, com esse desejo.

Procurarei assim mostrar, a partir de uma panorâmica das linhagens da filosofia contemporânea, que nos conduzirá desde Kant até aos nossos dias, que a filosofia é precisamente uma anatomia ou análise de segundo grau das várias medidas do mundo tomadas pelo pensamento, tendo sempre no horizonte compreender o que significa ‘pensamento’, o que envolve conceber em conjunto as vertentes lógico-epistemológicas, ético-políticas, estéticas e fenomenológicas deste. No mapa conceptual das linhagens da filosofia contemporânea que procurarei desenhar tornar-se-á eventualmente possível localizar as tendências filosóficas de Abel Salazar. Será para isso necessário caracterizar as doutrinas do Positivismo Lógico quanto ao sentido e ao sem-sentido no pensamento, doutrinas em grande parte inspiradas no *Tractatus Logico-Philosophicus* (1921) de L. Wittgenstein, obra em que o autor procurou apresentar um mapa do dizível e do pensável que permitisse conceber a relação entre os aspectos lógicos, científico-naturais, éticos, estéticos e místicos do pensamento, bem como traçar uma fronteira entre o sentido e o sem-sentido. Embora os positivistas lógicos tenham tomado o mapeamento de Wittgenstein como pretexto para expurgar o sem-sentido, que ligavam nomeadamente à metafísica, muito do que de então para cá aconteceu em filosofia afastou, em grande medida, e de forma determinante para aquilo que hoje se faz, a ambição clarificadora apriorista do Positivismo Lógico.

### **1. Introdução: Filosofia, Abel Salazar, Positivismo Lógico, Wittgenstein, ultrapassagem do positivismo.**

Fazer a conferência deste Ciclo dedicada à Filosofia colocou-me desde logo um problema: Abel Salazar é por vezes apontado como um filósofo

<sup>90</sup> Agradeço à Professora Maria José Cantista o facto de me ter passado a incumbência de representar a Filosofia no Ciclo de Conferências organizado pela Reitoria da Universidade do Porto em que esta conferência se inseriu.

e eu não sou especialista em filosofia portuguesa, as minhas áreas são a filosofia da mente, do conhecimento e da linguagem. Decidi então falar de Filosofia contemporânea, usando como chave as relações de Abel Salazar com a filosofia. A conferência não vai por isso incidir sobre o pensamento filosófico de Abel Salazar - embora eu vá ter sempre em mente as ligações de Abel Salazar com um movimento filosófico do século XX, o Positivismo Lógico do Círculo de Viena - mas antes directamente sobre o tema do Ciclo, o pensamento como medida do mundo, na medida em que esse tema tem uma relação directa com aquilo que a filosofia faz. Por isso tomando como pretexto o 'desejo filosófico' - vou chamar-lhe assim - que animou o pensamento de Abel Salazar, o que vou procurar fazer é mostrar a relação da filosofia, como disciplina, com essa ambição do pensamento humano de abordar o mundo a partir de vários ângulos (nomeadamente científicos, morais e artísticos). Vou defender, a partir de uma descrição histórica esquemática das linhagens da filosofia contemporânea, que nos conduzirá desde Kant até aos nossos dias, que a filosofia é uma anatomia ou análise de segundo grau dessas medidas várias que o pensamento toma do mundo, uma análise que tem sempre no horizonte compreender o que significa 'pensamento', o que envolve conceber em conjunto as vertentes lógico-epistemológicas, ético-políticas, estéticas, metafísicas e fenomenológicas disso que é pensar. Na medida em que questões relativas à natureza do pensamento interessam pessoas de áreas exteriores à filosofia, gostaria em última análise de incentivar essas pessoas a procurar na filosofia algo que lhes sirva para trabalhar nas suas próprias áreas.

O meu primeiro objectivo prático é desenhar um mapa conceptual das linhagens da filosofia contemporânea. Nesse mapa será eventualmente possível localizar as tendências filosóficas de Abel Salazar. Abel Salazar foi, nos anos 30, o divulgador do Positivismo Lógico em Portugal. O Positivismo Lógico foi um movimento filosófico de cariz lógico-analítico, anti-metafísico, nascido na Europa germanófona, pelas mãos de homens como Moritz Schlick, Rudolf Carnap, Otto Neurath, que consideravam, em geral, que aquilo que interessava fazer em filosofia deveria fazer-se em grande proximidade com o trabalho científico. Abel Salazar considerava o Positivismo Lógico uma 'filosofia científica' e um importante antídoto para aquilo que considerava ser «o filosofismo mole, retórico, verboso e sentimentalíde dominante entre nós»<sup>91</sup>. Vou a certo ponto caracterizar as doutrinas do Positivismo Lógico que interessaram Abel Salazar, que o fizeram ver essa virtude de antídoto no movimento. As doutrinas em causa são doutrinas quanto ao sentido e ao sem-sentido no pensamento, e gostaria

<sup>91</sup> A. Salazar 1937, *Contra o filosofismo*, III, *O Trabalho*, 19 Agosto.

de mostrar de onde vieram e o que lhes aconteceu, na própria sequência de desenvolvimento da filosofia analítica de então para cá (o Positivismo Lógico pode ser considerado um movimento dentro da chamada ‘filosofia analítica’; mais à frente explicarei o que se entende por tal denominação). As doutrinas quanto ao sentido e ao sem-sentido no pensamento defendidas pelos positivistas lógicos foram em grande parte inspiradas pelo *Tractatus Logico-Philosophicus* de Ludwig Wittgenstein. Wittgenstein (1889-1951) é um dos maiores filósofos do século XX e o *Tractatus* (1921) é uma obra em que ele procura apresentar um mapa do dizível e do pensável, que precisamente permita conceber a relação entre os aspectos lógicos, científico-naturais, éticos, estéticos e místicos do pensamento, bem como traçar uma fronteira entre o sentido e o sem-sentido. Embora os Positivistas Lógicos tenham tomado o mapeamento de Wittgenstein como um pretexto para expurgar o sem-sentido – a interpretação que fizeram do *Tractatus* vai por aí – muito do que de então para cá aconteceu em filosofia afastou, em grande medida, e de forma determinante para aquilo que hoje se faz, a ambição clarificadora cientificista que o Positivismo Lógico reclamava para a filosofia. É isso que vou procurar mostrar mais à frente, em grande medida através de algumas explicações relativas a outro dos grandes filósofos do século XX, Willard Van Orman Quine (1908-2000). O objectivo da referência a Quine é mostrar o estado actual das questões que levaram Abel Salazar a procurar a filosofia, sem ficarmos presos às referências que foram as suas.

Como disse, vou falar mais da forma como a filosofia trata da natureza do pensamento do que de Abel Salazar. Mas a verdade é que o interesse da figura de Abel Salazar excede em muito a sua filiação filosófica no Positivismo Lógico: Abel Salazar é um óptimo caso prático para ilustrar o que significa a ambição ou desejo filosófico do pensamento, já que precisamente se trata de alguém que não se limitou a um único ângulo de abordagem do mundo (por exemplo à sua prática de cientista ou de artista).

Uma vez que pretendo falar de pensamento, tenho que introduzir algumas distinções, que ultrapassam um pouco as descrições históricas que se seguem mas que são básicas para o ponto que quero defender acerca da filosofia, como disciplina, e do pensamento. Parece-me muito importante distinguir três coisas: cognição, conhecimento e pensamento. Penso que o termo ‘cognição’ deve ser aplicado a processos informacionais e computacionais em agentes cognitivos, considerados ao nível subpeçoal. ‘Conhecimento’ é um termo que em filosofia, particularmente em epistemologia, significa *crenças verdadeiras e justificadas* (esta é a chamada definição tripartida de conhecimento, a qual chama a atenção para os aspectos mentais-psicológicos, lógico-metafísicos e normativos daquilo

que consideramos conhecimento). Ao contrário de ‘cognição’, ‘conhecimento’ envolve, pelo menos para os filósofos, aspectos normativos de ‘justificação’ e ‘verdade’ e portanto questões do nível da consciência explícita dos nossos procedimentos de inquérito. Temos depois o terceiro termo, o que nos interessa, ‘pensamento’. Ora, o conhecimento é um âmbito específico do pensamento que cobre nomeadamente o que se faz nas ciências; na palavra ‘pensamento’ não existe essa limitação ao conhecimento científico, estão nela naturalmente incluídas vertentes estéticas, éticas, metafísicas e fenomenológicas (trazendo naturalmente questões quanto ao lugar do pensamento na realidade e quanto à compreensão daquilo que é conhecido). É de pensamento que vou falar, não de cognição ou de conhecimento.

## **2. De onde vem a filosofia contemporânea? – Kant e as questões O que é representar? O que é agir? O que é apreciar?**

À filosofia enquanto disciplina coube desde sempre pensar sobre o pensamento, compreender o que é isso de termos pensamento, entendimento, formularmos abordagens teóricas em várias áreas. É por terem tentado compreender a compreensão, entender o entendimento, analisar conceitos que normalmente usamos de forma não problemática, tais como ‘eu’, ‘mundo’, ‘linguagem’, ‘certo’, ‘errado’, ‘belo’, ‘racional’, ‘irracional’, que ao longo da história os filósofos têm feito coisas tais como oferecer teorias da ciência, da arte, da moralidade, da religião, etc. Poderíamos ir muito atrás na história da filosofia para falar desta ambição do filósofo (já Platão procurou saber o que é o conhecimento, qual é a natureza da realidade, como é que nós nela nos inserimos, como devemos viver). Mas não pretendo aqui erudição e sim apresentar quadros conceptuais que nos sirvam para nos orientarmos face à filosofia que se faz no nosso tempo. Vou por isso apresentar apenas um extracto dessa história, tomando como referência Abel Salazar, um extracto dirigido especialmente àqueles que não são filósofos.

Parece-me que uma boa pergunta de partida é a seguinte: de onde vem a filosofia contemporânea? Uma resposta não muito polémica e que ainda mais nos faculta uma boa orientação teórica consiste em dizer que vem de Kant (1724-1804). Porquê? Porque Kant mapeou, de uma forma que ainda faz sentido para nós, nas suas três *Criticas*, aquilo a que podemos chamar as tarefas do pensamento e a forma como a filosofia lida com elas. Na *Critica da Razão Pura* caracterizou o problema de compreender o que fazemos com o pensamento quando conhecemos cientificamente a natureza, quando representamos cientificamente aquilo que existe e avançou uma resposta. Na *Critica da Razão Prática* caracterizou o problema de

compreender o que fazemos quando nos fazemos fazer, quando controlamos racionalmente as nossas acções, quando agimos moralmente. Na *Crítica do Juízo* procurou compreender o que acontece, o que se passa em nós, no nosso espírito ou mente, quando apreciamos esteticamente formas, sejam formas da natureza, sejam objectos de arte. Quando na sua primeira Crítica, a *Crítica da Razão Pura*, Kant procura caracterizar a natureza do conhecimento científico, está em causa a forma como este alcança ou não alcança a realidade em si (a que Kant chama númeno, por oposição a fenómeno = mundo representado). O núcleo da proposta é o conceito de juízo sintéticos a priori (o conhecimento a que chamamos científico, por contraste, nomeadamente, com os juízos de percepção que todos fazemos constantemente enquanto agentes cognitivos, é para Kant constituído por *juízos sintéticos a priori*, representações caracterizadas por uma universalidade e necessidade que por exemplo os nosso juízos normais de percepção não têm. Ora, de onde vêm essas características conceptuais fortes (com a universalidade e a necessidade)? Como bem viu, segundo Kant, David Hume, o filósofo empirista escocês do século XVIII, essas características – que distinguem um *take* perceptivo da realidade de uma pretensão (*claim*) de conhecimento científico não poderiam provir exclusivamente da experiência. O que Kant quer dizer é que para formularmos o conhecimento a que chamamos científico temos que dizer coisas tais como ‘sempre’, ‘todos’, ‘necessariamente’ e embora essas características estejam presentes nas nossas pretensões ao conhecimento, não podemos nunca, nenhum de nós, ter experiência directa do conteúdo desses termos. O argumento de Kant é que se elas estão presentes, e não vêm da experiência, só podem provir do sujeito. Só as estruturas a priori do sujeito podem justificar o carácter de ciência dos enunciados da ciência. As estruturas a priori são para Kant o espaço, o tempo, as categorias (tais como substância e causalidade), a ideia básica é que o nosso tipo de mente molda aquilo que é representado.

Na segunda Crítica, a *Crítica da Razão Prática*, algo de análogo se passa: o pensamento também se volta para si próprio, para analisar aquilo que faz, mas neste caso não se olha para o caso em que se faz ciência e sim para o caso em que existe acção racional. Kant considera que os humanos podem chegar a agir por sobre os desejos e inclinações naturais, moldando e direccionando estes de uma forma da qual podem dar razões – a sua intenção é compreender o que é esse agir racional por contraste com o agir instintivo que nos é propriamente natural, animal. As suas grandes propostas são o imperativo categórico (a ideia segundo a qual devemos agir de forma a que a máxima que rege a nossa acção seja universalizável



– i.e. susceptível de ser tomada como regra por qualquer agente racional) e a ideia segundo a qual embora do ponto de vista da construção científica só alcancemos a realidade (inclusive a nossa própria realidade física e mental) como fenómeno, como representação (através da mediação das nossas estruturas a priori, mentais diríamos hoje), quando nos experienciamos a nós próprios agindo temos aí uma forma mais directa de alcançar a realidade em si (a que Kant chama, recorde-se, númeno). Assim, segundo Kant, a nossa experiência de nós próprios enquanto agentes racionais revela-nos mais acerca de nós próprios do que as nossas representações cognitivas, nomeadamente revela-nos que a nossa natureza é liberdade. Muito do pensamento ético e político contemporâneo vem daqui (desta ideia de uma possibilidade de universalização do agir racional – pense-se em John Rawls, cuja teoria da justiça, uma teoria permeada de influenciada kantiana, está no centro da filosofia política contemporânea). No entanto, muito pouca gente aceitaria os dualismos kantianos entre liberdade e desejos, realidade representada e realidade em si. Na terceira Crítica de Kant, a *Crítica do Juízo*, trata-se em geral da importância da componente estética na totalidade do nosso pensamento, de formas e apreciação. As ideias principais são o prazer desinteressado (uma apreciação que não deseja incorporar ou utilizar o apreciado) e o pensamento da finalidade nas coisas, e a possibilidade deste depender da nossa natureza mental, uma vez que finalidade e teleologia não fazem parte segundo Kant da representação científico-natural do mundo.

Falei de Kant, recorde-se, para mapear as tarefas de análise do pensamento sobre si próprio. A partir de Kant, podemos considerar que se queremos pensar sobre o pensamento aquilo que devemos procurar compreender é: o que é representar e conhecer? O que é agir? O que é apreciar? Claro que, ao procurar compreender isto, o que nos interessa é compreender a nossa natureza, aquilo que somos, não aquilo que somos enquanto corpo físico, organismo biológico, porque é claro que também somos tudo isso, vamos conhecendo mais tudo isso, mas a nossa natureza pensante, da qual depende própria possibilidade de conhecimento matemático, físico e biológico, histórico, etc. Kant fez tudo isto falando sempre de um suposto Sujeito Transcendental e das estruturas a priori deste, que incluiriam ‘molduras’, *templates* de conceitos, para pensar em tudo o que pensamos (dou exemplos: substância, causalidade, espaço, tempo, que não seriam adquiridos empíricos mas imposições mentais ao que é pensado). Queria apenas salientar duas ideias importantes de Kant neste desenho de tarefas para compreender o pensamento:

- (1) É certo que nos três funcionamentos que lhe pareceram essenciais ao nosso tipo de pensamento (a que chamou sensibilidade, entendimento e razão) Kant viu componentes a priori. No entanto, ao contrário de um filósofo como Descartes, que considerou que pensamento é consciência e consciência é imaterialidade, Kant descreveu todo este funcionamento e todas estas componentes a priori sem presumir a imaterialidade do pensamento.
- (2) Além das componentes Sensibilidade e Entendimento – hoje falaríamos talvez, num primeiro momento pelo menos, de percepção e categorização – Kant considera essencial que notemos um outro funcionamento do pensamento, a que chamou Razão, e que envolve ‘Ideias’ – Ideia são grandes unificações para organizar a nossa vida mental, cuja função é heurística. Dou um exemplo: ninguém poderá alguma vez experienciar directamente por exemplo ‘o mundo como totalidade’, uma vez que somos locais e limitados espacial e temporalmente e é assim também todo o nosso input de dados sobre o mundo. ‘Mundo’ é portanto um exemplo daquilo a que Kant chama uma ideia da razão, uma Ideia de que necessitamos para organizar o nosso pensamento mas que transcende aquilo que alguma vez poderá ser experientialmente dado a cada um de nós e portanto conhecido (‘Alma’ e ‘Deus’ estão nas mesmas condições, e aquilo a que Kant chama pensamento comporta essa ‘tendência para as grandes unificações’ – uma tendência à qual, segundo Kant, definitivamente não devemos chamar conhecimento).

Imaginemos que aceitamos as tarefas de Kant para o pensamento. Poderíamos depois, um pouco maldosamente, dirigir a Kant a seguinte pergunta: Onde é que está tudo isso? (toda essa maquinaria cognitiva) No cérebro? E esse processo todo, o Sujeito Transcendental, a sensibilidade, o entendimento, a razão, não pode ser simplesmente estudado em terceira pessoa, pelas disciplinas científicas da cognição? Essas perguntas são e não são boas perguntas. São boas perguntas porque, se Kant viu bem as tarefas do pensamento, temos o todo o direito de as trabalhar contemporaneamente, e portanto de utilizar aquilo que sabemos a mais relativamente a Kant para lhes dar resposta. Não são boas perguntas tanto quanto confundem cognição e pensamento, que podemos e devemos distinguir. Noutras palavras, é perfeitamente possível interessarmo-nos por cérebro e cognição – e portanto pelas bases físicas, neuronais daquilo que é pensar, representar, desejar, e continuar a considerar pertinente a moldura

kantiana para pensar sobre o pensamento. Recorde-se que com cognição queremos dizer informação, computação, processamento – o que se passa no agente cognitivo (e sobre isso sabemos hoje muito mais do que Kant sabia) – mas com a palavra ‘pensamento’ evocamos a intencionalidade da representação (e o representado não é o mecanismo da representação), os aspectos normativos da representação (questões de verdade, justificação, provas) bem como aspectos a que para simplificar vou chamar fenomenológicos de subjectividade e compreensão, e aspectos metafísicos relativos à relação pensamento/realidade (ao facto de a realidade ser tal que inclui pensamento). E por isso não podemos responder a questões acerca do pensamento com dados sobre cognição, embora possamos e devamos obviamente utilizá-los. Dou breves exemplos: quando eu penso ‘Esta mesa mede 3 metros’, ou ‘Eu amo a pessoa X’, esses estados mentais dependem causalmente do meu cérebro, são instanciados por eventos no meu cérebro mas não são acerca do meu cérebro, no sentido em que não é o meu cérebro que neles é representado, não é o meu cérebro que mede 3 metros ou que é amado. Repito que a pergunta não é pertinente (a pergunta era: Onde é que está tudo isso – toda essa maquinaria cognitiva - está no cérebro?) porque desenhando as tarefas de pensamento Kant estava a falar de pensamento como medida do mundo e eventualmente da paisagem e arquitectura disciplinar do pensamento, i.e. estava a falar da forma como o nosso tipo de mente possibilita as nossas produções intelectuais, e da forma como estas se articulam entre si e não do substrato físico da cognição (a que chamaria sujeito empírico, e hoje chamaríamos mente-cérebro) e que constitui uma área particular das nossas investigações, nomeadamente via a ciência cognitiva.

Vou dizer onde quero chegar e porque quis localizar Kant como momento inicial na filosofia contemporânea

- (1) Com Kant, a filosofia é caracterizada como tarefa de segunda ordem, não se pode fazer directamente teoria geral da natureza da realidade porque a nossa mente, o nosso tipo de mente, se intromete e temos que perceber primeiro como é que a mente funciona, e a nossa mente não é um espelho ideal, antes é caracterizada, por exemplo, pelo nosso particular tipo de percepção sensorial e de (para usar o termo de Kant) animalidade. É neste quadro de análise da razão pela razão que podemos compreender o programa transcendental kantiano para a filosofia. Chama-se ‘transcendental’ não ao ‘conhecimento em 1º grau’ (por exemplo matemático ou físico ou outro) mas à análise das condições de

possibilidade desse conhecimento, análise essa que se faz, segundo Kant, investigando as condições a priori do sujeito. A ideia de filosofia como análise das condições a priori substitui, com Kant, a ideia de filosofia como metafísica ‘transcendente’ (esta metafísica transcendente seria precisamente a tentativa de alcançar de forma puramente conceptual e apriorista objectos que nunca poderiam ser dados na experiência (os grandes objectos que interessam a razão humana, tais como Deus, a Alma, o Mundo). Tal tentativa não pode senão, segundo Kant, falhar.

- (2) É importante repetir, para distinguir Kant de filósofos anteriores e sublinhar a contemporaneidade do seu pensamento, que o apriorismo defendido por Kant não tem de maneira nenhuma que ser um inatismo imaterialista: aquilo que existe a priori no sujeito, a marca da mente naquilo que é conhecido, não é apátnio de uma alma com ideias inatas (esta ideia é por exemplo cartesiana, mas não tem lugar na *Crítica da Razão Pura*). Kant nega que possamos provar de forma puramente apriorista a imaterialidade, imortalidade ou sequer a unidade disso em nós a que podemos chamar alma ou também a existência de Deus. Isto não quer dizer que Deus ou a imortalidade da alma não possam ter uma função no pensamento – para Kant têm uma função prática, moral – quer apenas dizer que estaremos a pensar mal no que é pensar sobre Deus se pensarmos que podemos argumentar e demonstrar racionalmente a sua existência. Segundo Kant isso não pode fazer-se relativamente a objectos dos quais não temos qualquer experiência.
- (3) Sobretudo interessa reter que, para Kant, compreender a nossa natureza pensante não acaba com a tarefa de compreender o que fazemos quando pensamos cientificamente – temos que colocar num mesmo quadro de análise do que é pensar a forma como agimos moralmente e apreciamos esteticamente. Nenhuma dessas componentes do pensamento pode ser reduzida. Sugiro então que aceitemos que Kant nos deu uma tarefa concreta, definiu como que uma direcção para o pensamento contemporâneo, a tarefa de analisar a forma como se relacionam numa única paisagem de pensamento, as produções intelectuais, teóricas, de um mesmo Sujeito. Bom mas se a filosofia contemporânea começa com Kant, depois de Kant o que aconteceu em filosofia ?

### 3. Após Kant.

Algumas pessoas (estou a pensar por exemplo no filósofo americano Richard Rorty (n.1931), que diz isto como crítica) vêm Kant como um exemplo daquilo a que chamam uma ‘centração epistemológica da filosofia’. Isto quer dizer que Kant, como filósofo, consideraria a ciência como a preocupação maior da filosofia. Tal posição é, por um lado, contestável pelo que acabo de dizer acerca dos interesses de Kant como filósofo. Por outro lado, se Kant presta de facto, na sua filosofia, bastante atenção à ciência, isso é o resultado de uma mais do que legítima preocupação com a natureza da representação e do conhecimento. De resto, basta olhar para a história da filosofia para ver imediatamente que foi frequentemente o trabalho relativo a tópicos científicos que gerou, nos filósofos, a necessidade de elaborar uma metafísica, uma teoria geral da natureza da realidade (temos uma teoria científica e perguntamos: como será o mundo se é tal que pode ser pensado assim?). Os exemplos que costumo dar são as preocupações de Aristóteles, Descartes e Leibniz, respectivamente, com a classificação das espécies de seres vivos, a geometria analítica e o cálculo infinitesimal. Relativamente a Kant, é verdade que ele desenhou a sua filosofia em parte para compreender o que é representar cientificamente o mundo, e o que visou como exemplo foi a ciência sua contemporânea, a física newtoniana. Mas não foi essa sequer a única produção da razão cujas condições de possibilidade pretendeu analisar no seu programa de investigação. É certo que a Modernidade como época é marcada pelo advento da ciência físico-matemática da natureza (também, certamente, por um fenómeno de laicização das estruturas sociais e políticas), é certo que a ciência é o modo de abordar o mundo que se tornou civilizacionalmente dominante a partir da época moderna, tanto que foi possível chamar à civilização ocidental a ‘Civilização da Ciência e da Técnica’. Mas se a Modernidade é de algum modo definível conceptualmente como ‘época’ é porque ela representa a realização civilizacional de uma (de certa forma nova) noção de verdade, ligada ao progresso e à razão, que inclui a ciência, mas não se restringe a ela. Dizer, como Rorty, que Kant exemplifica uma centração epistemológica da filosofia, é esquecer que Kant pensou também nas repercussões morais e políticas da centralidade na Razão. Foi de resto pelo facto de essa situação se ter alterado de então para cá (não a centração epistemológica, mas a centração na razão) que muita filosofia feita depois de Kant se afastou da ciência.

Uma vez que propus considerar o mapa kantiano das tarefas do pensamento como iniciador da filosofia contemporânea, quero perguntar agora o que aconteceu depois de Kant, que marcou a filosofia contemporânea

(já sugeri que se teria tratado de uma ‘descenração epistemológica’ da filosofia e de uma entrada em cena do irracionalismo). Para percebermos este movimento temos que perceber que Kant, apesar de propôr limites à forma como representamos conceptualmente o mundo, devido à nossa particular conformação mental e finitude, foi um racionalista, um *Aufklärer*, um iluminista, alguém que apostou nas potencialidades da razão em termos científicos, morais, sociais e históricos. Compreender que essa ideia tem estado sob ataque desde então é uma chave para vastas áreas do pensamento contemporâneo.

O primeiro grande movimento de oposição à centralidade da razão, defendida em termos ideológicos pelos iluministas, foi o romantismo, de que todos ouvimos falar nas artes, e que em filosofia esteve ligado ao idealismo. O Romantismo foi um movimento cultural e artístico contra-iluminista que representou uma fortíssima influência na cultura europeia e americana nas últimas décadas do século XVIII e nas primeiras décadas do século XIX. Mais do que exclusivamente um movimento artístico, o romantismo constituiu uma certa manifestação global de rejeição do pensamento racional. Porque é que isso nos importa aqui? Porque o expoente máximo do pensamento racional seria o pensamento científico. Ora acontece que a partir desta perspectiva a racionalidade científica é, considerada como prosaica e des-‘espiritualizadora’ do mundo. Curiosamente, é em parte à filosofia kantiana que os românticos vão buscar a própria definição de realidade como espiritualidade (mais especificamente às ideias de Kant, que referi atrás, acerca da nossa verdadeira natureza como liberdade). Daí passam, no entanto, para uma posição de certa forma pouco kantiana: a ideia segundo a qual a ‘verdadeira’ realidade não pode ser acedida de forma intelectualista, mas apenas através da intuição, da emoção, do sentimento.

Noutras palavras, por trás dos traços programáticos do romantismo em arte – a exaltação da natureza, a expressão dos sentimentos, da emoção, da intuição, da imaginação, o culto do exótico – está uma determinada rejeição da civilização enquanto relação racional com o mundo. Para os românticos, apenas acontecimentos de intuição e de emoção, não argumentos racionais, nos darão acesso ao ‘eterno enigma do mundo’. Para fundamentar filosoficamente este programa, os românticos alemães foram buscar a Kant a ideia de númeno ou realidade em si como liberdade. Da ideia de liberdade como realidade em si em cada um de nós passaram à ideia de realidade total como Espírito, aspirando à realização. O caso mais extremo desta tendência foi a filosofia de G. W. F. Hegel (1770-1831), um sistema do Espírito Absoluto, um grande sistema da Racionalidade Real, realizando-se na natureza e na história. Se Kant colocou para si próprio a obrigação intelectual de delimitar o funcionamento do pensamento a partir

de dentro, analisando as características do nosso tipo de mente, antes de nos pronunciarmos acerca de magnos assuntos, para não correremos o risco de nos pronunciarmos acerca daquilo de que não temos experiência, Hegel volta decididamente às grandes teses metafísicas, acerca da natureza da realidade – é um idealista, que defende que toda a realidade, natureza e história, é uma manifestação do Espírito no seu desenvolvimento racional. Uma grande parte dos filósofos do século XIX podem ser vistos como contestando Hegel: K. Marx (1818-1883) retoma de Hegel as ideias de dialética e racionalidade da realidade, substituindo o idealismo por materialismo, o filósofo dinamarquês S. Kierkegaard (1813-1855) nega a importância de um sistema global da realidade para cada um de nós e traz o indivíduo e a sua existência para o centro da filosofia. A. Schopenhauer (1788-1860) e F. Nietzsche (1844-1900) negam a racionalidade da realidade, Schopenhauer identificando o Númeno kantiano (realidade em si) com a vontade (nada de racional), e o fenómeno com a representação, Nietzsche pegando neste princípio para propor uma interpretação estética da existência, a única possível, sob a forma dos princípios Apolo e Dionísos.

Mas a consequência que queria sublinhar diz respeito a algo que acontece em grande parte da filosofia europeia, no período posterior a Kant: para usar a expressão de Rorty de forma invertida, trata-se de uma ‘descentração epistemológica da filosofia’ e da instalação da oposição a um certo racionalismo dentro da própria filosofia. O racionalismo encontramos-lo ainda em Kant, a oposição na verdade vem até nós (vários filósofos contemporâneos, estou a pensar por exemplo num dos maiores filósofos do século XX, Martin Heidegger, o autor de *Ser e Tempo*, podem ser radicados aqui). O que quero dizer é que a separação entre filosofia e ciência, devido à ligação entre a ciência e um certo racionalismo afastado pela filosofia, não pode ser considerada consumada no próprio Kant – por mais que Kant proclame a divisão do trabalho entre os teóricos da forma, os teóricos do sujeito, que analisam a forma como a mente ou espírito opera nas suas várias vertentes e os teóricos do conteúdo, que produzem conhecimento de primeira ordem (i.e. os cientistas), a *Crítica da Razão Pura* é uma obra de teoria da ciência e de teoria do conhecimento, de acordo com a qual fazer teoria da ciência é um empreendimento central para compreendermos o que significa representar e conhecer racionalmente o mundo.

Bom, temos então, após Kant, a entrada em cena de certa ideia de filosofia como crítica ético-político-existencial e como prática mais ou menos poética. Abel Salazar diria que esta filosofia tanto pode ser boa arte – é o que ele pensa acerca de autores como Nietzsche ou Kierkegaard – como mau filosofismo. De qualquer forma esse estado pós-kantiano da filosofia interessa-nos para compreendermos a cisão entre as duas grandes

tradições filosóficas nossas contemporâneas, a filosofia analítica, que é a tradição filosófica dominante no mundo de língua inglesa, e a filosofia a que os filósofos analíticos chamam ‘continental’, que é a mais praticada entre nós e que tem origens sobretudo francesas e alemãs. Sintomático da localização pós-kantiana dessa cisão é o facto de Hegel ser considerado um grande filósofo na tradição continental, enquanto os filósofos analíticos o consideram, maioritariamente, um autor obscuro e incompreensível. Kant, pelo contrário, constitui uma referência para ambas as tradições. O que me parece importante para compreender as linhagens da filosofia contemporânea é notar que houve danos laterais do desvio anti-racionalista da filosofia do século XIX: na prática, a filosofia nascida desta época, na tradição franco-alemã, deu por si divorciada de um outro tipo de pensamento, que continuava a decorrer ao seu lado, e que se ia tornando social e civilizacionalmente cada vez mais omnipresente e incontornável, o pensamento científico. Ao contrário do que Kant ainda fazia, essa filosofia desistiu de pensar na questão ‘o que é representar’ tomando como discurso de primeiro grau a ciência. A filosofia hoje chamada ‘continental’ (se por tal nome entendermos aquela que tem como figuras de referência autores como Nietzsche, Heidegger, Foucault ou Derrida) está historicamente ligada, mesmo que longinquamente, à rejeição da imagem racionalista do mundo que se observa neste período.

#### 4. As principais tendências da filosofia no século XX.

Sugeri que podemos reportar ao movimento romântico alguns dos factores que teriam estado na origem da cisão entre filosofia analítica e filosofia continental (a ligação é entre romantismo e filosofia dita continental, produzindo o afastamento entre filosofia e ciência). Mas de facto e oficialmente a instituição destas duas grandes correntes da filosofia contemporânea deu-se mais ou menos nos finais do século XIX, inícios do século XX, com as obras de Gottlieb Frege (1848-1925) e Edmund Husserl (1859-1938)<sup>92</sup> O que é interessante é que estes dois autores, iniciadores de duas linhagens metodológicas que se mantiveram até hoje, e que estruturaram a filosofia do século XX, além de terem sido contemporâneos, ambos filósofos e matemáticos, não se interessavam por temas muito diferentes. Basicamente, ambos se interessavam pelo estudo dos conteúdos do pensamento, e ambos pensavam que esse estudo não deveria reduzir-se à disciplina que então se estabelecida na sua legitimidade e autonomia, deslocando a filosofia: a psicologia enquanto ciência natural empírica. Aliás, o

<sup>92</sup> Do lado de Frege poder-se-ia referir as obras *Begriffsschrift* (1879) e *Fundamentos da Aritmética* (1884), do lado de Husserl, as *Investigações Lógicas* (1900-1913) e a *Ideia de Fenomenologia* (1907).



problema para a filosofia no início do século XX era sobreviver como disciplina enquanto a psicologia se estabelecia como ciência – porque, afinal, o que é que filósofos e psicólogos fazem de diferente? Será que os psicólogos não estavam a utilizar finalmente, e muito bem, os métodos da ciência natural para tratar do pensamento?

A fenomenologia e a filosofia analítica, que são as duas grandes linhagens e metodologias da filosofia do século xx, foram geradas por uma resposta negativa esta questão (os métodos da ciência natural não podem ser utilizados para tratar a natureza do pensamento) e por uma mesma intenção: a despsicologização do estudo do pensamento. Sem desprestígio para a psicologia, o que isto significa é que a psicologia é uma ciência empírica, uma ciência natural entre as ciências naturais (da cognição, do comportamento), e a filosofia não o é nem quer sê-lo. Os problemas da filosofia não são problemas empíricos, relativos à cognição e comportamento de indivíduos, mas problemas de segundo grau relativos ao que é pensado e à natureza do pensamento, problemas que, de acordo com a filosofia analítica e a fenomenologia, respectivamente, devem ser abordados através de uma análise da linguagem, quer por uma análise da consciência.

De acordo com a fenomenologia, o objecto da filosofia é a descrição dos conteúdos da consciência-tanto-quanto-esta-representa-o-que-quer-que-seja, suspendendo qualquer referência ao mundo físico e ao estudo deste feito pelas ciências naturais (em termos de causalidade). O mundo natural é o terreno adquirido e legítimo das ciências naturais. O que cabe à filosofia é a descrição daquilo que é pensado, por si só (é claro que se acaba assim por encontrar de novo o mundo todo, e o próprio conhecimento científico desse mundo, mas agora considerado enquanto pensamento).

De acordo com a Filosofia Analítica, que deve muito ao desenvolvimento da lógica formal no início do século XX, o objecto da filosofia é o estudo (englobante, geral) do pensamento. Esse estudo deve ser feito (só pode ser feito) através da análise (lógica) da linguagem. Porquê a análise da linguagem? Porque só a linguagem torna o pensamento público e objectivo (por oposição às imagens e sentimentos que povoam as vidas mentais dos indivíduos), só a linguagem possibilita inquéritos racionais comuns. A linguagem pode ser um ‘espelho distorcido do pensamento’ mas é o único espelho que temos, a única ‘moeda comum’. Os inquéritos epistemológicos, metafísicos, éticos, estéticos, etc, da filosofia assumem assim, de acordo com a filosofia analítica, a forma de investigações lógico-linguísticas.

Podemos portanto ver as duas grandes linhagens da filosofia do século XX como uma tentativa de restaurar o racionalismo afastado no período pós-kantiano, e de defender que não é possível estudar o pensamento com os métodos das ciências naturais, por causa de aspectos essenciais do

que entendemos por pensamento: a representação, a normatividade, e a subjectividade da compreensão. Em Frege e Husserl encontramos esta resposta comum, a grande diferença sendo basicamente o uso da lógica na filosofia.

Antes de avançar, vou procurar ser um pouco mais específica quanto àquilo de que se fala quando se fala de filosofia analítica e de filosofia continental, dando como exemplo autores que provavelmente conhecem.

Quanto à filosofia analítica, as figuras individuais mais influentes na sua constituição, são, para além de Gottlob Frege, Bertrand Russell (1872-1970) e Ludwig Wittgenstein (1889-1951). Mais longinquamente, a filosofia analítica tem origens no empirismo clássico, ligado aos nomes de J. Locke (1632-1704), G. Berkeley (1685-1753) e D. Hume (1711-1776). Mas o empirismo que mais directamente a marcou foi o Positivismo Lógico do Círculo de Viena, de que falarei a seguir, ligado a Abel Salazar. É com esta forma de fazer filosofia – baseada nos métodos da lógica formal e na ideia de tarefa filosófica como análise da estrutura das teorias e da linguagem, de modo a conseguir a depuração da metafísica e a unidade da ciência – que Abel Salazar vai considerar ter afinidades. Se consideramos o facto de os autores do movimento verem aquilo que fazem como uma oposição à mitologia romântica do solo, da terra e da inefabilidade do ser omnipresente em filósofos influentes nessa altura no mesmo espaço cultural, sobretudo M. Heidegger, certas afinidade tornam-se compreensíveis. Pelo menos o próprio Abel Salazar não se coibiu de ver uma correspondência entre por um lado o ‘filosofismo retórico’ e as autocracias que no tempo (anos 30) se estabeleciam na Europa, e por outro lado entre o que chamava ‘filosofia científica’ e a democracia. Esta é uma outra chave para compreendermos a sua proximidade relativamente ao Positivismo Lógico.

Bom, mas afastando os olhos de Portugal, vemos que é no seio desta tradição constituída nas primeiras décadas do século XX que vêem a surgir filósofos tais como W.O. Quine, H. Putnam, N. Goodman, D. Davidson, M. Dummett, J. Searle, J. Rawls, S. Kripke, J. Fodor, D. Dennett, T. Nagel, etc.

Por seu lado, a linhagem da filosofia continental pós-husserliana comporta o trabalho em filosofia contemporânea que se reconhece sob os títulos de fenomenologia, hermenêutica, estruturalismo, desconstrução, abrangendo filósofos e teóricos tais como M. Heidegger, J.-P. Sartre, M. Merleau-Ponty, E. Lévinas, H. G. Gadamer, P. Ricoeur, M. Foucault, R. Barthes, J. Lacan, J. Derrida, etc. Esta tradição é a mais conhecida e praticada em Portugal na filosofia académica.

A relativa artificialidade da separação entre filosofia ‘analítica’ e ‘continental’ é provada pelo facto de ser muito difícil classificar como

‘analíticos’ ou ‘continentais’ filósofos contemporâneos como por exemplo J. Habermas ou R. Rorty, que vêem as ‘vantagens’ (diferentes) das duas tradições e se servem delas.

A partir de Heidegger (e se juntarmos um pouco de Nietzsche) podemos compreender a genealogia conceptual dos grandes nomes da filosofia continental, como Derrida, Foucault e Lévinas. Mas vou dar mais atenção ao outro ramo da filosofia contemporânea, a filosofia analítica, nesta recta final do mapa conceptual, precisamente por causa do interesse de Abel Salazar pelo Positivismo Lógico. Como afirmei no início, Abel Salazar viu no Positivismo Lógico um antídoto para aquilo a que chamou ‘o filosofismo degenerado e o *pathos* metafísico da filosofia portuguesa’. Pareceu-lhe que esse filosofismo português, essa tendência para um filosofar ‘poetizante saudosista’ (pensando em Teixeira de Pascoaes) e ‘metafísico’ (pensando em Leonardo Coimbra) era ainda muito pior do que os originais (que Abel Salazar considerava serem Bergson e Heidegger). Parece-me por isso interessante, para compreender não apenas Abel Salazar como momento e sintoma na vida intelectual nacional mas os rumos da filosofia contemporânea que por ele passaram, olhar e procurar ver (1) de onde vieram e (2) o que aconteceu às ideias nucleares do Positivismo Lógico que o seduziram, na própria sequência de desenvolvimento da filosofia analítica.

Antes de mais, o que era o Positivismo Lógico? Como já disse, o Positivismo Lógico, ou Círculo de Viena, foi um movimento intelectual dos anos 20-30 do século XX protagonizado por um grupo de filósofos e cientistas, entre os quais se encontravam M. Schlick, R. Carnap, H. Feigl e O. Neurath. O positivismo lógico é ‘positivista’ porque considera que as ciências são a única via para o conhecimento propriamente dito: qualquer incursão ‘para além’ dos limites e dos métodos da ciência arrisca-se a ser cognitivamente vã, sem sentido, e o discurso produzido arrisca-se a ser insusceptível de verdade ou falsidade. Para os positivistas lógicos seria esse o caso de grande parte da metafísica tradicional. O positivismo lógico é ‘lógico’ porque a definição daquilo que segundo este grupo se deve fazer em filosofia dependeu de desenvolvimentos na lógica e na matemática, e nomeadamente de criações e inovações do âmbito da lógica formal. A agenda do movimento é combater a metafísica, conseguir a unidade da ciência, e reclamar para a filosofia a análise lógica de linguagens e teorias.

Ora, o que aconteceu às ideias nucleares do Positivismo Lógico quanto ao sentido e ao sem sentido foi um ataque ao fundamento destas, e esse ataque está muito presente na filosofia que hoje se faz na tradição analítica. O que está basicamente em causa é o facto de os positivistas lógicos assumirem, na sua teoria do conhecimento, a separação entre a forma, a linguagem, a lógica, e o conteúdo (vulgo distinção analítico/sintético).

O nome que dá voz a essa crítica é Quine, um filósofo e lógico americano, que é provavelmente o mais importante epistemólogo do século XX. Quine continua a ser um epistemólogo e um lógico mas rejeita totalmente as ideias metodológicas e anti-metafísicas do Positivismo Lógico, e o seu ataque à distinção analítico-sintético tem consequências muito importantes na forma de os filósofos pensarem no que fazem quando fazem filosofia (é esse ataque que traz para a ordem do dia a ‘naturalização da epistemologia’ e a ‘continuidade filosofia-ciência’). Pode parecer surpreendente, mas parte do que está em jogo na discussão em torno dessa aparentemente insignificante dicotomia analítico-sintético, que diz respeito a frases e ao sentido que elas formulam, é a forma como a partir de dentro da filosofia é vista a relação da filosofia com as ciências e também a possibilidade (ou impossibilidade) de ‘expurgar’ a metafísica do inquérito científico.

Mas voltemos um pouco atrás – acabei de afirmar que Abel Salazar de certa forma reencontrou em Portugal um combate que, no mundo germanófono, os Positivistas Lógicos travavam contra algo que era expenciado por Heidegger. Exactamente o quê? Heidegger afirmou por exemplo, e essa afirmação ficou célebre, que «a ciência não pensa» (*Wissenschaft denkt nicht*)<sup>93</sup>. Transferido para Portugal, esse espírito de sobrançeria produz resultados que Abel Salazar detestava, e a que chamou *pathos* metafísico, filosofar poetizante, retórica oca, pedantismo, ininteligibilidade. Essas características pareciam-lhe comuns por cá, uma atracção fatal ligada ao atraso cultural do povo português.

Podemos triangular os interesses filosóficos de Abel Salazar da seguinte forma: (i) ele sente uma afinidade com o Positivismo Lógico, de que se torna um divulgador entre nós, (ii) a concepção metodológica de filosofia do Positivismo Lógico foi inspirada pelo *Tractatus Logico-Philosophicus* de Wittgenstein e (iii) ergueu-se contra o tipo de filosofia expenciado por Heidegger (R. Carnap tomava frases de Heidegger, tais como ‘o nada nadifica’, como o próprio exemplo do sem-sentido que o Positivismo Lógico queria expurgar do pensamento e Abel Salazar cita-o).

Olhemos para as concepções de filosofia de Heidegger e Wittgenstein. Parte do que se joga nas concepções de filosofia de Heidegger e Wittgenstein é a relação da filosofia com as ciências. Heidegger afirmou, e essa afirmação polémica ficou célebre, que a ciência não pensa. Embora a afirmação tenha sido ridicularizada, recebida com irritação (por ser pretensiosa, intolerante, irracionalista) ela não é imediatamente absurda. O que é certo é que Heidegger não era nem queria ser um epistemólogo, ele não aborda a ciência como epistemólogo, i.e. procurando compreender a

<sup>93</sup> Heidegger 1952: 151.

ciência como conjunto de afirmações sobre aquilo que a realidade é, em termos de verdade, justificação, provas. Se Heidegger se interessa pela ciência é para a pensar como forma de ‘abordar o ser’, como acontecimento civilizacional. E aquilo que lhe interessa como filósofo é o sentido do ser, a história do Ocidente, a crítica da modernidade, o destino da civilização. E se, para Heidegger, a ciência não pensa é porque ela não é ontologia, pensamento do sentido do ser. As ciências são apenas ‘um modo de ser da existência’, os métodos científicos foram criados para explorar os entes, não para explorar o ser dos entes, e pensar é, para Heidegger, pensar o Ser, não conhecer os entes. Note-se de passagem que o que está aqui em jogo (quando se afirma que a ciência é um modo de ser da existência) é a seguinte ideia: aquilo a que chamamos o mundo não está aí só para, ou principalmente para, ser conhecido, mas para ser ‘lidado’, e as ciências vêm nitidamente depois (note-se que mesmo do ponto de vista da teoria do conhecimento, e não do ponto de vista do pensamento do ser e da avaliação das civilizações, é possível dizer algo de semelhante: é o que fazem todos os filósofos pragmatistas, por exemplo).

E Wittgenstein, como é que faz filosofia? Ao contrário de Heidegger, que chega à filosofia através da teologia, Wittgenstein começou por estudar engenharia aeronáutica e chega à filosofia através da lógica e da matemática. O *Tractatus Logico-Philosophicus* resulta dos estudos, feitos com B. Russell, sobre os fundamentos da lógica e da matemática. A linguagem é o tema essencial do *Tractatus* (o que não impede que na obra se trate também daquilo que é ‘da ordem do místico’, e do lugar das questões éticas e estéticas no pensamento). De acordo com o autor do *Tractatus*, podemos pensar o que podemos ‘modelizar’ em linguagem, e a lógica é o denominador comum das linguagens, forma o quadro de estruturação do conhecimento do mundo. Por isso, a análise do esqueleto lógico do sentido é o primeiro passo em filosofia. É esta teoria que os positivistas lógicos vão buscar ao *Tractatus*, embora retirando, se assim se pode dizer, aquilo que para Wittgenstein era ‘o mais importante’. Vou procurar descrever a doutrina acerca de sentido e sem sentido apresentado no *Tractatus*, na medida em que é essa doutrina que inspira o chamado «critério de significação» do Positivismo Lógico, que enunciarei também a seguir.

Importa recordar o que os filósofos e cientistas do Círculo de Viena pretendem para a filosofia: eles pretendem que a filosofia consista (apenas) numa análise da linguagem e das teorias e que assim contribua para a unidade da ciência e para a depuração da metafísica (esta seria em grande medida ‘discurso sem sentido’ – e, como afirmei, Heidegger foi um exemplo abundantemente utilizado para mostrar o que se queria dizer com discurso sem sentido). As ideias de Wittgenstein vêm apoiar a referida intenção

na medida em que permitem fazer uma ‘cartografia’ daquilo que é pensável e dizível.

Temos então a seguinte estruturação do que podemos dizer e pensar:

- (1) O ‘esqueleto’ do sentido são frases lógicas ‘vazias’, a que Wittgenstein chama tautologias. Tautologias não são frases acerca do mundo, não são acerca de relações que se sustentam independentemente do pensamento em qualquer domínio de objectos. A sua verdade é formal, vazia: elas são verdadeiras em virtude da sua significação e não de qualquer conteúdo empírico. Este esqueleto do sentido é, ele próprio, segundo Wittgenstein (e isto é uma posição específica de Wittgenstein,, no *Tractatus*, acerca da natureza da lógica com a qual não temos que concordar) vazio de sentido (*Sinnlos*).
- (2) Temos depois frases com conteúdo empírico, nas quais algo com sentido é dito acerca do mundo, e que podem ser verdades ou falsidades
- (3) Temos transgressões do sentido, frases que não são propriamente falsas mas que não têm significado formal ou factual (são *Unsinnig* ou sem sentido). É aqui que fica Heidegger.

Aquém disto, é importante notar que para Wittgenstein temos que considerar como é sequer possível que a linguagem seja representacional, e a sua ideia é que isso é possível porque pela linguagem nos fazemos imagens dos factos (*Wir machen uns Bilder der Tatsachen*). Como o fazemos? Impondo uma forma. Que forma? Aquela que Frege e Russell mostraram existir no nosso pensamento - a forma lógica. Fazer análise lógica será então expôr a natureza e limites da linguagem, e a estrutura de linguagem é aquilo de que nos servimos para fazer imagens dos factos. Isto é assim no pensamento em geral (e não apenas no uso da matemática em ciências várias). Wittgenstein formula esta ideia de uma forma muito kantiana, dizendo que a lógica é transcendental, que as suas proposições não dizem nada, apenas mostram as propriedades formais do mundo, o que significa entre outras coisas que ele pensa que a lógica não é, ao contrário do que defendeu por exemplo B. Russell, uma ciência dos componentes mais abstractos e gerais da realidade mas vazia de sentido.

De acordo com a doutrina do *Tractatus*, tudo o que se deixa pensar deixa-se pensar claramente (é este o âmbito das ciências naturais). No entanto, a teoria do pensamento não pára aqui –Wittgenstein acrescenta que

acerca daquilo de que não podemos falar claramente devemos calar-nos, mas acrescenta ainda (embora esse aspecto do *Tractatus* não interessasse aos Positivistas Lógicos) que ‘aquilo de que não se pode falar’ pode ser ‘mostrado’. É certo também que a intenção do próprio *Tractatus* (mapear a fronteira do dizível e do indizível) transgride essa mesma fronteira, corresponde àquilo que de acordo com a doutrina do próprio *Tractatus* não pode ser feito (daí a célebre imagem da escada que utilizamos para subir e que depois deitamos fora, que aparece no fim do *Tractatus* e que se refere ao que é dito no próprio *Tractatus* -basicamente refere-se ao estatuto da teoria da linguagem feita pela linguagem).

Na verdade não é preciso procurar muito para ver no *Tractatus* aspectos que não encaixariam facilmente na agenda anti-metafísica e cientificista dos Positivistas Lógicos – basta olharmos por exemplo para duas concepções de mundo que aparecem no *Tractatus*. Por um lado temos o *mundo como tudo aquilo que acontece*, o mundo da ordem de factos, é esse mundo que é modelizado pela linguagem. Temos depois o *mundo como totalidade* e aí ‘é da ordem do místico que o mundo (sequer) seja’.

O *Tractatus* não é uma obra de um filósofo-cientista (como eram e queriam ser os positivistas lógicos) mas de um leitor de Schopenhauer e seguidor de Tolstoi, alguém que busca a claridade racional mas a delimita ao âmbito científico-natural e não quer deixar uma suposta racionalidade interferir com o âmbito do que nos importa mais mas não pode ser formulado claramente. Esse é para o autor do *Tractatus* o âmbito do ético, do estético, do místico. (É isto que justifica que no *Tractatus* sejam ditas coisas como ‘Se o bem e o mal alteram o mundo, então alteram a totalidade do mundo, não o que pode ser expresso em linguagem e conhecido cientificamente’ ou ‘O mundo pode ser o mesmo do ponto de vista dos factos e totalmente transfigurado ética e esteticamente’<sup>94</sup>). À parte o beco sem saída em que afirmações como estas deixam a tentativa de uma ética ou política racional – um dizer o que fazer argumentado – torna-se no mínimo claro que para Wittgenstein o pensamento não se reduz ao âmbito científico-natural.

Foi então este livro, o *Tractatus*, que inspirou a ‘filosofia científica’ do Positivismo Lógico que veio a seduzir Abel Salazar e foi a partir destas ideias acerca de linguagem e pensamento que os positivistas lógicos formularam o célebre critério de significação. A doutrina acerca daquilo que é cognitivamente significativo no que nós podemos pensar – o chamado critério de significação - resume o intuito programático do Positivismo Lógico. Interessa por isso compreender o critério de significação. (Assume-se

<sup>94</sup> Ver especialmente as proposições 6.

que aquilo a que chamamos conhecimento toma a forma de ‘frases’, asserções, declarações de que p é o caso). O critério de significação afirma o seguinte:

Uma frase é cognitivamente significativa SSE é ou analítica ou em princípio empiricamente verificável ou falsificável. Senão, trata-se de uma pseudo-enunciação, sem sentido.

Este critério está sujeito a contestações várias (a mais célebre é a que pergunta pela forma como o critério se aplica à sua própria formulação), mas convém notar antes de mais que ele está moldado de modo a incluir como cognitivamente significativas frases aparentemente vazias de conteúdo empírico mas que são importantes para o conhecimento, nomeadamente as frases lógicas. Estas não são verificáveis nem sequer em princípio, pois não têm conteúdo empírico. No entanto não é possível pensar conteúdos empíricos, nomeadamente científicos, sem estruturas desse tipo.

## 5. Do Positivismo Lógico a Quine

Sob a influência do Positivismo Lógico, a filosofia torna-se sobretudo teoria do conhecimento, e a teoria do conhecimento envolve a filosofia da ciência e a filosofia da linguagem, uma combinação de princípios empiristas com análise lógica. Áreas éticas e estéticas são ‘expulsas’ como emotivistas, sem valor cognitivo. Mas porque é que Abel Salazar, um artista, um homem eticamente empenhado, se deixou seduzir por este movimento e foi um grande divulgador? Já sugeri que sobretudo devido ao efeito de antídoto relativamente ao ‘filosofismo’ dominante entre nós.

Nada nos obriga a pensar, no entanto, que o que Abel Salazar procurava na filosofia (e que se pode talvez formular assim: é necessário analisar e clarificar o sentido daquilo que é pensado) foi bem satisfeito pelo Positivismo Lógico. Aliás, olhemos de novo. O critério de significação dos positivistas lógicos leva a crer que podemos separar, nos enunciados da cognição, os momentos em que ‘respondemos’ à linguagem (seria o caso das frases analíticas – há certas coisa que pensamos que dependem exclusivamente da linguagem, como dizer ‘Um triângulo tem três lados’) dos momentos em que ‘respondemos’ à experiência (seria o caso das frases sintéticas, com conteúdo empírico, verificáveis, como ‘Esta mesa é vermelha’). É esta forma de conceber a relação entre linguagem e experiência que Quine irá atacar num artigo que é talvez o mais importante texto de epistemologia do século XX, um artigo intitulado *Two Dogmas of Empiricism* (1953). As teses avançadas no artigo definem o ambiente adverso às teses do Positivismo Lógico que caracteriza a filosofia analítica contemporânea.



Antes de dizer quais são os dogmas a que Quine se refere em *Two Dogmas of Empiricism* e quem os defende, quero sublinhar que apesar de se poder reponder à pergunta «O que aconteceu depois do positivismo lógico?» com a referência a uma crítica ao empirismo (Dois Dogmas do Empirismo parece um título *crítico* do empirismo), Quine é um empirista. Só que Quine é um empirista sofisticado, que não concorda com o aspecto *fundacionalista* do empirismo clássico e do Positivismo Lógico (o fundacionalismo é uma posição em epistemologia, uma posição acerca de justificação, de acordo com a qual existem crenças básicas auto-justificadas, que fundam todo o conhecimento, e que são, se formos empiristas, crenças sobre sensações ou experiência sensorial imediata (impressões, proposições protocolares)). Basicamente Quine pensa que essa ideia não passou por uma forma correcta de pensarmos sobre a linguagem, e sem a linguagem nós não podemos pensar. Quine é então um crítico do empirismo do Positivismo Lógico, empirismo esse que está ligado ao ‘critério da significação’. Olhemos para o critério de significação: está nele incorporada uma distinção entre o analítico e o sintético. De acordo com essa distinção há momentos/passos do nosso pensamento/discurso em que nos movemos em função apenas do próprio pensamento/discurso (= o analítico. Exemplo: ‘Um triângulo tem três lados’) e momentos/passos do nosso pensamento/discurso em que nos movemos em função da experiência (= o sintético. Exemplo: ‘Esta mesa é vermelha’). De um ponto de vista histórico, a separação analítico-sintético é uma reafirmação da separação entre forma e conteúdo do pensamento, expressa agora em termos de linguagem. Quine não concorda com a distinção radical entre o analítico e o sintético e escreve o célebre artigo *Two Dogmas of Empiricism* (1953) a explicar por que razão não concorda. Antes de mais, Quine considera que a distinção analítico-sintético é um dogma (Dogma-1). Esse Dogma-1 está dependente de um Dogma-2 a que Quine chama ‘reducionismo’. O reducionismo é a ideia segundo a qual quando pensamos e conhecemos existem relações um-a-um entre frases e experiência e todas as enunciações significativas são traduzíveis em enunciações acerca da experiência imediata. Para falar como Quine, de acordo com o Dogma-2 cada frase significativa ‘compareceria sozinha ao tribunal da experiência’. Ora, Quine pensa que isso não é o caso. A oposição de Quine ao reducionismo baseia-se no seguinte. Quine concebe a totalidade do conhecimento humano como uma teia de crenças. Essa teia de crenças ‘toca na experiência apenas nos bordos’. Um efeito desta situação é o facto de a actividade cognitiva não ser uma constante e sistemática re-avaliação de cada crença à luz da experiência. Pelo contrário, existe, de acordo com Quine, bastante ‘espaço de manobra’ quando se decide que crenças se vai deixar cair e que crenças se vai manter de cada

vez que acontece ('nos bordos') uma experiência que entra em conflito com as crenças do agente. Quine considera que nenhuma crença é imune à revisão (nem as crenças melhor entrincheiradas na teia das crenças, correspondentes nomeadamente à lógica). Mas essa revisão, à qual nenhuma crença é imune, não é nunca feita com apoio num veredicto incontestável do 'tribunal da experiência'. Em conclusão: criticar a distinção analítico-sintético redundna na defesa de um tipo de empirismo *holista*, de acordo com o qual quando pensamos não temos relações directas nem com 'factos isolados' nem com 'significados isolados'. As consequências disto que parece um mero ponto epistemológico podem ser vastas para a forma como pensamos sobre o pensamento. Por exemplo, e para o dizer directamente, criticar a distinção analítico-sintético é em última análise criticar a legitimidade da própria separação entre a tarefa da filosofia e a tarefa das ciências no inquérito racional.

E realmente, para Quine, a filosofia e as ciências são contribuições para um mesmo inquérito racional: não existe uma fronteira entre 'metafísica especulativa e ciência natural'. Uma conclusão geral para a filosofia é que não pode haver filosofia puramente analítica. Em termos de teoria do conhecimento isso significa que não pode existir uma 'pura análise apriorista ou formal' do conhecimento e que toda a epistemologia (toda a teoria do conhecimento, da mente, da linguagem) será 'naturalizada'. Ora, parece-me que este talvez seja o dado mais importante para percebermos o estado de várias áreas da filosofia analítica contemporânea (obviamente, e pela razão que mencionei, uso a expressão 'filosofia analítica' muito mais como um termo sociológico, aplicado a uma prática da filosofia, do que como caracterização daquilo que a filosofia faz).

### **Conclusão: Naturalização e Kant.**

A naturalização é quase uma palavra de ordem na filosofia contemporânea – pelo menos na filosofia analítica – ouvida desde a epistemologia à filosofia moral. Fala-se de epistemologia naturalizada por tudo e por nada, quando se lida com teorias filosóficas da mente, da linguagem, da cognição. Isto não quer dizer que a filosofia não pense em si própria como continuando a fazer uma análise conceptual, argumentativa e normativa do pensamento. O que significa é que a filosofia já não des-naturaliza os seus objectos de análise, como quiseram por razões diferentes por exemplo Descartes e os Positivistas Lógicos. Os filósofos já não se eximem de perguntar qual é o lugar na natureza do sujeito transcendental de Kant, mesmo se a pergunta correcta não é acerca de localizações no cérebro. Mas admite-se formulá-la em termos de evolução do cérebro e no cérebro, e é por isso que áreas tais como a filosofia da mente e a filosofia moral,

que continuam a tratar os problemas de Kant (o que é representar?, o que é agir?, etc) estão hoje fortemente ligadas à Ciência Cognitiva e aos estudos que nesta se fazem por exemplo acerca de racionalidade, emoções, conceitos, etc. Mas então porque é que a filosofia continua a ser filosofia e não ciência? Por um lado devido à especificidade da análise conceptual e argumentativa, por outro porque a tarefa de análise do pensamento é a nossa condição definitiva, no sentido em que não poderemos nem ter uma *view from nowhere*<sup>95</sup> sobre a realidade (isso seria só realidade, não pensamento), nem vir a pensar no pensamento a partir de fora, ‘em terceira pessoa’, como se diz usualmente. Não podemos porque o pensamento, a perspectiva, é uma parte essencial daquilo que há. Este é portanto um problema que pensar sobre o pensamento coloca, que é mais propriamente uma condição: nunca poderemos fazer isso a partir de fora, não podemos pensar sobre o mundo como se não houvesse pensamento.

Não me refiro apenas ao facto de existencialmente e pessoalmente estamos envolvidos com o nosso pensamento, comprometidos mesmo, mas ao desdobramento fundamental pensamento / mundo, que gera todas as questões conceptuais, todas as questões acerca de verdade, bem, belo, etc, que não poderemos apagar sem deixarmos de existir. Noutras palavras, mesmo tendo na mão uma anatomia ou análise de segundo grau das várias ‘medidas do mundo’ tomadas pelo pensamento (teorias científicas, morais, abordagens artísticas, etc), tendo sempre no horizonte compreender o que significa ‘pensamento’, nenhuma perspectiva global das formas de pensar como medida do mundo alguma vez coincidirá com o mundo. Porquê? Porque o pensamento, sendo uma ‘medida do mundo’ nunca poderá ser uma *view from nowhere* (Thomas Nagel), uma visão a partir de lado nenhum, uma visão sem perspectiva.

O que é que tudo isto tem a ver com Abel Salazar e com o pensamento como medida do mundo? Vou recapitular o que fiz aqui até aqui. Afirmo que os filósofos sempre quiseram dar (e agora vou dizê-lo nos termos de Wittgenstein) um mapa conceptual do que é dizível e pensável em ciência, moralidade, arte, sempre procuraram perceber como pensamos e qual é a natureza do nosso pensamento. Sugeri que a forma como basicamente hoje gerimos essas tarefas de pensamento vem de Kant. Tracei depois as linhagens da filosofia pós-kantiana que conduziram ao movimento filosófico do qual Abel Salazar se sentiu mais próximo, o Positivismo Lógico. Explicito a inspiração desse movimento em Wittgenstein e procurei esquematizar a forma como a crítica ao movimento, feita por outro dos grandes filósofos do século XX, W. V. Quine, se reflectiu na forma como

<sup>95</sup> A expressão é do filósofo americano Thomas Nagel.

grande parte das tarefas filosóficas são hoje em dia levadas a cabo – tudo isso se sedimentou em torno do termo ‘naturalização’. Sugeri que essa naturalização não nos faz afastar mas sim obriga a reformular as tarefas que Kant prescreveu para a análise do pensamento, e que são, recorde, *o que é representar e conhecer? O que é agir racionalmente? O que é apreciar esteticamente?*. Abel Salazar foi obviamente um pretexto para esta panorâmica do pensamento contemporâneo.

### Referências

Cunha, Norberto, 1997, *Génese e Evolução do Ideário de Abel Salazar*, Lisboa, INCM.

Heidegger, Martin, 1988, *Que veut dire penser?, Essais et conférences*, Paris, Gallimard.

Kant, Immanuel, 1985 (1781/ 1787), *Crítica da Razão Pura*, Lisboa, Fundação Calouste Gulbenkian.

Kant, Immanuel, 1994 (1788), *Crítica da Razão Prática*, Lisboa, Edições 70.

Kant, Immanuel, 1998 (1790), *Crítica da Faculdade do Juízo*, Lisboa, INCM.

Quine, Willard Van Orman, 1953, *Two Dogmas of Empiricism*, in *From a Logical Point of View*, Cambridge MA, Harvard University Press.

Salazar, Abel, 1937, *Contra o filosofismo*, III, *O Trabalho*, 19 Agosto.

Wittgenstein, Ludwig, 1987, *Tractatus Logico-Philosophicus*, Lisboa, Fundação Calouste Gulbenkian.



MENTE



## A MENTE SEGUNDO CHOMSKY

Para além do muito interesse que a figura heterodoxa de Noam Chomsky tem como intelectual—trata-se de um académico que é ao mesmo tempo *insider* e criador teórico, *outsider* e crítico político — é um facto que a sua investigação em linguística, entendida como uma teoria computacional funcionalista da mente-cérebro, abriu caminho para as actuais ciências cognitivas.

A investigação de Chomsky em gramática generativa, que tem como objectivo fazer “ciência natural de um módulo da mente humana”, concebe a teoria da linguagem no quadro de um mentalismo funcionalista e é um exemplo central do cognitivismo simbólico.

O mentalismo característico das ciências cognitivas não envolve qualquer apelo a estruturas estranhas ao mundo físico e impôs-se com Chomsky na linguística como um “anti-behaviorismo” necessário ao estudo da linguagem humana. Este foi um dos primeiros passos da “revolução cognitiva” em psicologia no fim dos anos 50. O ponto de Chomsky era então que se pode continuar a admitir que um comportamento (por exemplo o uso de língua natural num humano) é causado sem necessariamente o ver como resposta provocada pelo ambiente externo. Na sua célebre polémica com B.F. Skinner, Chomsky insiste em distinguir o seu problema em teoria da linguagem (“Quais são os determinantes causais do comportamento verbal?”) da questão skinneriana (“Quais são os estímulos que provocam o comportamento verbal?”). A abordagem skinneriana impede por princípio o acesso da teoria ao conhecimento acerca da estrutura que deve ser segundo Chomsky o *explanandum* básico da teoria linguística. E para Chomsky a compreensão do conhecimento acerca da estrutura só pode ser ganha postulando estados e processos mentais.

Antes de dar alguns exemplos dos estados e processos existentes na mente humana para possibilitar o conhecimento e uso de linguagem de acordo com Chomsky, convém notar que com a proposta recente de um Programa Minimalista (cf. Chomsky, *Minimalist Program*, 1995) Chomsky veio deixar os seus apoiantes de longa data praticamente “sem



armas para responder ao inimigo”, ao pretender reduzir todo o complicado mecanismo do sistema da linguagem, que os sintaticistas passaram décadas a aperfeiçoar, a uma interface entre sistemas de performance, o sistema conceptual-intencional e o sistema articulatorio-perceptual, cuja estrutura é constrangida por considerações pragmáticas e de *design*. No espírito do programa minimalista a linguagem não tem “propriedades intrínsecas”, sendo antes uma solução para um determinado problema de exteriorização, pensada portanto como uma questão de “engenharia” resolvida pela evolução no cérebro humano.

De facto Chomsky sempre esteve interessado em questões de princípio acerca da mente e da linguagem, mesmo se isso significa arruinar algum do minucioso trabalho em modelos dos seus seguidores.

Mas voltando atrás, aos inícios e ao esqueleto filosófico do pensamento de Chomsky (cf. *Syntactic Structures*, 1957, *Aspects of the Theory of Syntax*, 1965, *Rules and Representations*, 1980 e *Knowledge of Language, its Nature, Origin and Use*, 1986), a originalidade de Chomsky foi a admissão na teoria da linguagem de estados da mente-cérebro (sendo a mente o cérebro tomado num certo grau de abstracção nos termos da tese funcionalista) que sustentam algo de semelhante à dedução, i.e., transformações/computações de representações abstractas segundo regras. A explicação causal far-se-á nesse nível de regras e representações: nesse sentido, a teoria linguística chomskyana é o exemplo paradigmático do tipo de funcionalismo que é o cognitivismo simbólico (simbólico na medida em que as representações postuladas - por exemplo, os SNs, SVs, SPs - são discretas e as regras algorítmicas).

Um humano, neste caso a “parte” do cérebro do humano que é o Sistema de Linguagem, é uma máquina simbólica física que cumpre manipulações de símbolos especificadas por regras que constituem um programa. A Faculdade de Linguagem (FL) é um órgão mental (um órgão na medida em que é “inato”, pois a linguagem não se ganha ou adquire a partir do ambiente: deve-se falar do crescimento ou desenvolvimento das estruturas de linguagem até um estado estável e não de aquisição por generalização indutiva a partir da experiência) e um órgão-programa. Uma explicação empírica/indutiva das estruturas de linguagem é trivial por evitar a questão crucial da natureza das estruturas da mente-cérebro que constituem o Conhecimento de Língua em cada indivíduo.

Falar uma língua natural é portanto um comportamento regido por regras e a explicação desse comportamento envolve estados e processos mentais cujas regras formais constituem conhecimento inconsciente e inacessível à consciência (é para isto que Chomsky usa o verbo *cognize*). Isto significa que o melhor método para conhecer as regras formais da língua

do falante nunca é perguntar-lhe. Aquilo que é *cognized* no indivíduo é estudado via Gramática Formal pelo linguista (a expressão ‘Gramática Formal’ pode ser entendida como conscientemente ambígua, na medida em que tem dois sentidos: a “teoria do linguista” e “aquilo que existe na mente”).

O conceito de Gramática Formal vem suprir a insuficiência que Chomsky viu na linguística de raiz saussuriana e a que se refere como o “esquecimento do aspecto criativo da utilização de linguagens”, a possibilidade de fazer uso infinito de meios finitos. Chamou a essa propriedade generatividade e procurou modelizá-la com Gramáticas Formais. I.e., segundo Chomsky, as línguas naturais não podem ser bem pensadas como inventários de itens de dupla face som/sentido: o conceito de língua é teoricamente pouco interessante se não fôr completado com o conceito de Gramática Generativa, conceito este nascido de uma confluência de preocupações relativas à mente com o entendimento provindo das ciências formais.

A Gramática Formal, objecto matemático, é ao mesmo tempo um modelo da mente, e portanto do cérebro. Modeliza estruturas que são propriedades de sistemas físicos, e não de um corpus de enunciados: ao contrário do que se passava anteriormente na linguística, o objecto do programa de investigação chomskyano foi desde o início, e continua hoje a ser, a mente. Em suma Chomsky acredita na legitimidade daquilo a que chama, seguindo Husserl, o “Estilo de Galileu”, que caracteriza como uma construção de modelos matemáticos a que se atribui maior realidade do que ao mundo normal das sensações.

Segundo o modelo actual dos Princípios e Parâmetros (que não é substituído mas apenas disciplinado pelo Programa Minimalista) são então *cognized* num humano adulto os Princípios e Parâmetros da Gramática Universal (GU), no que constitui um estado estável da faculdade de linguagem. À teoria de FL - 0 (faculdade de linguagem no estado zero) chama-se GU (Gramática Universal), e esta tem que permitir um número indefinido de línguas humanas possíveis. A partir daí em cada indivíduo será atingido o estado estável do conhecimento de uma língua natural particular (cf. para exemplos de Princípios e Parâmetros, Paiva Raposo, 1992, *Teoria da Gramática. A Faculdade da Linguagem*).

A proposta de Chomsky em termos filosóficos é portanto a de uma naturalização da significação via a dimensão formal, algorítmica dos fenómenos de linguagem. A abordagem considera as línguas naturais à imagem das linguagens formais, envolvendo o postulado da existência mental de estruturas formais. O seu ponto de referência é a psicologia individual, o que coloca um problema a que o próprio Chomsky chama o “problema de Wittgenstein” e que considera ser a mais interessante crítica ao enquadramento conceptual da gramática generativa: segundo a teoria

de Chomsky, é possível caracterizar como “regras” os processos da mente-cérebro de um indivíduo isolado, que estará então a seguir regras privadamente. Wittgenstein contesta esta possibilidade (cf. Parágrafo 202 das *Investigações Filosóficas*).

A ideia de “seguir uma regra” é a analogia central do cognitivismo simbólico uma vez que é supostamente na observância de regras que consiste a identidade entre os mecanismos cognitivos num humano e o funcionamento de um programa numa máquina simbólica. Mas exactamente o que é seguir uma regra? Se a minha observância de regras é cega como Wittgenstein pensa ser o caso, por exemplo no uso de conceitos, seguir uma regra não é um saber que eu possuo e sou capaz de explicitar, é antes uma prática “impenetrável” ao (suposto) agente. Que razões restam para afirmar então que uma regra está a ser seguida privadamente por alguém? Para Wittgenstein, ao contrário da pretensão implícita na atribuição da observância privada de regras, são impossíveis asserções factuais acerca de regras, as atribuições de observância de regras a indivíduos são legítimas apenas no contexto de uma comunalidade de respostas e de práticas numa comunidade. Pode-se afirmar no entanto que fazer asserções factuais acerca de regras que estão a ser seguidas sem serem estarem a ser pensadas é, dado o compromisso naturalista, a própria finalidade da gramática generativa. A legitimidade ou não do conceito de regra para processos físicos na mente-cérebro funcionalistamente caracterizados e inacessíveis à consciência é portanto um conflito nuclear na fundamentação conceptual do programa generativista.

Também Quine e Searle atacam o estatuto filosófico que Chomsky pretende dar à gramática generativa. Segundo Quine, a adequação descritiva do funcionamento das regras postuladas pela teoria aos comportamentos linguísticos não é razão suficiente para supor que são esses os mecanismos, e não outros, que existem num cérebro humano. Aliás, para Quine, quando se fala de linguagem não é razoável pensar em propriedades fisicamente codificadas. Se os sistemas de regras que descrevem o comportamento linguístico podem admitidamente ser vários e diferentes entre si, mantendo cada um a adequação descritiva, não será razoável supor que algum deles descreve mecanismos no cérebro. Segundo Quine devemos abster de atribuir qualquer tipo de realidade “mental” ao sistema de regras, falando apenas de adequação dos comportamentos linguísticos ao sistema. A crítica de Quine a Chomsky é uma consequência da sua tese da subdeterminação das teorias pela evidência.

Se segundo Quine Chomsky, com o seu realismo quanto a regras e representações, afirma mais do que é possível, por princípio, afirmar quanto a questões de linguagem, segundo Searle Chomsky não dá provas suficientes daquilo que afirma e seria necessário fazê-lo para poder fazer

afirmações acerca de regras que estão de facto a ser seguidas. Estas afirmações são segundo Searle possíveis. Quando um ser humano, ao contrário de uma máquina, segue uma regra, ele é, segundo Searle, guiado pelo significado ou conteúdo efectivo dessa regra. Por isso, mesmo se várias regras descrevem os acontecimentos e têm o poder preditivo certo, Searle pensa que há apenas uma regra que está a ser seguida, aquela que está a funcionar causalmente. Ora, dado o tipo de regras consideradas por Chomsky não é possível chegar a fazer esta distinção, e portanto essas regras são meras hipóteses descritivas e não verdadeiras regras, i.e., elementos constituintes da mente-cérebro, causalmente eficazes, como Chomsky pretende.

### Referências

Chomsky, Noam, 1957, *Syntactic Structures*, [The Hague], Mouton Publishers.

Chomsky, Noam, 1965, *Aspects of the Theory of Syntax*, Cambridge MA, MIT Press.

Chomsky, Noam, 1980, *Rules and Representations*, Oxford, Blackwell.

Chomsky, Noam, 1986, *Knowledge of Language, its Nature, Origin and Use*, (New York), Praeger.

Paiva Raposo, Eduardo, 1992, *Teoria da Gramática. A Faculdade da Linguagem*, Lisboa, Caminho.



## *DENNETT E A INTELIGÊNCIA ARTIFICIAL*

Dada a sua concepção funcionalista do mental, segundo a qual o facto de um sistema físico ter mente consiste na realização por esse sistema físico de uma determinada organização funcional, Dennett considera que não existe nenhuma diferença essencial entre a inteligência natural e a inteligência artificial. O campo disciplinar da Inteligência Artificial (IA) pode por isso servir como fonte de vários esclarecimentos em filosofia da mente e as experimentações possíveis em IA via a criação de programas e robôs devem ser consideradas como contribuições importantes para a compreensão das mentes naturais.

Pondo em prática esta convicção, Dennett tem-se servido ao longo da sua obra de exemplos de criações da IA para desafiar e ilustrar pontos na teoria da mente. Fê-lo já com o General Problem Solver de Newell, Simon e Shaw (um programa capaz do tipo de raciocínio heurístico chamado análise meios-fins) nos anos 60, continua hoje a fazê-lo com os robôs móveis autónomos de Rodney Brooks (cf. <http://www.ai.mit.edu/people/brooks/brooks.html>) e de Maja Mataric ( cf. <http://www.ai.mit.edu/people/maja/maja.html>) capazes de comportamentos cognitivos no mundo e mesmo de estratégias de cooperação em grupo, e fá-lo, finalmente, mantendo uma proximidade prática com o campo, através da sua participação como “filósofo residente” no projecto da criação de Cog, o robot humanóide do Laboratório de Inteligência Artificial do MIT, dirigido por Rodney Brooks, (cf. [www.ai.mit.edu/projects/cog](http://www.ai.mit.edu/projects/cog)).

Estes programas e robôs são para Dennett experiências mentais reais ou realizadas, mesmo se são experimentações sobre o “mental-possível” que não simulam ou explicam imediatamente o “mental-natural”. São experiências mentais no duplo sentido de partirem da imaginação sobre o que poderia acontecer em determinadas circunstâncias e de serem acerca da mente, e são reais na medida em que conduzem à criação de programas e máquinas, uma vez que a IA é um campo afim da engenharia.

Mas a IA é para Dennett mais do que uma fonte de exemplos e experiências mentais. Dennett entende a IA como uma maneira de fazer

filosofia, mesmo se não é isso que estes cientistas pensam que estão a fazer (cf. Dennett, *Cog as a Thought Experiment*). Essa actividade, fazer filosofia, é entendida por Dennett à maneira de Kant como uma investigação das condições de possibilidade da cognição em geral. Isto quer dizer que a epistemologia pura dos filósofos é para Dennett apenas o caso limite num *continuum* que envolve também as criações da IA. As questões da filosofia e da IA são identicamente abstractas e gerais e visam a emergência no mundo físico de percepção, acção, inteligência e consciência. A IA partilha ainda com a filosofia o método das experiências mentais. Aliás, segundo Dennett, os métodos da IA são um paraíso para filósofos, precisamente porque todo o campo consiste em experiências mentais. Esta afirmação parece falhar o alvo em relação a um domínio de engenharia, no entanto aquilo que Dennett pretende pôr em relevo é o possível desprendimento da questão do mental em relação ao mental biologicamente baseado e actual: este não tem privilégios por princípio, ou só os teria como consequência de um essencialismo acerca das origens que para Dennett é indefensável. Os fenómenos mentais biologicamente baseados e actuais situam-se por isso conjuntamente com outras formas possíveis de mentes num mesmo espaço do *design*.

Mas estas experiências mentais da IA têm uma vantagem em relação à da filosofia pura: como são experiências mentais controladas via “próteses” (os computadores) as hipóteses podem ser refutadas em concreto, no sentido próprio daquilo que acontece numa experimentação, enquanto que nas experiências mentais dos filósofos nada pode nunca correr mal, pois não existindo o constrangimento que representa o facto de se ter que construir o modelo que se imagina: não há assim lugar para as surpresas com que se aprende. A Inteligência Artificial e Vida Artificial têm, na sua maneira de colocar a questão de Kant, uma vantagem não apenas em relação à filosofia pura mas também em relação às investigações cognitivas a que Dennett chama “abstémias”, por exemplo aquelas que são estritamente neurobiológicas e procedem “neurónio a neurónio”. Essa vantagem é a seguinte: é mais fácil deduzir competências comportamentais dos mecanismos que se contruiu do que elaborar hipóteses acerca de mecanismos internos de caixas negras cujo comportamento se observou (o que é a posição do cientista “abstémio”, por exemplo do neurocientista, por comparação com o teórico da IA).

Este louvor das vantagens da passagem à acção que a tentativa de constituição de uma IA representa por contraposição à investigação de uma inteligência “natural” tem ligações com uma ideia directriz de Dennett quanto à investigação da mente: para Dennett as Ciências Cognitivas enquanto ciências da inteligência “natural” ganharão se se assumirem como uma prática de Engenharia Invertida (*reverse engineering*).

Segundo Dennett, as ciências cognitivas, enquanto ciências da mente-cérebro, investigam soluções de *design* biológico, soluções já realizadas e frequentemente imperfeitas, muito distantes da melhor solução possível, inclusive porque partem da necessidade de aproveitar o material pré-existente na evolução biológica, nunca partindo obviamente do zero. Ao sublinhar esta imperfeição e oportunismo das soluções biológicas de *design*, Dennett pretende combater aquele que considera ser um defeito característico de filósofos e outros cientistas cognitivos, a tendência a considerar que, em relação ao estudo da mente e na exclusão do dualismo, restam a física e a matemática, ocultando a perspectiva teoricamente menos pura da biologia-como-engenharia (e por isso parente da IA...) como chave para o estudo da mente.

A maneira de trabalhar da IA é segundo Dennett - e isto é um louvor - ambiciosa e “saltadora de etapas”: é uma abordagem *top-down* que nos fará avançar mais rapidamente na compreensão da cognição do que as descrições *bottom-up*, nomeadamente as neurobiológicas.

Por tudo isto, Dennett nunca deixa de recomendar aos filósofos a familiaridade com os problemas de implementação relacionados com os seus problemas mais abstractos acerca da mente. Por exemplo, aconselha sempre a familiaridade com o Jogo da Vida de John Horton Conway, baseado na teoria dos autómatos celulares e antecessor das investigações no campo da Vida Artificial (cf. [http://www.yahoo.com/Science/Artificial\\_Life/Conway\\_s\\_Game\\_of\\_Life/index.html](http://www.yahoo.com/Science/Artificial_Life/Conway_s_Game_of_Life/index.html)) como uma maneira prática de considerar problemas relevantes para a filosofia da mente, tais como a identidade através do tempo, a causação, os níveis de explicação, (cf. Dennett, *Real Patterns*, <http://www.tufts.edu/as/cogstud/publist.htm>). Correlativamente, aos cientistas que trabalham em IA, Dennett recomenda algumas leituras filosóficas (como Hume, Ryle, Wittgenstein, Millikan...) oferecendo a possibilidade de pela consciência histórica serem capazes de discernir as más ideias perenes no tratamento da cognição que certamente os perseguirão.

Como foi dito, a fundamentação da posição de Dennett em relação à IA é o facto de não admitir a possibilidade de uma distinção de essência entre Inteligência “natural” e inteligência não natural. Na medida em que a caracterização dos fenómenos mentais é funcional, a discriminação entre os substractos que os realizam não está justificada. A IA pode não ser natural por não ser resultado da evolução biológica mas, se a definição de mental como funcional é legítima, ela é genuína e pode-nos fazer aprender muito sobre os problemas específicos dos sistemas de representação de conhecimento dos quais a nossa mente-cérebro é um caso particular.

No entanto para alguém como Fodor (cf. Baumgartner e Payr, 1995 *Speaking Minds, Interviews with 20 Eminent Cognitive Scientists*), esta



fascinação de Dennett pela IA como “maneira de investigar a mente” está completamente mal dirigida: segundo Fodor, a ciência da mente deve ser psicologia cognitiva e não IA. A IA é engenharia, construção de máquinas e como tal tem interesse científico por si, mas não é o bom caminho para o estudo da mente. Uma coisa é tentar entender o pensamento, outra é construir máquinas inteligentes: como Fodor diz, também não se faz física simulando o universo. Essa seria uma ideia muito pouco razoável, porque os fenômenos que observamos no mundo são efeitos de complicadíssimas interações nos mecanismos subjacentes e na prática impossíveis de reconstituir e isto também se aplica à física da cognição. Fodor pensa que fazer ciência é simplificar para compreender e não simular, e assim deve fazer-se também a ciência da mente. Mas Dennett quer andar depressa e a IA parece-lhe o atalho ideal.

O diferendo entre as posições de Fodor e Dennett quanto à IA é recondutível à questão de saber se há ou não casos em que a simulação é replicação. Para quem defende uma teoria funcionalista da natureza de alguma coisa, nomeadamente do mental, é difícil resistir à conclusão de que há casos em que a simulação é replicação.

Além do mais é preciso decidir até que escala uma simulação teria que fazer-se para obter propriedades mentais. No caso de as propriedades mentais serem funcionais e de nível mais elevado do que a estrutura física da matéria que Fodor considera irreprodutível, a sua objecção poderia não ser eficaz.

### Referências

Baumgartner, P. & Payr, S. *Speaking Minds*, Princeton NJ, Princeton University Press.

Dennett, D., 1978, *Brainstorms*, MIT Press.

Dennett, D., 1998, *Brainchildren*, Penguin.

Dennett, D., 1998b, *Real Patterns*, in Dennett 1998.

Dennett, D. Cog as a Thought Experiment, <http://www.tufts.edu/as/cogstud/publist.htm>

Dennett, D., Things about things, <http://www.tufts.edu/as/cogstud/publist.htm>

Haugeland, J., 1985, *Artificial Intelligence, The Very Idea*, MIT Press

## *ALGUNS PROBLEMAS DE FILOSOFIA DA INTELIGÊNCIA ARTIFICIAL*

O objecto de investigação da IA é a natureza da cognição, e ter uma teoria da cognição – nomeadamente de fenómenos como a categorização e identificação de objectos, a resolução de problemas, a decisão, a consciência — sempre foi a ambição dos filósofos. Por isso, desde os inícios da disciplina, os filósofos interessam-se pelas investigações em IA e pelo que elas podem mostrar quanto à inteligência humana e quanto à inteligência em geral.

Curiosamente, os filósofos têm adoptado posições extremas quanto à IA, desde a defesa de uma impossibilidade por princípio da criação de inteligência e consciência não naturais, por razões várias, até á convicção de que através da IA poderá surgir uma concepção mais geral e mais abstracta sobre a natureza da inteligência, que coloque os humanos e todos os seres biológicos inteligentes como apenas casos particulares de um fenómeno geral. Basicamente, a ideia é que se qualquer sistema com o tipo correcto de organização funcional pode ser inteligente e até mesmo consciente, e se essa condição pode ser formulada independentemente da matéria de que fôr constituído o sistema e independentemente das suas origens, outros sistemas que não os humanos e outros seres biológicos poderão, pelas mesmas razões que estes, ser inteligentes e conscientes.

### 1. A crítica filosófica à IA

Hubert Dreyfus (<http://socrates.berkeley.edu/~hdreyfus>) e John Searle (<http://socrates.berkeley.edu/~jsearle>) são dois filósofos com posições críticas face à IA.

Hubert Dreyfus

Hubert Dreyfus foi talvez o primeiro filósofo a notar a sobreposição de interesses da IA e da filosofia. Escreveu como resultado desse encontro

textos polémicos nomeadamente um relatório intitulado *Alchemy and Artificial Intelligence* (Rand Corporation, Santa Monica, California, 1965) e depois o livro *What Computers Can't Do* (1972), criticando as pretensões da IA. Dreyfus parte do princípio de que a IA trabalha os mesmos problemas que a filosofia, problemas como a natureza do entendimento e do conhecimento. A sua crítica insiste sobretudo na relação entre os modelos então desenvolvidos na IA, baseados na ideia de mente como sistema simbólico e na ideia de inteligência como resolução de problemas e a tradição racionalista e intelectualista em filosofia, tradição essa que Dreyfus critica. Para Dreyfus, a IA repetia portanto os erros 'intelectualistas' da filosofia, erros já apontados por filósofos como Heidegger, Wittgenstein e Merleau-Ponty.

A crítica de Dreyfus tinha como centro a ideia de que o pensamento não pode ser considerado como consistindo em representações simbólicas do mundo, nem a mente como um sistema governado por regras ocupado com a resolução de problemas, pois essas definições excluem partes importantes e básicas do mental. As 'partes excluídas' do mental eram para Dreyfus por exemplo os movimentos corporais e o reconhecimento de padrões e estas estariam subjacentes à própria possibilidade das habilidades (skills) explícitas envolvidas nas representações e na resolução de problemas.

Os escritos de Dreyfus foram muito mal recebidos, mas apresentavam uma perspectiva sóbria sobre alguns exageros de previsão que acompanharam os primórdios da IA. Apesar de algumas das previsões do próprio Dreyfus terem sido simplesmente desmentidas (por exemplo a ideia de que um computador não poderia vencer um humano num jogo de xadrez), o que Dreyfus queria criticar era a suposição apriorista não argumentada – defendida, na sua opinião, por exemplo por A. Newell, H. Simon e M. Minsky – segundo a qual regras e representações seriam a natureza da mente. De acordo com o intelectualismo dessa visão, até a percepção seria pensamento e resolução de problemas através da manipulação de regras. Esta concepção ocultava o conhecimento do fundo (*background knowledge*) ou de senso comum, que na perspectiva de Dreyfus era fundamental para a própria cognição. O conhecimento do fundo não é conhecimento de factos mas sim, no caso das pessoas, aquilo que elas sabem sem saberem que sabem e que nunca foi aprendido (como por exemplo que as pessoas se movem mais facilmente para a frente do que para trás, num exemplo de Dreyfus, ou que se se entornar água em cima da toalha ela passará para as pernas de quem está por baixo).

Dreyfus insistiu ainda noutras características do pensamento humano, ausentes nas simulações computacionais que então considerava como

exemplos: as margens da consciência (*fringe consciousness*), a tolerância da ambiguidade, a ligação a um corpo no mundo que organiza e unifica a experiência de objectos e as impressões subjectivas, a capacidade de aborrecimento, cansaço e perda de motivação e os propósitos e interesses que gerem o confronto do sujeito com a situação no mundo, fazendo com que nem todos os factos no mundo sejam igualmente relevantes num dado instante, com que o mundo não seja 'liso'.

O teor das críticas de Dreyfus foi então para cá incorporado no natural desenvolvimento da IA e actualmente Dreyfus admite a proximidade entre os princípios do conexionismo e a tradição anti-racionalista em filosofia.

### John Searle

John Searle foi conduzido às Ciências Cognitivas a partir do seu trabalho em filosofia da linguagem. É um crítico famoso dos limites do modelo computacional da mente. Searle pensa que as ciências cognitivas são um campo de investigação excitante mas fundado num erro conceptual acerca da natureza da mente, e que a 'hipótese da IA forte' (a ideia segundo a qual qualquer sistema que implemente o programa correcto poderá ter realmente uma mente e consciência, logo não será apenas uma simulação), é refutável via o Argumento do Quarto Chinês. O Argumento do Quarto Chinês pretende ainda 'refutar' o Teste de Turing (Turing 1950).

O artigo *Minds, Brains and Programs*, onde o argumento é defendido, apareceu na revista *Behavioral and Brain Sciences* em 1980, juntamente com vinte e oito respostas de críticos. Desde então a experiência mental do Quarto Chinês é incontornável na filosofia da mente, associada à ideia de que não se pode falar de processos mentais sintácticos sem falar de semântica (e, de facto, Searle liga a semântica à consciência).

A experiência mental do Quarto Chinês (QC) consiste no seguinte: alguém, que não fala chinês, está fechado dentro de um quarto onde há símbolos chineses em caixas. Tem um livro de instruções em inglês, que explica como combinar os símbolos chineses e como enviar sequências de símbolos chineses para fora do quarto, quando são introduzidos no quarto outros símbolos chineses, através de uma pequena janela. A pessoa que está dentro do quarto não sabe nada acerca disso, mas as pessoas que estão fora do quarto chamam aos símbolos que introduzem 'perguntas' e aos símbolos que saem 'respostas'. O sistema fala portanto chinês, na perspectiva das pessoas que estão fora. Então, o sistema passa o Teste de Turing, embora a pessoa lá dentro saiba que não percebe uma palavra de chinês. Searle afirma que a experiência mental do QC torna clara a possibilidade

de um sistema que tem ‘intencionalidade atribuída’ mas não ‘intencionalidade intrínseca’ ou ‘semântica genuína’.

O problema é saber exactamente que é que o argumento de Searle mostra. Antes de mais, note-se que o QC é mais propriamente uma parábola, e que posto em forma de argumento daria qualquer coisa como o seguinte:

Os programas são sintácticos  
 A sintaxe não é suficiente para a semântica  
 As mentes têm semântica  
 Implementar um programa é insuficiente para haver mente

O QC mostraria então que a mente não é um programa e que por isso programar apropriadamente alguma coisa nunca poderia dar-lhe uma mente, já que as propriedades formais não constituem a ‘intencionalidade genuína’.

O filósofo David Chalmers parodia este argumento do seguinte modo (Chalmers 1996: 327):

As receitas são sintácticas  
 A sintaxe não é suficiente para ser-saboroso  
 Os bolos são saborosos  
 As receitas não são suficientes para fazer bolos

Chalmers pretende mostrar que o argumento de Searle não distingue entre sintaxe e implementação de sintaxe, entre um programa ‘na prateleira’ e um programa correndo numa máquina física.

O próprio Searle sublinha sempre que o seu argumento não tem nada a ver com um estágio evolutivo particular da tecnologia, mas antes diz respeito a princípios conceptuais. O que o QC pretenderia fazer ver seria que o cognitivismo – no qual é central a ideia de que ‘a mente está para o cérebro como o software para o hardware’ – está errado ao considerar que não existe nada de essencialmente biológico acerca da mente humana. É esta posição que tem como corolário a defesa da IA forte, a ideia, recorde-se, de que qualquer coisa que implementasse o programa correcto poderia ter realmente uma mente. Para Searle, pelo contrário, a mente é essencialmente consciência, e a existência de consciência é um facto biológico.

De facto, na formulação inicial do Quarto Chinês, Searle fala mais propriamente de semântica, e afirma que ‘a sintaxe não é suficiente para a semântica’. Mas o que Searle pensa é que não há possibilidade de considerar qualquer coisa como mental a não por relação com a consciência.

A fraqueza da posição de Searle é não dar um critério claro do que se entende por intencionalidade originária ou intrínseca por oposição a intencionalidade atribuída, as noções de que se serve para analisar o QC. O único critério que Searle dá são os poderes causais que alguns sistemas físicos (nomeadamente os cérebros humanos) teriam e outros não.

Variadíssimas objecções foram levantadas ao Quarto Chinês, algumas das quais as a seguir listadas:

- a. Com o QC Searle pretende provocar a nossa identificação com o ser humano que simula ‘à mão’ um programa de IA. Ora, isso não faz sentido pois Searle oculta a questão da escala: um ser humano não poderia simular à mão um programa de IA suficientemente complexo para a tarefa em causa. Isso envolveria meses ou anos, para não falar de horrível aborrecimento. Searle passa ‘por baixo do pano’, com o apelo à identificação, uma concepção irrealista da relação entre inteligência e manipulação simbólica, ao impedir que se considere a enorme diferença de complexidade entre o nível do programa e o nível da pessoa e ao sugerir que ‘sentamos’ a falta de entendimento da pessoa ao levar a cabo um programa.
- b. A resposta básica à situação do QC é a resposta dos sistemas: é um erro imputar o entendimento ao executor do programa (incidentalmente animado). O entendimento *pertence ao sistema como um todo*, que inclui os “pedaços de papel” com as regras/símbolos (aliás, nos nossos neurónios também não há entendimento do português que falamos, e no entanto, nós, o sistema global, falamos português, e a falta de intencionalidade genuína dos nossos neurónios não constitui prova da falta de entendimento genuíno em nós).

Vários autores (cf. Hofstadter & Dennett 1981) contrapuseram a Searle, situações semelhantes à do QC, mas em que a nossa intuição vai em sentido contrário ao que se passa no QC.

O psicólogo canadiano Zenon Pylyshyn propôs o seguinte: e se se fosse substituindo célula a célula as células do nosso cérebro por chips programados para manterem exactamente as mesmas funções? Neste caso continuaríamos a falar e agir do mesmo modo, mas Searle teria que dizer que deixaríamos de significar (*mean*) progressivamente (a nossa interioridade desapareceria de nós, sem ninguém notar, nem nós próprios). O problema é que neste caso não ‘tendemos’ a ter essa intuição – o que mostra,

de novo, que Searle não dá critérios para dizer quando é que a intencionalidade genuína está presente ou ausente do sistema.

O filósofo americano John Haugeland propõe a seguinte caso: o cérebro de uma mulher real tem uma lesão, e há um ‘agente artificial’ que intervém para “conduzir” os neurotransmissores de forma idêntica ao que antes da lesão acontecia naturalmente. O cérebro funciona então exactamente como se a mulher estivesse saudável. Haugeland pergunta a Searle se esta mulher pensaria ou não teria entendimento nenhum do seu próprio pensamento. O próprio Searle concorda que nesse caso devemos olhar para o ponto de vista da mulher e não para o do agente (embora este pudesse, por hipótese ter o seguinte ponto de vista: ‘Loucos! Não lhe dêem atenção! Ela é um fantoche cujas acções são causadas por mim!’)

O ponto destes contra-exemplos imaginários é mostrar que o QC depende exclusivamente da nossa intuição, e as nossas intuições podem ser feitas variar (é o caso das intuições do próprio Searle). Por isso nenhum bom argumento acerca daquilo a que Searle chama intencionalidade intrínseca deverá depender exclusivamente de intuições.

Ninguém nega que a experiência de Searle é provocadora, sobretudo porque ela diz respeito ao facto de no estudo da cognição estarem sempre em causa vários níveis. Aliás, Hofstadter sugere que o que precisamos para pensar no QC é precisamente do conceito de níveis de implementação, da ideia de que um sistema pode emular outro sistema e, assim sendo, o ‘encaixe’ de uma ‘máquina’ noutra dá origem a ‘máquinas virtuais’. Debaixo de cada máquina virtual há sempre outra máquina, sendo apenas num certo sentido a máquina de baixo a ‘máquina real’. Em princípio os níveis estão selados uns aos outros (como a pessoa que não fala o chinês que o sistema fala, como os neurónios que não falam o português que nós falamos). Mas os níveis poderiam ‘comunicar’ e Hofstadter sugere que talvez isto seja o que se passa por exemplo quando um sistema humano aprende uma nova língua, que quando é bem aprendida não corre simplesmente sobre a primeira, como uma espécie de parasita de software, antes deixa de precisar de ser traduzida e torna-se mais fundamental. Hofstadter associa, aliás, esta ‘comunicação’ entre níveis de um sistema à consciência (que, note-se, é o verdadeiro problema de Searle no QC).

Em suma, saber o que quer dizer ‘nível’ parece essencial para conceber a natureza do pensamento e da consciência. Enquanto os níveis estão isolados uns dos outros, como no Quarto Chinês, tudo é claro, mas se eles interferem e se confundem deixa de o ser. Note-se que o próprio Searle admite que há dois níveis no QC, mas não admite que poderia haver dois pontos de vista (ou abrir-se-ia uma caixa de Pandora e a experiência e a

mente começariam a espalhar-se por toda a parte - de facto Searle pensa que o pampsiquismo é o risco do cognitivismo).

Aqueles que como Dennett e Hofstadter se recusam a admitir que o QC mostre qualquer coisa de importante (i.e. que evidencie a ausência do que quer que seja) defendem que mentes existem em cérebros e podem vir a existir em máquinas programadas. Se e quando essas máquinas vierem a existir os seus poderes causais não derivarão das substâncias de que elas são feitas mas dos programas que instanciam. E poder-se-á saber que elas têm mentes falando com elas e ouvindo com atenção o que têm para dizer...

Em *A Redescoberta da Mente* Searle apresenta um segundo argumento anti-cognitivista. Esse argumento é, grosseiramente, o seguinte:

A sintaxe não é uma propriedade física  
 O cognitivismo supõe o tratamento de fenómenos físicos como sintáticos  
 O cognitivismo incorre na falácia do homúnculo  
 (i.e., o cognitivismo descreve como se fossem propriedades naturais propriedades que só existem 'para um observador' - no caso, a sintaxe, que não é para Searle uma característica do mundo natural mas uma interpretação por um observador)

As críticas de Searle ao cognitivismo podem ser unificadas notando que para Searle, o princípio da conexão é fundamental para pensarmos na mente. O princípio da conexão é o princípio segundo o qual só entendemos como mental o que é actualmente inconsciente porque o entendemos como conteúdo possível da consciência. A consciência é a essência da mente, i.e., só entendemos alguma coisa como mental e não física porque a reportamos à consciência (senão como distinguiríamos um neurónio de uma memória?).

Searle chama à sua posição em teoria da mente um materialismo não reducionista, embora os seus críticos afirmem que não é possível distinguir esta solução anti-reducionista, segundo a qual os Estados Mentais são causados por e são características de, mas não são idênticos a, Estados Cerebrais de um dualismo de propriedades segundo o qual os fenómenos mentais envolvem propriedades que não são físicas. No entanto, Searle de modo algum admite ser um dualista.

Searle é anti-reducionista porque defende que a consciência, por ser ontologicamente subjectiva, torna a redução impossível. Assim, todas as teorias reducionistas do mental falhariam na distinção entre um *zombie* e um ser consciente. A posição de Searle depende portanto da noção de 'redução'.



A redução é um termo e um problema da filosofia da ciência, um termo para a análise de alguma coisa identificada num nível de descrição em termos de outro nível, mais fundamental, permitindo-nos dizer que a primeira coisa não é mais nada além da segunda. O problema da redução aplicado ao mental consiste em saber se o mental poderia ser descrito em termos totalmente não mentais. Searle pensa que não, devido à subjectividade ontológica da consciência.

Alguma coisa é ‘ontologicamente subjectiva’ para Searle se não conseguimos descrever as suas características em terceira pessoa, que é o que estamos a fazer mesmo quando fazemos por exemplo neurofisiologia da consciência. Para Searle, a consciência é uma propriedade física do cérebro apesar da sua subjectividade ontológica e ela é irreduzível a qualquer outra característica física. A subjectividade ontológica não deve ser confundida com a subjectividade epistemológica. Uma coisa é a objectividade como a boa tentativa de eliminação das pre-concepções subjectivas, eliminação essa que faz parte do espírito da ciência, outra é a afirmação de que o mundo não contém elementos irreduzivelmente subjectivos. Não há razão para a segunda afirmação. A primeira noção de objectividade é epistemológica, a segunda é ontológica.

Ora, se se aceita esta distinção, a verdadeira questão é a seguinte: Como é que podemos ter uma concepção objectiva dos factos ontologicamente subjectivos da consciência? A isto Searle responde com o naturalismo biológico: a ideia de que a consciência é uma característica biológica do cérebro humano e de outros animais e uma propriedade emergente (como a liquidez, a partir da energia cinética molecular).

Searle põe a questão nestes termos porque pensa que muitos materialistas estão errados quando pensam que sem admitir a redução se aceita necessariamente o dualismo.

## 2. Os computadores, a filosofia e o funcionalismo

Se através de Dreyfus e Searle se apontou algumas críticas aos fundamentos filosóficos da IA, é preciso por outro lado sublinhar que, não especificamente a IA mas a própria existência de computadores, foi extremamente importante para toda a filosofia da mente dos últimos 40 anos, e central na ideia de funcionalismo.

O funcionalismo é um posição filosófica hoje muito espalhada quanto à natureza da cognição (teorizada por exemplo por H. Putnam). De acordo com o funcionalismo, é essencial para conceber a natureza da cognição a distinção entre hardware e software. I.e., o funcionalismo acentua aquilo a que L. Moniz Pereira (Moniz Pereira 1990) chama a «não obriga-

toriedade de correspondência entre o processamento de uma certa função cognitiva e o suporte material que executa esse processamento». Para além de possibilitarem exemplos reais de intencionalidade e racionalidade em sistemas físicos sem necessidade de ‘homúnculos’ ou observadores cartesianos, os computadores desligaram conceptualmente a ideia de inteligência da particular realização biológica das inteligências nos humanos, e na medida em que a inteligência foi ao longo do tempo considerada definida por aquilo que é ser humano, mudaram a concepção do humano.

### 3. Limites?

Actualmente, o debate filosófico em torno da naturalização da cognição, centra-se em dois aparentes limites a uma teoria integralmente científica, feita exclusivamente em terceira pessoa: os estados qualitativos, i.e., o experienciar, o *what-it’s-like-to-be* nageliano e a racionalidade, com os seus aspectos normativos.

### Referências

Boden, Margaret (ed), 1990, *Philosophy of Artificial Intelligence*, Oxford, Oxford University Press.

Chalmers, David, 1996, *The Conscious Mind*, Oxford, Oxford University Press.

Dennett, Daniel e Hofstadter, Douglas (eds.), 1981, *The Mind’s I*, New York, Bantam Books.

Dreyfus, Hubert, 1972, *What Computers Can’t Do*, New York, Harper & Row.

Minsky, Marvin, 1985, *The Society of Mind*, New York, Simon & Schuster.

Moniz Pereira, Luís, 2000, *Inteligência Artificial, Intelectu*, 3.

Newell, Allen, 1990, *Unified Theories of Cognition*, Cambridge MA, Harvard University Press.

Pinto, João Alberto, 2007, *Materialismo, Superveniência e Experiência*, Porto, Campo das Letras.

Putnam, Hilary, *Minds and Machines*, 1960, in *Philosophical Papers*, vol. 2, Cambridge, Cambridge University Press, 1975.

Searle, John, 1989, *Minds, Brains and Programs*, *Behavioral and Brain Sciences*, 3.

Searle, John, 1998, *A Redescoberta da Mente*, Lisboa, Instituto Piaget.

Simon, Herbert, 1969, *The Sciences of the Artificial*, Cambridge Mass., MIT Press.

Turing, Alan, *Computing Machinery and Intelligence* (1950), in Dennett & Hofstadter 1981.

Para uma bibliografia relativa às questões filosóficas da IA, cf. a 4ª parte da Bibliografia da Filosofia da Mente organizada por David Chalmers, em <http://www.u.arizona.edu/~chalmers/biblio.html>. (Atualmente (2009): *A Bibliography of Philosophy of Mind and the science of consciousness* em <http://consc.net/mindpapers>, Parte 6)

## *CRITÉRIOS DE EXPERIÊNCIA (WITTGENSTEIN, MALCOLM E DENNET ACERCA DOS SONHOS)*

### **1. Os sonhos e os problemas da epistemologia da consciência**

Não vou falar directamente da dor e do sofrimento<sup>96</sup> mas de alguma coisa que eles exemplificam: os problemas da epistemologia da consciência. Mais especificamente, vou tratar as asserções de experiência interior e o sentimento de autoridade introspectiva que as acompanha<sup>97</sup>. Para abordar a questão escolhi um exemplo polémico para os filósofos da psicologia (o exemplo dos sonhos), e três filósofos (Wittgenstein, Malcolm e Dennett).

O problema que se põe quanto aos sonhos é saber se os sonhos são experiências. Numa primeira definição, aceite quase generalizadamente, os sonhos são experiências que ocorrem durante o sono e que são posteriormente recordadas e relatadas em vigília<sup>98</sup>. O problema da definição é que ela simplesmente pressupõe, sem justificação, a possibilidade de experiência num indivíduo de alguma forma inconsciente. Se não pressuusermos essa possibilidade, somos conduzidos a abordar a natureza dos sonhos a partir das relações entre apresentação, memória (nos sentidos de inscrição e recuperação) e expressão.

À primeira vista parece possível distinguir claramente as coisas para as quais é como alguma coisa ser e as coisas para as quais não é como nada ser (esta mesa, por exemplo). Só as primeiras têm consciência. O nosso problema (saber se os sonhos são experiências) traduz-se portanto assim: Será como alguma coisa sonhar? Podemos com segurança saber se

<sup>96</sup> Este texto corresponde a uma conferência apresentada no âmbito do Colóquio Para uma Antropologia da Dor e do Sofrimento, que teve lugar no âmbito do Projecto de Investigação com o mesmo título dirigido por Maria José Cantista (FLUP).

<sup>97</sup> Numa primeira caracterização, o sentimento de autoridade introspectiva marcaria a infalibilidade das crenças acerca da natureza e existência de estados de consciência próprios. Assim sendo, apenas as expressões que as pessoas produzem dos seus estados privados poderiam constituir evidência da existência e da natureza destes estados.

<sup>98</sup> Cf. por exemplo Hobson 1999a e Hobson 1999b.

o sonhador se sente ser durante o sono – se por exemplo sente dor ou sofrimento quando o sono é acompanhado de gemidos e gritos – como podemos garantir que existem durante o sono reacções fisiológicas que usualmente acompanham experiências de medo, ansiedade, ou excitação? Um problema idêntico coloca-se em relação a formas e graus de (in)consciência de humanos e outros animais como a anestesia, o coma e mesmo a morte – pelo mero facto de se necessitar de um ‘critério de morte’ está-se a admitir que é por vezes necessário decidir se um corpo que mantém reacções e funcionamentos ainda se sente ser.

Para saber se existe ou não existe experiência de algum conteúdo, seja no sonho ou em vigília<sup>99</sup>, em coma ou sob anestesia, o problema é saber a que evidência se pode justificadamente apelar. Ora, a questão dos sonhos interessa o filósofo da psicologia porque mostra a plausibilidade (mas também as consequências) de um certo verificacionismo na teoria da consciência, a que D. Dennett chama um operacionalismo<sup>100</sup> em primeira pessoa. Este é a ideia segundo a qual são impossíveis por princípio:

1. a consciência de um estímulo na ausência de crença nessa consciência
2. a distinção feita pelo sujeito, relativamente à sua consciência de *x*, entre aparência e realidade.

A pertinência do operacionalismo ( por vezes também chamado verificacionismo) em primeira pessoa quando se trata de experiência interior deve-se a duas razões:

1. o-em si do experienciar, a sua existência, só pode acedido como um para-mim.
2. só o próprio ser sensiente ou consciente pode saber como é ser e – pelo menos no caso humano – expressar como é para si ser. Daí, aliás, a sua autoridade.

Do ponto 2 surge no entanto um problema grave:

<sup>99</sup> No caso da vigília a questão põe-se por exemplo em relação a alguma coisa vista durante um muito curto intervalo de tempo (imagine-se: 40 mseg). A questão que se põe é saber o que é necessário para afirmar que o sujeito teve consciência desse conteúdo.

<sup>100</sup> O operacionalismo é uma posição em filosofia da ciência de acordo com a qual os conceitos teóricos são definidos por meio de operações empíricas. O operacionalismo está próximo de um empirismo radical segundo o qual os conceitos teóricos devem ser pensados como construções lógicas a partir de experiências. Assim, do ponto de vista operacionalista, as proposições que seriam à primeira vista acerca de entidades teóricas (por exemplo partículas físicas) são de facto acerca de experiências e operações determinadas. No caso de Dennett, a aplicação do operacionalismo le portanto da ideia segundo a qual o discurso explicativo não é acerca de complicadas entidades teóricas de facto in-experimentáveis directamente mas acerca de experiências ao problema da natureza dos estados conscientes liga-se à consideração segundo a qual aquilo que não deixa marca na ‘experiência da consciência’, não existe como conceito na/da consciência ou conceito de estado consciente. A significação de qualquer frase acerca da consciência é equacionada com a evidência disponível, e a evidência só está disponível nas circunstâncias mencionadas. Note-se desde já que perante esta proposta de ‘ligação conceptual’ da consciência ao rastro (e portanto à memória) é possível pelo menos possível objectar, como faz por exemplo N. Block, que existem coisas, perfeitamente reais, que aparecem e desaparecem sem rastro. (Block 1997: 177).

1. Para que alguém conheça e exprima a sua experiência interior, tem que haver inscrição e recuperação de informação, i.e. tem que haver memória<sup>101</sup>. Ora, seja o que fôr recordar, não é verdadeiro por definição que o recordado tem que ter acontecido tal como é recordado. A natureza da memória não é tal que as memórias devam ser verídicas. No entanto o sujeito ele próprio não tem qualquer possibilidade de distinguir entre as suas memórias verídicas e as não verídicas.

A vulnerabilidade dos sonhos a este problema reside no facto de eles apenas serem acedidos - já que não dão lugar a expressão em acto - enquanto memórias de experiências<sup>102</sup>. Como diria Malcolm, o uso da frase 'Estou profundamente adormecido e a sonhar<sup>103</sup>' como asserção provocaria um sentimento de estranheza, e seria de algum modo auto-contraditório.

Mas se os sonhos só são acedidos mediante recordação subsequente, impõe-se saber se o facto de alguém recordar que sonhou basta ou não basta para concluir que sonhou (que experienciou, portanto). Norman Malcolm, um discípulo de Wittgenstein, e um verificacionista radical, defendeu que não basta. Foi mesmo levado a concluir em *Dreaming* (1959) — uma obra que tinha como objectivo uma análise conceptual wittgensteiniana do conceito 'sonhar' — que os sonhos não são experiências. Malcolm concluiu ainda que as investigações científicas dos sonhos — que visavam estabelecer a relação do sono REM<sup>104</sup> (um estágio do sono assinalado por movimentos oculares rápidos) com o sonho e confirmar assim a concepção dos sonhos como experiências que ocorrem durante o sono<sup>105</sup> - incorriam em violações e pecados conceptuais e em última análise não diziam sequer respeito aos sonhos.

Em função de investigações sobre o sono e o sonho afirma-se hoje que todas as pessoas sonham, todas as noites. Apesar disso, algumas pessoas afirmam convictamente não sonhar, outras sabem que sonharam mas não se recordam dos detalhes, e todas elas, mais ou menos rapidamente, perdem as recordações dos detalhes. Como é fácil imaginar, o que é problemático aqui é a possibilidade de alguém, a partir de critérios externos,

<sup>101</sup> Cf. Roedinger & Goff 1998 e Allen & Reber 1998.

<sup>102</sup> As memórias de experiências podem, como memórias, não ser verídicas, i.e., serem memórias-de-experiências sem que tais experiências tenham existido.

<sup>103</sup> Cf. Malcolm 1956:18: «Suppose that I am in bed and you come and shake me and ask "Are you asleep?" and that I reply "I am sound asleep". It would be amusing if you took me as claiming that I am sound asleep and then concluded from this that I am sound asleep»

<sup>104</sup> O sono REM é um estágio do sono em que existem activação cerebral das áreas responsáveis pelo processamento visual e dos sistemas motores responsáveis por movimentos oculares rápidos. Supostamente durante o sono REM também existem sensações e percepções vívidas, internamente geradas, pensamentos ilógicos e bizarros e movimentos comandados mas inibidos. O sono REM alterna com a vigília e o sono não REM de acordo com uma sequência recorrente.

<sup>105</sup> Para a perspectiva científica contemporânea sobre a questão cf. Hobson 1999a e 1999b.

como EEGs e movimentos oculares, garantir a outra pessoa que ela sonhou (experienciou, portanto), quando esta, uma vez interrogada, garante que tal não aconteceu. Ou então, numa hipótese futurista, a possibilidade de garantir a alguém que acorda e relata que sonhou um sonho A que esse sonho não foi sonhado mas inserido, e que o sonho experienciado, um sonho B, não está acessível. Isto para não falar em casos mais complicados de dissociação ou deslocamento ou virtualização da experiência de uma pessoa, como aqueles que são explorados em ficção em filmes como *Matrix* (Irmãos Wachovsky) ou *Existenz* (Cronenberg).

Afastar a possibilidade de contestação da autoridade subjectiva nestes casos é o que está no horizonte de Malcolm quando declara que as investigações do sono REM não dizem respeito aos sonhos. Malcolm não diz que as pessoas não sonham (o 'critério' dos sonhos é a existência de relatos de sonhos) mas que os sonhos não são experiências<sup>106</sup>.

A origem do veredicto de Malcolm está numa afirmação de Wittgenstein nas *Investigações Filosóficas* (IIª Parte, 11. 213) «A questão de saber se a memória do sonhador o engana quando ele relata um sonho depois de acordar não pode surgir a não ser que introduzamos um critério completamente novo para o relato concordar com o sonho, um critério que nos dá um conceito de verdade por oposição a sinceridade do reportar (*truthfulness*)»<sup>107</sup>. Wittgenstein está a sublinhar que nós não temos um espaço interior de avaliação que nos dê a capacidade de distinguir entre ter realmente sonhado e recordar (enganosamente) que sonhámos.

A partir daqui, Malcolm defende que não faz sentido falar de uma verdade acerca da experiência, obtida por verificações exteriores que contradiga o testemunho do sujeito: apenas as expressões que as pessoas produzem dos seus estados privados são critério da existência e natureza destes estados.

Para salvar a autoridade introspectiva Malcolm terá no entanto que negar que os sonhos sejam experiências. O argumento é o seguinte: podemos *parecer enganosamente* que tivemos uma experiência sem a termos tido, uma vez que temos maneiras de *confirmar ou infirmar* as asserções de experiência independentemente da inclinação subjectiva à afirmação. Mas se nos parece que tivemos um sonho, tivemos um sonho. Ter tido realmente

<sup>106</sup> Como se verá, os sonhos não são experiências porque não 'cumprem os critérios' para o serem (os critérios em causa são critérios de sentido).

<sup>107</sup> A dificuldade ou impossibilidade de comparar a verdade (truth) com a sinceridade (truthfulness) do reportado acerca da experiência vale para toda a experiência. Poderíamos parafrasear Wittgenstein e afirmar que 'a questão de saber se a memória do experienciador o engana quando ele relata a experiência não pode surgir a não ser que introduzamos um critério completamente novo para o relato concordar com a experiência, um critério que nos dá um conceito de verdade por oposição a sinceridade de expressão'. Isto envolveria no entanto a abdicção da distinção [que é para Malcolm uma distinção 'conceptual'] entre 'sonhos' e 'experiências'.

um sonho e estar sob a impressão de ter tido um sonho é de facto a mesma coisa. O único critério do uso do conceito 'sonhar' é o testemunho de uma pessoa acordada, cujas asserções de experiência interior não podem, a partir de fora, ser confirmadas ou infirmadas.

O argumento de Malcolm, depende de uma epistemologia (muito contestável) que conjuga o verificacionismo com uma teoria da significação que tem como referência a linguagem comum. Assim, ele pensa que (1) existem critérios para conceitos; (2) os conceitos têm que ter métodos de verificação ou não têm sentido, (3) existem casos paradigmáticos que cumprem esses critérios e nos quais podemos decidir claramente que os conceitos se aplicam. Daí as conclusões quanto aos conceitos 'sonhar' e 'experiência'. No entanto estas conclusões só se justificam se:

1. a análise apriorista for um método defensável para a abordagem do significado de termos mentais (e da linguagem em geral)
2. a competência semântica das pessoas (i.e. a forma como elas usam conceitos) pressupuser de facto aquilo a que Putnam (Putnam 1962), analisando Malcolm, chama uma 'gramática de profundidade'. Esta gramática, não apercebida pelo falante comum, revelaria a dependência lógica de conceitos em relação a critérios (revelaria por exemplo a dependência lógica do conceito 'sonhar' em relação ao testemunho de uma pessoa acordada), tendo como referência as fronteiras do sentido estabelecidas pelo uso comum (a alternativa seria uma linguagem privada, cuja possibilidade Malcolm, como wittgensteiniano, obviamente nega).

É devido a esta epistemologia que Malcolm defende que é uma extensão sem sentido do conceito 'sonhar' aquela que toma um critério fisiológico (por exemplo os movimento oculares rápidos) para o conceito. O critério fisiológico envolve não apenas uma mudança conceptual radical como o uso de um conceito sem métodos de verificação.

Malcolm salva, com a sua conclusão, o sentimento de autoridade introspectiva quanto ao que se passa na mente. A salvação do sentimento de autoridade dá-se no entanto à custa de um conservadorismo epistemológico inaceitável. Parece absurdo defender, como faz Malcolm, que as investigações do sono REM nem sequer dizem respeito aos sonhos. Embora o em-si da experiência seja um para-mim, isso não nos obriga a afirmar (como assume o filósofo da linguagem erigido em guardião dos conceitos comuns) que o conhecimento dos fenómenos mentais só pode ser apriorista. Se o auto-acesso não for um acesso epistémico a tipos naturais,



parece pelo contrário desejável que os conceitos mentais tenham caracterizações técnicas, resultantes de investigação empírica. É claro que o problema com este movimento é que ele coloca fora da pessoa a autoridade acerca da natureza da sua própria experiência, produzindo uma situação estranha. Se as considerações objectivas não são pertinentes para decidir acerca da presença de experiência, já que o em-si desta é um para-mim<sup>108</sup> e se ao mesmo tempo, muitas vezes as convicções do sujeito não são suficientes para determinar a existência ou natureza da experiência (o que é o caso do sonho, se não se estiver disposto a decretar por critérios linguísticos aprioristas que os sonhos decididamente não são experiências), arriscamo-nos a ter que admitir fenómenos de subjectividade sem sujeito, *qualia* não apercebidos, fenomenologias sem amarras<sup>109</sup>. É claro, aliás, que é aqui que reside a pertinência do verificacionismo em primeira pessoa para aqueles que o defendem.

Acontece que a grande diferença entre filósofos da linguagem como Wittgenstein e Malcolm e os teóricos da mente contemporâneos é o facto de estes acreditarem que a análise conceptual é um mau método para a teoria da mente. A análise conceptual é um mau método precisamente porque impede arbitrariamente a entrada dos dados empíricos nos conceitos mentais. H. Putnam apontara já (Putnam 1962) que o principal erro da análise de Malcolm era não prever que os 'critérios da significação', inclusive da significação de termos mentais, são modificados ao longo do tempo, através da investigação empírica e dos novos conhecimentos<sup>110</sup>. É claro que se se permitir a entrada de dados empíricos no sentido dos termos mentais, o sentimento de autoridade acerca da natureza da sua própria experiência do sujeito para quem é como alguma coisa ser não poderá ser explicado como um acesso epistémico. Terá que ser justificado de outra maneira, nomeadamente como incorrigibilidade.

Dennett procura explicar esta incorrigibilidade através da relação entre memória, apresentação e expressão. Vou concretizar a ideia através da sua análise dos sonhos em *Are Dreams Experiences?*(1978). Esta análise insere-se na tentativa de elaboração de uma teoria fisicalista da consciência e o interesse de Dennett pelos sonhos deve-se a dois factores:

1. os sonhos permitem analisar uma certa substituição da apresentação por memória (que Dennett quer alargar a toda a teoria da experiência);

<sup>108</sup> Utilizando a formulação de J. Searle (Searle 1998) poder-se-ia dizer que a fenomenologia não é redutível à neurofisiologia devido à sua subjectividade ontológica.

<sup>109</sup> Dennett 1978: 143.

<sup>110</sup> Aliás um dos fundamentais propósitos de Putnam 1962 é defender que a competência semântica das pessoas prevê ou inclui competências 'indutivas', o que se traduz na projectibilidade dos predicados.

2. os sonhos permitem clarificar a distinção entre consciência e outras formas de apercebimento (*awareness*) com as quais a consciência é por vezes confundida.

Dennett considera nos sonhos apresentação, memória e composição. O ponto que pretende explorar é o seguinte: mesmo que se admita que acontece uma apresentação (entenda-se: consciente) das imagens e narrativas do sonho, correspondente aos eventos neuronais durante o sono REM, é a memória, no sentido de inscrição e armazenamento de informação, que possibilita o relato do sonho. Ora a inscrição e o armazenamento podem por princípio ser inconscientes e são independentes da (hipotética) apresentação. Assim, parece possível interferir na recordação sem se ter interferido na “apresentação”. Interferindo no armazenamento podem por exemplo ser introduzidas memórias de sonhos não sonhados, que serão recordadas como sendo de sonhos sonhados. A recuperação para relato dá acesso a conteúdos mas não a garantia de que uma apresentação aconteceu e nem isso pode ser avaliado pelo sujeito. Foi isso que Malcolm percebeu quando defendeu que de ‘Parece-me que foi como alguma coisa sonhar’ não é permitido concluir que ‘Foi como alguma coisa sonhar’.

Esta condição (segundo a qual a recuperação para relato dá acesso a conteúdos mas não a garantia de que uma apresentação aconteceu) aplica-se aliás à experiência em geral. Analisemos por exemplo uma recordação, reportada por uma pessoa A, das sua experiências relativas à entrada numa sala cheia de gente. Pergunta-se: Quantas pessoas estavam na sala? A pessoa B estava presente?

No relato da pessoa A, como em qualquer relato de experiência, estarão presentes conclusões tiradas com base em informação armazenada. Mas esta não tem que ter sido apresentada, apenas que tem que ter sido inscrita e que estar acessível. Se por exemplo a pessoa A afirma que não se recorda da presença da pessoa B na sala cheia de gente onde entrou, não se pode daí concluir que ela se apercebeu realmente de uma ausência, a ausência de B. Como diria Wittgenstein, a ausência de percepção não é percepção de uma ausência. A ausência de B não foi experienciada mas apenas ‘concluída’ por implicação a partir da informação disponível.

Quanto ao terceiro processo envolvido nos sonhos, a composição das frequentemente elaboradas narrativas e imagens do sonho, embora se considere usualmente que ela acontece durante o sono, nos períodos do sono REM, poder-se-ia descobrir que não é assim. Pense-se nos casos em que alguém sonha algo que encaixa perfeitamente no que está a acontecer quando se acorda (dá-se um grande estrondo exterior, por exemplo, e o sonhador acorda de um sonho em que sonha que morreu numa explosão). Parece haver uma preparação narrativa do encaixe com a realidade.

de exterior. Excluindo a existência de pré-cognição, esse encaixe poderia explicar-se, sugere Dennett, por hipóteses semelhantes a estas: talvez os sonhos sejam compostos e apresentados muito rapidamente no intervalo entre acordar e estar plenamente desperto, sendo o efeito conseguido com um sistema de atraso da percepção exterior. Ou talvez, nesse intervalo, os sonhos sejam compostos e armazenados em ordem invertida e depois recordados na sequência própria. Ou talvez exista uma biblioteca de sonhos não sonhados no cérebro, com vários fins indexados, prontos a serem inscritos na memória de forma adequada à percepção exterior, para virem a ser relatados. O que importa é que para o sonhador, uma vez acordado e relatando o que lhe parece ser, será impossível por princípio perceber qualquer diferença entre memórias inseridas de sonhos que não foram seus e sonhos experienciados.

Dennett não pretende infirmar com um exemplo avulso e com hipóteses *ad hoc* a ideia segundo a qual os sonhos são experiências. O seu propósito é apenas analisar as implicações que teria a confirmação de uma hipótese empírica semelhante a estas. Tal confirmação atestaria que o sentimento de autoridade subjectiva do sonhador não lhe dá qualquer autoridade para decidir entre hipóteses. Pelo contrário, outro tipo de dados (nomeadamente neurofisiológicos) seriam mais relevantes para a decisão. E note-se que de acordo com pelo menos a última hipótese sonhar não teria sido como nada, embora viesse a parecer ter sido como alguma coisa.

Dennett defende por isso que a questão ‘são os sonhos experiências?’ é uma questão teórica em aberto, i.e. os sonhos poderão ser ou não ser experiências mas isso não será decidível por apelo à autoridade subjectiva. O caso dos sonhos mostra que apesar do sentimento de autoridade subjectiva que as acompanha, o conteúdo das asserções de experiência interior é vulnerável à infirmação empírica. Dennett estende a conclusão a toda a fenomenologia (nomeadamente às descrições de imagens mentais, decisões voluntárias, experiência de cores, dores, etc, pelo sujeito). O sentimento de autoridade subjectiva não assinala qualquer conhecimento privilegiado da experiência mental.

## **2. O sentimento de autoridade subjectiva e os critérios de experiência**

Trata-se então de explicar o que é o sentimento de autoridade subjectiva, uma vez que de facto e em geral nos humanos a consciência se liga ao sentimento de autoridade subjectiva. Este sentimento não é no entanto

uma marca confiável do apercebimento num sujeito: para Dennett o conteúdo do apercebimento é tudo o que entra na memória<sup>111</sup> e os indícios deste apercebimento não estão necessariamente ou exclusivamente ligados à convicção subjectiva. A haver uma ‘relação conceptual importante’ como queria Malcolm esta dar-se-ia entre experiência e memória e não entre experiência e convicção subjectiva.

O caso dos sonhos mostra que pode existir memória de algo apercebido sem sequer ter havido apresentação: basta ter havido inscrição na memória de um conteúdo que (1) guie o comportamento actual. Este é um primeiro indício do apercebimento de x num sistema.

Por outro lado, a informação que guia o comportamento actual poderá ser objecto de acesso público, i.e., estar disponível para ser expressa. Um segundo indício do apercebimento será então: (2) poder ser expresso. Mas quando alguém dá por si querendo dizer, querendo emitir uma asserção, por exemplo uma asserção de experiência interior, isso acontece via a memória e o ‘querer-dizer’, sem qualquer espaço de observação interno. A incorrigibilidade que caracteriza estas asserções é função dos acessos sub-pessoais inconscientes entre a memória e o sistemas de produção da fala: os relatos de experiência interior são os próprios dados e não relatos de dados. Precisamente os sonhos levam Dennett a alargar este segundo indício a um terceiro: a capacidade retrospectiva de relatar (3). O quarto indício serão as (4) saliências funcionais que a investigação empírica descobrir (no nosso exemplo, a caracterização do sono REM). Estes indícios podem inclusivamente chocar entre si.

Assim, de acordo com esta teoria fiscalista e verificacionista<sup>112</sup> da consciência, a consciência não esgota o apercebimento. No entanto, os juízos esgotam a consciência imediata. Não pode haver consciência de x sem crença do sujeito na consciência de x, e o sujeito será incorrigível nessa

<sup>111</sup> *A memória é uma condição necessária da experiência, concedendo-se que as memórias podem não durar muito (cf. Dennett 1978: 144).*

<sup>112</sup> *Convem sublinhar que para Dennett sem o verificacionismo o teórico da consciência seria levado a admitir a possibilidade de zombies, de espectros invertidos, e a verdade do epifenomenalismo. Dennett considera todas estas ideias absurdas (cf. Dennett 1991: 461). Daí que afirme que o ‘superficialismo’ do verificacionismo não é uma resposta superficial aos problemas da teoria da consciência. ‘Consciência’ é apercebimento de que [factos], e algo que tem resultados [nomeadamente ser expresso]. Mas de facto o verificacionismo e a correlativa definição de consciência como um fenómeno de segundo grau ocupam, justificada ou injustificadamente, o lugar da consciência fenomenal. Por exemplo F. Dretske [Dretske 1994] contrapõe a esta restrição da consciência ao apercebimento de factos e à expressão aquilo a que chama ‘percepção não epistémica’, que não requer conhecimento ou crença. Para Dennett um fenómeno dessa natureza não existe ou pelo menos não se trata de ‘consciência’ mas sim de outro tipo de apercebimento. Dretske resume o defeito da teoria da consciência de Dennett da seguinte maneira: ela faz com que Dennett confunda ‘aperceber maçãs’ com ‘aperceber-se de que existem maçãs’. É por esta razão que Dretske classifica Dennett como cognitivista, i.e. alguém para quem a percepção é cognição, e ver ou ouvir ou sentir alguma coisa são espécies de juízos. Dennett torna mais confusa a questão ‘É (ou não é) a experiência constituída por entidades da natureza do pensamento?’ (thought-like entities) quando a mistura com a questão os qualia, que Dennett define como propriedades de segunda ordem das propriedades da experiência consciente.*

crença. Mas ele é incorrigível porque não tem ‘espaço de manobra’ e não devido a qualquer autoridade epistémica<sup>113</sup>.

Uma particularidade da teoria de Dennett é defender que no limite, em muito do apercebimento que acontece em sistemas como os humanos, a pura apresentação desaparece e no seu lugar fica a memória, concebida como registo físico; em virtude desta substituição de apresentação por memória ter-se-á que somos muitos mais ‘apagados’<sup>114</sup> do que aquilo que nos parece ser o caso.

Repita-se que neste quadro o apercebimento tem um âmbito muito maior do que a consciência, um âmbito a ser determinado pela investigação empírica. A autoridade subjectiva marca um certo tipo de acesso, e não a presença de um tipo natural, a ‘consciência’. Os tipos naturais são da competência da investigação empírica, e nada garante a priori que a consciência seja um.

Aprendendo com o absurdo da posição de Malcolm, que pretendia decidir a partir do uso linguístico presente se os sonhos são experiências, conclui-se que a noção de critérios de experiência é epistemologicamente ilegítima, derivada de uma insustentável posição apriorista perante questões empíricas abertas.

### 3. Conclusão.

Suponho que não é necessário sublinhar que as questões epistemológicas em torno da consciência são importantes para a ética, nomeadamente para a ética aplicada.

Um modelo como o de Dennett conduz-nos a concluir que não podemos dispôr de critérios aprioristas que garantam a existência de experiência num sistema. Assim, não podemos desde já distinguir claramente as coisas para as quais é como alguma coisa ser das outras. Não podemos fazê-lo porque por um lado é o próprio mundo que é vago, sem divisão abrupta entre consciência e apercebimento inconsciente, por outro porque se trata de questão empíricas abertas. Isto significa que nas ocasiões em que se deve considerar se há aí ou não alguém para quem é como alguma coisa ser se é obrigado a decisões.

<sup>113</sup> *De facto não existe um espaço interior para a apreciação do aparecer da vida interior. É por isso também não existe espaço para um verificacionismo no sentido próprio, como Wittgenstein viu. Seria interessante comparar o recuo wittgensteiniano perante o verificacionismo (por não ser possível cometer erros em asserções de experiência interior, que são *Äusserungen* e não asserções empíricas) e o verificacionismo que Dennett mantém.*

<sup>114</sup> *Cf. Pinto 1999: 133. O termo ‘apagado’ traduz o termo *zombie* e nomeia uma réplica física de um sistema consciente na qual a experiência fenomenal está ausente. A pertinência de qualquer destes termos é a captura da situação em que tudo decorreria no escuro, enquanto Dennett pensa que quando se trata de experiência nunca se está perante situações assim tão claras.*

Aliás, já é isso mesmo que as práticas mostram. Por exemplo, desde há algumas décadas que se admite como critério de morte a morte cerebral. Várias leis declaram que só existe morte cerebral quando se dá a cessação de todas as funções do cérebro (uma justificação seria o facto de o cérebro ser responsável pelo funcionamento do corpo como totalidade<sup>115</sup>). No entanto, tratamentos médicos são descontinuados e colheitas de órgãos são feitas em situações em que o funcionamento integrado do organismo ainda persiste, artificialmente mantido. O que se passa é que é implicitamente admitido que existem certas funções do cérebro que importam mais, nomeadamente a consciência ligada às actividades cognitivas superiores. Este é o factor que rege a decisão quanto à existência de alguém capaz de sofrer<sup>116</sup>. É portanto perfeitamente possível estar perante um corpo humano vivo, que mantém um funcionamento integrado (e por exemplo reacções de ‘dor’), e decidir que se trata de morte, porque o que importa realmente é a perda irreversível da consciência (logo, da possibilidade de sofrimento). Aquilo que normalmente se oculta é que semelhantes juízos sobre qualidade de vida supõem decisões e não descobertas, e se opõem àquilo a que os filósofos chamam a determinação absoluta de pessoas<sup>117</sup>. Eles são incongruentes com a ideia de que a existência de uma pessoa é uma questão necessariamente determinada no mundo. É uma decisão ética que lhes subjaz e não a descoberta de um abismo natural entre as coisas para as quais é como alguma coisa ser e as outras.

### Referências

Allen, R. & Reber, A., 1998, “Unconscious Intelligence”, in Bechtel, W. & Graham G., *A Companion to Cognitive Science*, Oxford, Blackwell, pp. 314-323.

Block, N., Flanagan, O., Guzeldere, G, 1997, *The Nature of Experience*, Cambridge MA, MIT Press.

<sup>115</sup> Critério usado por exemplo pelos filósofos católicos G. Grisez e J. Boyle nas suas análises de bio-ética. Cf. Grisez & Boyle 1979. A centração no ‘funcionamento integrado’ e não na actividade deve-se ao facto de reconhecer, a actividade poder continuar depois de as ‘funções’ terem cessado.

<sup>116</sup> Dennett defende a utilidade, neste ponto, de uma distinção entre a dor e o sofrimento, fundamentada na ligação entre a consciência e o importar [cf. Dennett 1996: 162]. Sem um sujeito persistente não pode haver alguém para quem alguma coisa importe, e sem consciência não pode haver esse sujeito persistente. Por outro lado, pode haver estados sensíveis (de dor por exemplo) sem reportação um sujeito unificado. Cf. Miguens, *Qualia e Razões*, neste volume.

<sup>117</sup> Cf. Parfit 1984. A determinação das pessoas corresponde à possibilidade de uma resposta clara à questão ‘Existe aí ou não uma pessoa?’, em todas as circunstâncias. D. Parfit defende a impossibilidade dessa resposta, e portanto a ‘indeterminação’ das pessoas. Se a Identidade Pessoal [a Pessoa] não é um facto a mais além da continuidade psicológica não é legítimo pressupor a pessoa independentemente da efectuação dessa continuidade psicológica. Parfit liga a sua posição [a que chama reducionista] acerca da Identidade Pessoal à definição lockeana de Pessoas (no *Essay Concerning Human Understanding*, 2º Livro, Cap. XXVII) pela auto-consciência e pela memória de actos passados.

Block, Ned, 1997, Begging the Question Against Phenomenal Experience, in Block, N., Flanagan, O., Guzeldere, G, 1997, *The Nature of Experience*, Cambridge MA, MIT Press.

Dennett, Daniel, 1978, *Are Dreams Experiences?* in Dennett, Daniel, *Brainstorms*, Cambridge MA, MIT Press, 1981.

Dennett, Daniel & Kinsbourne, Marcel, Time and the Observer in Block, N, Flanagan, O & Guzeldere, G. 1997.

Dennett, Daniel, 1991, *Consciousness Explained*, New York, Little Brown and Company.

Dretske, Fred, 1994, Differences that make no difference, in *Philosophical Topics* 22,1-2.

Flanagan, Owen, 1998, "Consciousness", in Bechtel, W., & Graham, G., *A Companion to Cognitive Science*, Oxford, Blackwell, 176-185.

Flanagan, Owen, 1992, *Consciousness Reconsidered*, Cambridge MA, MIT Press.

Grisez, G. & Boyle, J, 1979, *Life and Death with Liberty and Justice*, Indiana, University of Notre Dame Press.

Hobson, J.A., 1999a, *Dreaming*, in Wilson, R. & Kueil, F., (eds.) *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge MA, MIT Press, 242-244.

Hobson, J.A., 1999b, *Sleep*, in Wilson, R. & Keil, F., (eds.) *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge MA, MIT Press, 772-775.

Malcolm, Norman, 1956, Dreaming and Skepticism, in *Philosophical Review* LXV.

Malcolm, Norman, 1959, *Dreaming*, London, Routledge and Kegan Paul.

Parfit, Derek, 1984, *Reasons and Persons*, Oxford, Clarendon Press.

Pinto, João Alberto, 2007, *Materialismo, Superveniência e Experiência*, Porto, Campo das Letras.

Putnam, Hilary, 1962, *Dreaming and the Depth Grammar* in Putnam, Hilary, 1975, *Philosophical Papers*, vol. 2, Cambridge, Cambridge University Press.

Roedinger, H. & Goff, L., 1998, "Memory", in Bechtel, W. & Graham, G., *A Companion to Cognitive Science*, Oxford, Blackwell, 250-264.

Searle, John, 1998, *A Redescoberta da Mente*, Lisboa, Instituto Piaget.

Singer, Peter, 1999, *Brain Death and the Sanctity of Life Ethic*, Conferência no Departamento de Filosofia da New York University.

Wittgenstein, Ludwig, 1987, *Investigações Filosóficas*, (trad. Manuel Lourenço), Lisboa, Fundação Calouste Gulbenkian.

## 1. Introdução: A interioridade natural.

Vou supor que para falar de dor e sofrimento é preciso ter uma teoria da interioridade natural<sup>118</sup>. Esta interioridade natural não é o interior físico de um sistema físico, mas não há razões para considerá-la extra-mundana.

A teoria incluirá, por um lado, uma explicação da existência de entidades físicas capazes de *qualia* (i.e., de estados qualitativos, subjectivos, privados, sentidos), e por outro lado, uma explicitação das relações da interioridade com razões. Estas razões não são apenas razões explicativas da existência de interioridade. O que está em causa é que uma certa sofisticação da interioridade envolve a ascensão de razões à consciência e a presença da representação de razões para si próprio num controlo de si.

A dor é, na literatura<sup>119</sup>, o *quale* prototípico de experiência subjectiva. É o caso exemplar de sensibilidade como sensiência, a partícula de mundo sentido por excelência. O facto de a dor ter a vantagem (teórica...) de não representar nada, de ser um *raw feeling*, um sentimento cru, por oposição à *aboutness* ou intencionalidade, o traço usualmente central na

\* Este texto corresponde a um seminário apresentado no âmbito do projecto *Para uma Antropologia da Dor e do Sofrimento*, coordenado por Maria José Cantista.

<sup>118</sup> Também se poderia dizer uma teoria do "mental". Os termos "interioridade", "mental" e "alguém" serão intersubstituíveis ao longo do texto.

<sup>119</sup> Alguns exemplos na literatura da filosofia analítica do uso da dor como estudo de caso para asserções de experiência interior e suas virtualidades epistémicas são: Wittgenstein, *Investigações Filosóficas*, 1953, em praticamente toda a obra, mas, cf., por exemplo, o §244 ["A expressão verbal da dor substitui o grito, não o descreve"] e o §257 ["Como seriam as pessoas que não exteriorizassem as suas dores?"]; Anscombe, *Intention*, 1957, onde se defende que os relatos de dor são casos em que não temos conhecimento do que dizemos mas meramente somos capazes de dizê-lo, cf. p.57; Quine, *Word and Object*, 1960, p.5 ["Ouch is a one word sentence, which a man may volunteer from time to time by way of laconic comment on the passing show. The correct occasions of its use are those attended by painful stimulation"]. Isto para não referir as discussões no âmbito da Teoria da Identidade, nas quais o destino do problema da mente parece jogar-se em torno da questão: "A dor é/não é idêntica a disparos das fibras-C?" [cf. Kim 1996 para uma descrição histórica específica]. Os artigos de H. Putnam, fundadores do funcionalismo retomam aliás e naturalmente o mesmo exemplo, [cf. *Minds and Machines* (1960), *The Nature of Mental States*, 1967]. Posteriormente, também S. Kripke, em *Naming and Necessity*, (1972) o retoma, na consideração da necessidade/contingência envolvidas no estabelecimento de identidades teóricas. O ponto de grande parte destas discussões é a avaliação da nossa autoridade introspectiva sobre a natureza do objecto de introspecção, a comparação daquilo que somos capazes de dizer que se passa conosco, aquilo que nos sentimos ser, com o que é cientificamente relevante [por exemplo neurofisiologicamente] acerca de nós mesmos.



consideração da interioridade, faz com que ela abra a porta à teoria da interioridade como sensiência. E os problemas do mental como sensiência (*sentience*) são diferentes dos problemas da intencionalidade. Estes últimos dizem directamente respeito à mente como ‘representação do exterior’, enquanto os primeiros envolverão uma representação ou percepção daquele que representa, e por isso estarão ligados ao teor da interioridade.

Só há mundo sentido, importar (*matter*) se há *mind*, mentes ou proto-mentes. E a génese e complexificação d’ “aquilo que importa” (vou a partir de agora utilizar a expressão de Derek Parfit, “*what matters*”<sup>120</sup>) é paralela à génese e complexificação dos tipos de mentes. Sem um espaço para aquilo que importa, o que seria dor? Dor inconsciente? Dor que não é de ninguém, que não importa a ninguém? Os contornos da questão podem ser evidenciados por uma constatação como esta de D. Dennett<sup>121</sup>: enquanto dormimos, o processamento dos estímulos vindos dos nossos nociceptores impedem-nos de colocar os nossos membros em posições prejudiciais para as articulações. Mas será que esses processos são dores? Se respondemos que sim, teremos que afirmar que existem dores inconscientes, que não supõem um sujeito. Para Dennett, esses estados de protecção de um corpo são certamente estados de uma certa sensiência sem que possam por isso ser considerados como experiências de um eu. De alguma maneira eles não ‘importam’ (por exemplo, não têm qualquer significado moral), pois para que tais estados importem e sejam por exemplo fonte de sofrimento tem que haver um centro-sujeito que perdure e a quem importem. Há todo um projecto nesta observação: é implícito que criaturas naturalmente dissociadas, incapazes da centralidade que caracteriza a interioridade de tipo humano, seriam, embora sensientes, de certa forma incapazes de sofrimento, possuindo embora circuitos nervosos para a dor. Embora Dennett defenda a inexistência de fronteiras nítidas entre a sensiência e o que virá ser (por exemplo em nós) a consciência, poderia dar-se mesmo assim o caso de dor e sofrimento não serem o mesmo fenómeno em diferentes graus de intensidade, mas algo cuja existência só poderia ser determinada em função dos tipos de mentes e das suas características. Outro exemplo de Dennett: alguém teria ‘sofrido uma dor’ durante

<sup>120</sup> Cf. Parfit 1984. Este é o fio condutor do artigo e a razão para a unificação das suas partes.

<sup>121</sup> Em Dennett 1996, p.162. O uso da dor como exemplar na análise do mental vem de longe na obra de Dennett, desde Dennett 1969, Cap 11, a, por exemplo, *Why You Can't Make a Computer That Feels Pain*, in Dennett 1978, pp. 209-210. Este último artigo inicia-se remetendo a diferença na simulação de pensar e sentir para uma diferença entre programação e hardware, respectivamente (p.190). Dennett enuncia ainda as suas reservas quanto à possibilidade de simulação de dor, que se ligam à parvoqualidade do nosso pensamento da dor. Esta está, humanamente, ligada a uma carne ‘mole, frágil e quente’ e às lutas desta para sobreviver e evitar a morte (p.197). Um certo carácter parvoqual do conceito de dor não é necessariamente irracional e provavelmente exerce uma função na definição de uma comunidade moral. Apenas se torna objectável quando se traduz na ideia segundo a qual *qualia* apenas poderão ter como base física uma bioquímica semelhante à humana. Uma das conclusões de Dennett neste texto é que independentemente do seu carácter natural ou artificial a dor nos conduz ao problema moral da mente [se existisse um robot que sente dor artificial as pessoas cuidadosas fariam bem em não o pontapear, diz Dennett].

uma intervenção cirúrgica sob o efeito paralisante do *curare*<sup>122</sup>, estando portanto impossibilitado de exprimi-la corporalmente; posteriormente (mas não imediatamente a seguir) teria sofrido uma amnésia, cobrindo esse fragmento de experiência de dor. Neste caso, se ninguém recorda uma dor que ninguém exprimiu, terá existido de facto dor?<sup>123</sup> Perguntar se esta dor existiu conduz-nos a querer saber se foi de alguém, se foi sentida. Desejaríamos dizer que sim, mas a nossa ligeira dúvida prende-se ao facto de que acreditar que sentimos dor, exprimi-la e recordá-la ainda que brevemente, são componentes essenciais do sentir dor como experiência consciente<sup>124</sup>, pelo menos em nós.

Para que servem estes dois casos? O *esse est percipi* do mundo sentido exemplificado pela dor, parece supor alguém (i.e. alguma consciência parece ser uma condição necessária para a existência de dor). Mas que tipo de alguém e que tipo de consciência? Quão constituída, centralizada e continuada, tem que ser essa interioridade? O que é que está envolvido na passagem à sensiência a partir da sensibilidade como reacção local nem sequer necessariamente viva (há sensores artificiais, por exemplo)? Será toda a interioridade desde logo um centro?

Procurarei analisar algumas destas questões no ponto 2 deste artigo, no qual vou procurar ligar as características das interioridades naturais, mesmo que ainda não humanas, ao sentir, à sensiência e suas razões, fazendo apelo por um lado a algumas considerações de Dennett acerca da evolução da consciência<sup>125</sup> e por outro lado as sugestões de um psicólogo inglês, Nicholas Humphrey, que avança uma teoria acerca do teor afectivo-emocional da consciência fenomenal.

No entanto, pelo menos humanamente, a interioridade alargou-se da sensiência e do seu teor fisiológico primeiro, quase apenas reagente, não abertamente consciente, para domínios onde ‘aquilo que importa’ se dá a uma auto-consciência capaz de voluntaridade. Uma interioridade natural humana não surge *ex nihilo* e por isso deve ser considerada a partir da passagem de um *self* biológico sensiente, mesmo que minimal, para um eu humano.

Mas o aspecto ético da questão da interioridade não aparece imediatamente ao nível da sensiência (pense-se no primeiro exemplo que dei). Ele envolverá aquilo a que vou chamar o nível pessoal, as condições de pessoalidade

<sup>122</sup> Dennett vai buscar à história da medicina (aos anos 40 do século XX) o uso do *curare* como anestésico. Veio posteriormente a concluir-se que o suposto efeito anestésico era um efeito paralisante.

<sup>123</sup> Para Dennett a nossa intuição diz-nos que apesar de tudo houve dor [não gostaríamos que um suposto efeito de anestesia geral fosse assim conseguido em nós próprios]. Cf. Dennett 1978, p. 210. No entanto, segundo Dennett, alguns ‘enfiteçados’ wittgensteinianos ter-se-iam declarado dispostos a aceitar a substituição.

<sup>124</sup> Dennett pede-nos para imaginar a situação inversa: acreditamos e recordamos que sentimos, mas sem haver sentir.

<sup>125</sup> Cf. Dennett 1991, *The Evolution of Consciousness* e Dennett 1996, 3 [*The Body and Its Minds*] e 6 [*Our Minds and Other Minds*].

(*personhood*) e a Identidade Pessoal. Nos pontos 3 e 4, procurarei começar a analisar, exclusivamente a partir da filosofia, alguns problemas do discurso sobre a interioridade natural humana. Primeiro analisarei o que está em causa na relação entre o nível pessoal (onde há agentes conscientes, *qualia* e razões) e o nível sub-pessoal (onde se trata por exemplo de neurofisiologia, onde não há ‘pessoas’ nem ‘ações’). A distinção entre nível pessoal e nível sub-pessoal é de Dennett e é básica em teoria da mente. A sua pertinência reside no facto de que impedir a transgressão teórica entre estes níveis quando se descreve a vida mental é, segundo Dennett, e como evidenciaram Wittgenstein, nas *Investigações Filosóficas*, e Ryle, em *The Concept of Mind*, essencial.

Mesmo supondo a não descontinuidade entre sensiência e auto-consciência é preciso explicar a diferença entre elas. Noutras palavras, dada a possibilidade de existência de vários tipos de interioridades, é importante saber justificar quando é que há aí uma mente, uma interioridade que é além disso uma pessoa, e portanto uma certa maneira de existir daquilo que importa. É isso que está em causa na análise dos critérios da pessoalidade (*personhood*)<sup>126</sup>. A pessoalidade é uma sofisticação da interioridade natural<sup>127</sup> e é possível defender que ela é uma condição de possibilidade da existência de sofrimento por oposição a dor. Para Dennett, esta pessoalidade supõe que a auto-consciência depende de um certo modo de existir das razões no controlo de si de um sistema físico.

Se a pessoalidade é o lugar da ética, aquilo que está nela em jogo são os componentes da interioridade e ao explicitar estes componentes ter-se-á que decidir pela existência ou não existência de uma ‘Razão Pura Prática’. Num último momento do artigo procurarei mostrar o quanto a moralidade e a política (domínios em que está em causa uma distribuição ponderada daquilo que importa, mesmo das dores e dos sofrimentos, e obviamente também dos prazeres e utilidades) dependem de decisões quanto a um outro componente da interioridade humana, a identidade pessoal. Este vocabulário moral de ressonância utilitarista<sup>128</sup> prepara a en-

<sup>126</sup> De novo um termo de Dennett.

<sup>127</sup> A teoria de Humphrey quanto às sensações, de acordo com a qual estas são centrais na explicação da consciência, não se conjuga facilmente com a postura deflacionista de Dennett quanto aos *qualia*, ou melhor, opõe-se-lhe muito claramente em alguns pontos. A minha ideia aqui é portanto colocar em conjunto contribuições que não se sequenciarão sem problemas, antes pelo contrário evidenciarão dilemas e escolhas teóricas necessárias, numa única pré-história da interioridade natural.

<sup>128</sup> Utilidade é o termo mais geral para aquilo que importa a alguém. A formulação mais actual é a da satisfação das preferências e não a da maximização do equilíbrio prazeres/dores. A razão da entrada do utilitarismo no trajecto que estou a traçar é o facto de a perspectiva utilitarista encarar directamente “aquilo que importa” em relação com o futuro e as decisões, quanto a este, de agentes racionais. Como é sabido, a grande crítica ao utilitarismo é o facto de este perder de vista as pessoas como pessoas separadas, individuais, na sua avaliação da bondade/qualidade global dos estados de coisas desejáveis, negligenciando ainda a autonomia e afeições (*attachments*) das pessoas. Mas o utilitarismo é sobretudo uma filosofia da política pública (*public*

trada no texto, para a sua conclusão, do pensamento de Derek Parfit, um filósofo moral inglês contemporâneo, para quem razões e pessoas<sup>129</sup> são as questões centrais da filosofia moral. Para Parfit um ética substantiva só pode nascer do cruzamento de uma teoria da racionalidade com uma teoria da identidade pessoal, (a parte da teoria da interioridade que explicita aquilo que acreditamos ser), pois a maneira como o mundo moral pode ser moldado voluntariamente e racionalmente por agentes (se é que estes são metafisicamente possíveis), depende absolutamente de uma teoria da identidade pessoal, por oposição à possível instantaneidade de uma auto-consciência, e por oposição a uma consciência que não seria própria, não existiria em entes individuados, se pudessemos imaginar tal coisa<sup>130</sup>.

O sentimento de agir é um componente essencial da interioridade humana, tanto quanto a sensiência e a auto-consciência<sup>131</sup>, e a questão da Identidade Pessoal pode enunciar-se dizendo que todo o agente tem que saber que coisa no mundo ele é para poder agir. Não conhecemos agentes de outro tipo, senão estes que são assim particulares. Saber o que no mundo se é significa ter que saber (descobrir) o que há aí que é controlado, e quanto. Mas este saber, além de poder ser perdido, fragmentado, mais ou menos possuído, além de poder oscilar desde a megalomania (do ‘controlo todo o universo’) até à desposseção absoluta (‘sou controlado, programaram-me’) (na esquizofrenia o sentimento do controlo de si seria caracterizável por uma oscilação entre estes dois extremos<sup>132</sup>), tem que ser especificado e esse é trabalho da teoria da identidade pessoal.

O problema da Identidade Pessoal ao longo do tempo existe porque esta tem que ser operada, não é dada (há situações em que seria um problema saber se eu ainda sou eu, se sou o mesmo, se sobrevivi). As investigações

*policy), e é, enquanto tal, uma teoria moral racional por excelência. Pode certamente ser irrealistamente racional, mas se não se considerar um ponto de vista semelhante àquele que o utilitarista procura, o ponto de vista da benevolência impessoal, imparcial, não se passará nunca de um micronível na consideração da qualidade moral de acções, nem aparecerão como questões para a filosofia moral racional problemas macroeconómicos como a fome (cf. Sen 1987, como guia para estas questões. Sen é no entanto um crítico do utilitarismo).*

<sup>129</sup> Cf. Parfit, Derek, *Reasons and Persons*, Oxford, 1984 e Dancy, Jonathan, *Reading Parfit*, 1997

<sup>130</sup> Cf. Gil 1998. “Aquele [a consciência de si] dá-se no instante, esta [a identidade pessoal] realiza-se pela permanência do eu através do tempo, através da duração [...] A permanência temporal é a maneira menos essencialista de considerar a unidade do eu”. Este perdurar, a crença nele, devem estar envolvidos, segundo F. Gil, na “confiança na acção”, pois “sem crer na duração não se vê como poderíamos ter confiança nos efeitos temporais das nossas acções. Se o futuro nos aparecesse como descontínuo, zenoniano, não estaríamos seguros de que eles continuariam a produzir-se após o primeiro instante da sua realização. Por outro lado, sem uma continuidade entre a acção e o seu efeito, que pressupõe também a duração, no caso de os efeitos desejados parecerem produzir-se, não saberia se eles são a consequência da minha acção”. A propósito da crença na duração assim envolvida na acção, F. Gil faz notar que “se a verdade do homem consistisse no ser-para-a-morte heideggeriano” a possibilidade de agir estaria comprometida [o que não parece ser o caso]. Nesse sentido, é “terrorismo intelectual pretender que não ser-para-a-morte signifique inautenticidade”.

<sup>131</sup> Cf. Gil 1998.

<sup>132</sup> Cf. Sass 1998, pp 543-65.

de Parfit conduzem-no por este caminho, e através dele a outra questão: saber se a Identidade Pessoal é sempre determinada.

Uma mudança na nossa concepção de Identidade Pessoal, tal como aquela que Parfit defende e a que chama a Visão Reduccionista, pode e deve afectar as nossas crenças acerca de racionalidade e moralidade na acção. Se deixamos de acreditar que a Identidade Pessoal é um facto a mais além da continuidade psicológica, deixamos de poder explicar a unidade da vida de uma pessoa dizendo que as experiências nessa vida são experiências tidas pela mesma pessoa<sup>133</sup> e estranhas consequências se seguem quanto àquilo que importa. Nomeadamente, a teoria clássica da racionalidade na acção (que é o egoísmo racional, segundo o qual o fim supremamente racional para cada pessoa é que a sua vida decorra tão bem quanto possível<sup>134</sup>), colapsa. Isto dá-se pois ele assume a separação das pessoas e tem como unidades os agentes ao longo de toda a sua vida, quando essas unidades e a sua determinação são precisamente aquilo que é problemático.

Se não somos unidades já determinadas, separadas e naturalmente continuadas, aquilo que devemos racionalmente decidir e desejar afigurar-se-á muito diferente. Isto importa quando consideramos questões tais como a felicidade e o sofrimento, o prazer e a dor que poderão vir a ser desejados ou evitados como experiências por alguém futuro. Esse alguém futuro posso ser eu mesmo, num estado muito futuro da minha vida com o qual a minha continuidade psicológica é fraca<sup>135</sup>, mas também, e é por isso que os problemas tratados por Parfit não são unicamente especulativos, podem ser pessoas futuras, afectadas por decisões presentes. É nesse sentido que teorias da racionalidade e da identidade pessoal são necessariamente pressupostas nas políticas públicas em geral (na economia, na política, no direito, etc<sup>136</sup>). Nestas áreas é obviamente decisivo saber quais são as unidades relevantes para saber que razões existem para agir quando se trata de distribuição, de compensação, de punição, de atribuição de culpa, de imparcialidade, de merecimento. Devemos considerar pessoas separadas como irreduzíveis ou compô-las, em tratamentos globais? Poderemos tratar as suas vidas inteiras como unidades se elas podem estar fracamente

<sup>133</sup> Para Parfit só podemos explicar essa unidade descrevendo relações entre experiências e um particular cérebro e as 'variações' deste.

<sup>134</sup> O egoísmo racional é *agent biased* e *temporally neutral*, de acordo com os parâmetros que Parfit utiliza para classificar as teorias da racionalidade na acção.

<sup>135</sup> Uma das (aparentes) falhas estruturais do egoísmo racional como teoria da acção é partir do princípio que o que importa é o agente ao longo de todo o tempo da sua vida (nos termos de Parfit, o egoísmo racional é *agent-biased* e *temporally-neutral*). Ora, nós não somos *temporally-neutral*, pelo contrário. Naturalmente damos muito pouca importância, por exemplo, ao nosso futuro muito distante. Então, ou essa inclinação favorável ao temporalmente próximo é irracional ou é a suposta neutralidade temporal que deve cair.

<sup>136</sup> Traduzo assim *public policy*.

ligadas<sup>137</sup> (tratar-se-á ainda da mesma pessoa)? Poderemos falar dos interesses de pessoas futuras, de modo a considerar por exemplo a questão da justiça não apenas como intra-geracional mas entre-geracional<sup>138</sup> (para chegar a acordo quanto a Princípios de Poupança Justa, por exemplo<sup>139</sup>)?

Há, segundo Parfit, inúmeras maneiras de estarmos enganados quanto ao que julgamos ser uma pessoa, e isto é obviamente importante quando tentamos evitar prejudicar outras pessoas ou fazer-lhes bem, pois podemos ser muito irracionais prosseguindo esses objectivos. Para Parfit, várias teorias da moralidade como racionalidade na acção que estão disponíveis são auto-refutantes<sup>140</sup>. E o ponto de Parfit ao congregar Razões e Pessoas como os temas da filosofia moral é o seguinte: embora haja muitíssimos cépticos morais não há assim tantos cépticos acerca da racionalidade. Portanto, argumentos quanto à racionalidade (de acções e de crenças) são a única possibilidade daquilo a que Parfit chama um progresso em ética. Poder-se-á avançar numa discussão objectiva da moralidade quando se fala de razões para agir em geral e não abruptamente de ‘razões morais’. O fim da história da interioridade que vai ser feita será então uma delimitação a partir de Parfit de alguns parâmetros para a ponderação racional daquilo que importa para além das pessoas particulares. Deste modo se chegará a uma impessoalidade daquilo que importa.

## 2. Primeiro problema para a interioridade natural: o sentir

“Frequentemente, quando alguém quer acentuar o facto de o assunto ter passado da (mera) neuroanatomia para a experiência, da (mera) psicofísica para a consciência, da (mera) informação para os *qualia*, o termo *enjoy* (gozar, usufruir) é apressadamente empurrado para o palco”

D. Dennett, *Consciousness Explained*<sup>141</sup>

A primeira coisa de que precisamos para começar a conceber a possibilidade de ‘alguma coisa que importa’ é de um espaço em primeira pessoa (seria melhor falar de uma proto primeira pessoa). Este poderia parecer ser imediatamente o problema da consciência – do *what is it like*

<sup>137</sup> Cf. Parfit 1984, p. 334: “Se a unidade [quando se trata de aplicar princípios distributivos] é a totalidade da vida de uma pessoa, como Rawls e muitos outros assumem, o Princípio da Igualdade dir-nos-á para tentar ajudar aquelas pessoas que estão pior [worst off]. Se a unidade é o estado de qualquer pessoa num particular tempo, um Princípio de Igualdade dir-nos-á para tentarmos tornar melhores não as vidas das pessoas que estão pior, mas os piores estados em que pessoas estão”.

<sup>138</sup> Kant pensava que a história era injusta entre gerações, que o destino diferente de diferentes gerações era injusto, cf a sua *Ideia de uma história Universal de um Ponto de vista Cosmopolita*.

<sup>139</sup> Cf. Rawls 1993, p.227. Kant, por exemplo, pensava que desde logo o destino [diferente] de diferentes gerações era injusto. Cf. *Ideia de uma História Universal de um Ponto de Vista Cosmopolita*, in Kant, *A Paz Perpétua e Outros Opúsculos*, Lisboa, Edições 70, 1995.

<sup>140</sup> Os três casos centrais em Parfit 1984 são o egoísmo racional (S), o consequencialismo (C) e a moralidade de senso comum (M).

<sup>141</sup> Dennett 1991, p.383.

to *be...*?, na célebre expressão de Nagel. Esta pergunta teria resposta por certas partes do mundo e por outras não. É objecto de acesa discussão na filosofia da mente saber se se pode ou não falar da consciência como uma questão de tudo ou nada, ou se é apenas legítimo falar de graus e de tipos de mentes, artificiais e naturais, desde a sensibilidade inconsciente ou quase inconsciente (como a de sensores artificiais ou da pele do nosso pescoço à luz do sol na praia) à consciência de tipo humano. Esta última posição continuista é defendida por Dennett, para quem, considerando por exemplo um humano, a consciência não é uma unidade nem um centro nem um *medium* para o qual seriam transduzidos os processamentos cognitivos inconscientes. O problema da consciência fenomenal separa as águas e os autores na teoria científica e filosófica da mente, mas não vou discutir aqui as razões envolvidas nesta cisão<sup>142</sup>. O que me importa aqui é o seguinte: quando se passa do tratamento da consciência como caixa negra, adoptado pelos fenomenólogos de vários tipos (no sentido em que se toma o ‘comportamento’ da consciência como dado, e se recolhem fenomenologias sem inquirições físicas) para o estudo dos mecanismos físicos e biológicos da consciência, o que é que faz sentido perguntar acerca da existência de consciência?

Um hipótese é que devemos pedir uma narrativa evolucionista que nos permita conceber as razões da passagem da não-consciência à sensibilidade, à sensiência e à consciência de tipo humano ou outro em organismos. E o que queria aqui acentuar é que os percursos nesta via de análise têm chamado a atenção de muitos teóricos, filósofos e psicólogos evolucionistas, para algo que se poderia formular assim: dada a ligação dos estado de consciência dos organismos a alarmes (filogeneticamente) antigos, aparentemente não poderia haver estados de consciência que fossem emocional e afectivamente neutros. A natureza não faria máquinas epistemicamente neutras, não faria máquinas epistémicas, *tout court*, máquinas de ‘pensar a partir de fora’, como se não fosse eu, pois não teria razões para isso. Como diz Dennett<sup>143</sup>, “criaturas biológicas não poderiam tornar-se subitamente recolhedoras desinteressadas de informação (pensantes-observadores) porque os repórteres nelas são os alarmes dos seus antepassados, e não podem mandar mensagens intactas, mas sempre positiva ou negativamente editadas”. Como Kathleen Akins, uma filósofa da neurociência, tem vindo a insistir acerca de várias propriedades sensoriais, como as cores, os sistemas sensoriais não são feitos para detectar categorias naturais mas sim para manter os organismos vivos; eles servem um propósito ‘narcisista’<sup>144</sup>.

<sup>142</sup> Para uma discussão completa, cf. Pinto 1999.

<sup>143</sup> Dennett 1991.

<sup>144</sup> Cf. as referências a Akins in Dennett 1991, páginas 179, 376, 382.

Assim sendo, a consciência, longe de poder ter tido *ab origine* qualquer finalidade epistémica, teria antes origem em episódios de interrupção por alarme do controlo de si não consciente pelo organismo (esse, assim como a sensibilidade, deve existir mal exista um organismo) e estaria portanto relacionada com o controlo de um si, supondo o apercebimento da distinção si/fora. A situação será então a seguinte segundo Dennett: nos organismos sensientes mais simples a consciência é simplesmente o facto de estar feito para (*wired*), tocar e recolher ou avançar em resposta ao que é ‘bom’ ou ‘mau’ no ambiente. Distinguir bom/mau é a razão da consciência. Pense-se nos exemplos que temos ouvidos mencionar por neurofisiólogos, como a persistência, sob anestesia ou mesmo em cadáveres<sup>145</sup>, de respostas reflexas a estímulos que seriam dolorosos –o que significa que o comportamento de dor existe pois os trajectos nervosos entre nociceptores e músculos estão funcionais, e o que falta é precisamente aquilo a que chamaríamos a consciência. Estes exemplos indiciam o problema das pequenas percepções inapercebidas não reportadas a centros de interioridade que podem ainda não existir (ou, no caso, já não existir). Este sentir não poderia ser necessariamente reportado a um centro.

Esta razão da consciência seria a base dos ‘traços mais terríveis e deliciosos’<sup>146</sup> da consciência: a consciência-que-virá-a-ser-consciência, começa em termos de evolução por ser nestes organismos uma ligação no sentido mais directo possível às fontes do bem e do mal; eles tocam-lhes. A ideia é que essa origem ‘contaminada’, essa ligação directa da proto-consciência às fontes do bem e do mal permanece de alguma maneira, na medida em que o ponto de vista do observador consciente humano é um descendente sofisticado dos pontos de vista primordiais dos primeiros replicadores, pelo mero facto de serem vivos, que por terem um interesse (replicar-se) tiveram razões para dividir o seu mundo em bom e mau. A consciência, enquanto física e biologicamente existente teria origem evolutiva numa distinção bom/mau para uma divisão dentro/fora: dir-se-ia que a ligação à individuação é a fonte do bem e do mal que será para sempre o teor da consciência em seres biológicos... eis por que as questões do prazer e dor remetem para ‘*playing around with one’s boundary*’<sup>147</sup>. A consciência, enquanto vigilância, interrupção, teria tido origem na necessidade de substituir, em ocasiões de alarme, um centro ao piloto automático do organismo.

No princípio não havia razões, só causas, costuma dizer Dennett<sup>148</sup>.

<sup>145</sup> Cf. Jorge Tavares, in M.J. Cantista ed, 2001. *Os cadáveres que reagem são considerados legalmente mortos segundo o critério da morte do tronco cerebral.*

<sup>146</sup> Dennett, 1991, p. 177-178.

<sup>147</sup> Dennett 1991, p. 413.

<sup>148</sup> Dennett 1991, p. 173.



Nada tem função enquanto nada que tem interesses: é com o vivo e as suas fronteiras ou bordos, i.e. com a distinção fora/dentro que nascem as razões, enquanto fins-para-o-dentro, enquanto dever evitar as coisas ‘más’ para si e procurar as coisas ‘boas’. Essas razões (para evitar, procurar) não são obviamente representadas explicitamente para o organismo, são-no *no* organismo. Dennett chama-lhes, *free floating rationales*, i.e. razões para um funcionamento, neste caso biológico, que existem mas não são apreciadas por nenhum pensamento. Interessa-nos notar que, quando nasce o mundo biológico do *design*, no sentido duplo de desígnio ou fim e organização, nasce o mundo das razões, e é neste mundo já ‘finalista’ que devemos procurar o proto-mental, e o proto *mind*ing. No princípio, dizem as narrativas evolucionistas de psicólogos e biólogos, todos os sinais causados por coisas no ambiente significavam ‘fugir daqui’ ou ‘lançar-se sobre isso’<sup>149</sup>: o proto-significar (i.e. ser de alguma maneira para o dentro) seria simplesmente ‘estar directamente ligado ao mecanismo biológico para fazer x’, uma destas acções. ‘Bom/mau’ seria portanto o primeiro aspecto da significação.

A sensiência e a capacidade de dar resposta estariam assim, é essa a hipótese, intimamente vinculadas, em termos de história biológica dos organismos. É aqui que eu quero introduzir alguns dados da teoria psicológica (bastante especulativa, mas que me parece reveladora nalguns pontos) acerca da origem da consciência, do psicólogo Nicholas Humphrey.

Esta ligação entre sensiência e capacidade de dar resposta é a pista para Humphrey. Ele que pensa que a questão ‘Quanto gosta (ou ‘desgosta’) um animal de ser estimulado?’ está intimamente ligada à questão ‘Como é para o animal, o organismo, responder ao estímulo?’ (este ‘como é?’ deve ser entendido no sentido do ‘*What is it like?*’ nageliano, no sentido de consciência, portanto).

Segundo Humphrey, estar consciente (por contraste com representar o que existe no exterior) é sentir; a consciência sensorial ter-se-ia desenvolvido a partir das respostas de aceitação/rejeição dos organismos. O caminho da investigação de Humphrey em *A History of the Mind* (1992) vai levá-lo através de zonas pouco trilhadas na psicologia teórica do último século. Segundo Humphrey, no que respeita a uma psicologia sensorial a situação hoje não é muito diferente daquela que Freud diagnosticou em 1905<sup>150</sup>, quando afirmou que ‘tudo o que se refere ao problema do prazer e da dor toca um dos pontos mais débeis da psicologia actual’.

O que aconteceu segundo Humphrey na história da psicologia foi o seguinte: o tratamento da percepção envolveu uma controvérsia acerca

<sup>149</sup> Dennett 1991, p.178.

<sup>150</sup> S. Freud, *Três Ensaios sobre a Teoria da Sexualidade*, citado e comentado em Humphrey 1992, p. 55.

da existência do tratamento serial ou paralelo do sentir e do representar o exterior. E o que esteve em jogo nessa controvérsia foi a possibilidade de admitir uma diferença entre o que Humphrey chama os modos de representar auto-cêntrico (sentir) e alo-cêntrico (perceber)<sup>151</sup>. O resultado (infeliz para Humphrey) foi que o interesse teórico pela especificidade do sentir caiu, rebatendo-se sobre o interesse pelo perceber. Humphrey quer ir atrás, em *A History of the Mind*, daquilo em que consiste sentir. No tempo de Freud ninguém dava atenção a isso, e, segundo Humphrey, no nosso tempo também não, a questão do representado na representação (o fora do organismo) sobrepôs-se à questão teórica acerca do que se passa agora aqui comigo. Mas o tratamento desta propriedade, desta eu-idade da representação não tem, segundo Humphrey, que envolver qualquer referência ao fora.

A diferença e possibilidade de disjunção entre sentir e perceber tem, segundo Humphrey, passado teoricamente despercebida mas seria central na questão da consciência. É como se, por comparação com um síndrome neuropsicológico que ocupou Humphrey na sua carreira (a visão cega), o estudo psicológico e neuropsicológico e filosófico da percepção enfermasse de visão cega (o caso em que, por lesão no cortex visual, pacientes vêem sem sentirem que vêem, i.e, sabem O QUE, sem sentirem QUE<sup>152</sup>). Como diz Humphrey, esta é a situação inversa daquela na qual pacientes agnósticos sentem, sem nada poderem pensar - como nós quando ouvimos uma língua desconhecida como sons sem sentido. No caso do pensamento aquilo a que Humphrey vai chamar “sentição” (*sentition*) estamos na primeira circunstância: pensamos na percepção como se ela fosse directamente uma questão de *aboutness*, intencionalidade, sem qualquer mediação sensorial. Aliás Dennett é um dos autores que não admite a especificidade teórica do sentir na teoria da consciência<sup>153</sup>. As especulações de Humphrey colocam-se contra a conhecida posição eliminativa de Dennett quanto aos *qualia*, e é exactamente por isso que me interessam. Dennett, com a sua teoria da consciência como máquina virtual (uma estrutura temporária, não capturável neurofisiologicamente porque funcional e ligada a auto-estimulação e auto-inspecção no controlo de um organismo), pretende eliminar

<sup>151</sup> Humphrey 1995, *La doble provincia de los sentidos*, p.50.

<sup>152</sup> Cf. Humphrey 1995, ‘Más acerca de la visión a cegas’, pp. 94-95 e seguintes.

<sup>153</sup> Para ser justa, a questão de Dennett com os *qualia* é antes acerca da relação entre conhecimento e propriedades conhecidas. Dennett não nega, antes subscreve, grande parte da história evolucionista da interioridade que Humphrey descreve. A grande crítica a Dennett será considerar que ele confunde “consciente” com “consciente de”, mesmo que o objecto do ‘de’ seja definido de forma muito abstracta. Contra Dennett, cf. Pinto, J.A 1999, *O difícil problema mente-mente*, p.118: “O que determina a existência de um problema mente-mente não é apenas a introspecção [...] mas que o introspectado seja fenomenal ou sentido”. O problema fica a ser o seguinte: pode existir auto-consciência sem *qualia*?

a ideia de *qualia*<sup>154</sup>, de incorribilidade no auto-acesso sentido, que é para muitos a marca do estar consciente. Fá-lo porque o primeiro ponto a considerar no pensamento do pensamento é que não há ninguém a ver: i.e. no cérebro não há outra interioridade senão o próprio funcionamento, e portanto não há lugar para uma apreciação a mais, que constituiria os *qualia*. Pelo contrário Humphrey pensa que estar consciente é exactamente sentir – pensa que há aí um conteúdo a especificar teoricamente através da ligação da eu-idade e agor-idade do sentir à actividade do organismo.

Para Humphrey, a função das sensações é, na economia mental, prover o organismo de representações do que se está a passar consigo, aqui e agora, dar uma eu<sup>155</sup>-idade e uma agor-idade, à experiência do mundo, característica da qual está privada a ‘percepção pura’ (ou aquilo para que ela serve: o que conta no perceber é aquilo de que o organismo é informado quanto ao exterior, não interessa de onde ou de quem, interessam por exemplo permanências exteriores de objectos).

Como já se percebeu, Humphrey defende a distinção teórica entre sentir e perceber. Os animais<sup>156</sup> teriam desenvolvido modos diferentes de representar aquilo que se passa na superfície do corpo: as sensações, representações carregadas de afecto do que se está a passar comigo agora, e as percepções, representações neutras quanto a afecto do que se está a passar fora. Há portanto uma dualidade entre sentir e perceber: embora usualmente indiscerníveis na nossa vida mental eles podem no entanto des-acoplar-se. Por exemplo, ‘vermelho’ é a sensação, mas também a cor percebida de uma pétala. Num sentido como a visão é muito difícil ver essa dualidade sensação-percepção (no olfacto por exemplo já não) Mas mesmo na visão Humphrey encontra vestígios do sentir enquanto desacoplável do perceber cuja função é representar o fora, nomeadamente na nossa reacção emotiva forte às cores<sup>157</sup>.

As sensações enquanto constitutivas da nossa vida mental são para Humphrey teoricamente especiais devido às seguintes razões: 1. devido à sua privacidade, 2. por estarem localizadas no espaço do corpo, 3. por serem modalmente específicas, 4. por serem no tempo presente. Elas auto-caracterizam-se aliás quanto a todos estes aspectos, e esse é o ponto central<sup>158</sup>.

Ora, esse tipo de características parece a Humphrey só poder aparecer em processos que têm muito em comum com actividades corporais.

<sup>154</sup> Cf. Dennett, para uma defesa desta posição, *Quining Qualia*, in Block, Flanagan & Guzeldere, *The Nature of Consciousness*, 1997, p.619.

<sup>155</sup> Seria melhor dizer *self-idade*.

<sup>156</sup> Humphrey fala directamente de animais pois a passagem entre proto-sentir e consciência que lhe interessa só se dá em animais com neo-córtex.

<sup>157</sup> E dos macacos. Humphrey relata experiências de curiosidade epistémica versus prazer sensorial.

<sup>158</sup> Cf. Humphrey 1995, *Cinco características em busca de uma teoria*, p.139.

As acções vulgarmente chamadas voluntárias (não no sentido filosófico mas neurofisiológico, i.e. de sistemas eferentes, efectuadores de controlo) são exactamente o único caso de reunião dessas características que Humphrey encontra. Humphrey especula então que a actividade a que chama sensação seria uma actividade que se dirige para fazer alguma coisa no local onde a sensação é sentida.

As sensações são portanto minhas, privadas, 'sei-as onde estão'. A questão é: como é que eu sei isso, sendo que no saber isso, eu ficarei a saber que um isto é o eu consciente? A resposta é: agindo. Eu fico a saber que sou aquilo que descubro controlar, mas não tem que haver aqui um eu explícito nem um saber explícito, apenas uma efectuação apercebida (os exemplos de Humphrey são estudos com gémeos siameses e com bebés). Sentir seria então para alguém emitir instruções que são requeridas para suscitar a actividade apropriada, e as actividades sensoriais actuais dos humanos ter-se-iam desenvolvido a partir dos inícios filogeneticamente primitivos das instruções para aceitar/rejeitar, até circuitos encurtados que se tornaram nos animais superiores (com neo-córtex) circuitos de *feed back* reverberantes no cérebro<sup>159</sup>. A actividade de sentir já não é então em nós como teria sido e é ainda em organismos mais primitivos – ir lá, à superfície do corpo, responder – mas 'ir ao modelo do corpo no cortex'. Sentir é essa actividade para Humphrey, e esta seria a essência do estar consciente. Esta teoria da resposta-actividade permite-lhe inclusive pensar as modalidades sensoriais à maneira de acções do organismo, explicando assim os estranhos abismos do sentir, i.e. a falta de continuidade entre as modalidades sensoriais<sup>160</sup> (por exemplo não conseguimos imaginar as passagens entre o amarelo e o agudo, entre o ácido e o verde). Em qualquer exposição neuroanatômica as modalidades sensoriais são simplesmente tomadas de forma bruta, mas Humphrey pergunta como podem ser elas ser descontínuas se todas consistem neurofisiologicamente em condução electroquímica. Dizer que ter sensações visuais ou sensações auditivas é ser recebido no cortex visual ou auditivo não é explicar nada, segundo Humphrey. É melhor conceber essa diferença à maneira de actividades de um corpo (à maneira das diferenças entre jogar, dançar, cantar). Por outro lado, a reverberação dos circuitos é para Humphrey o tempo subjectivo, o tempo da consciência, que não é o tempo físico<sup>161</sup>. Humphrey fundamenta essa afirmação com a possibilidade de contracção ou expansão do tempo subjectivo por exemplo com drogas, que aliás também servem para mostrar

<sup>159</sup> Cf. Humphrey 1995, *Tiempo presente*, p.195.

<sup>160</sup> Cf. Humphrey 1995, *Energias nerviosas específicas*, p.180.

<sup>161</sup> Cf. Humphrey 1995, pp. 199-207.

essa separabilidade entre sentir e perceber; por exemplo sob o efeito de drogas psicodélicas como o LSD os olhos deixam de se preocupar com as questões normais do ver, tais como: Onde? A que distância?.

Humphrey pensa portanto que o sentir como actividade/resposta é a chave da consciência, e que esta via é esquecida pelos filósofos quando tratam dos *qualia*. Aquilo que eu quis ir buscar à psicologia evolucionista e a Humphrey foi a hipótese de uma ligação do sentir à consciência através de uma eu-idade e agora-idade do mental que Humphrey especula ser actividade e caracteriza por oposição à ex-centração do perceber<sup>162</sup>.

Afirmei no início que a interioridade tem vários componentes. Ela não se resume portanto ao sentir, que pode ou não ser central na consciência. Pensando a partir da nossa própria interioridade, e citando Fernando Gil, é aliás surpreendente que “os conceitos do eu façam sistema”, já que os opostos de eu de nenhum modo se identificam<sup>163</sup>. A explicação da interioridade natural não deve portanto tratar apenas da oposição entre sentir e não-sentir, mas também das oposições entre (auto)consciência e não-consciência, entre unidade e cisão ou pluralidade, entre controlo e desposseção.

### 3. Segundo problema para a interioridade natural humana: os níveis pessoal e sub-pessoal.

Antes de abordar esta multiplicidade teórica do interior, importa notar que para muitos filósofos a dificuldade central do problema da interioridade mental, é saber como é que se relacionam, num humano, num corte presente, o nível pessoal (da interioridade já plenamente constituída - o mundo da consciência e do sentir, das razões e acções) e o mundo sub-pessoal das causas, movimentos, do corpo e de neurofisiologia. Para Dennett este é o problema central quer de Ryle, quer de Wittgenstein, dois dos autores que o orientam. O ponto de Ryle (em *The Concept of Mind*, 1949) e Wittgenstein (nas *Investigações Filosóficas*, 1953) é impedir a interferência indevida de descrições conceptuais da vida mental (nas quais se trata de dizer o que é pensar, sentir, agir) com explicações causais (de cérebros ou outro suporte). É evidente que todo o problema surge do facto de a descrição conceptual não visar um reino mental separado.

<sup>162</sup> Para a objecção de princípio a explicações evolucionistas dos *qualia*, cf. Pinto, 1999, p. 131. Fundamentalmente, nada numa explicação evolucionista excluiria a hipótese do zombie.

<sup>163</sup> Gil 1998: “É para nós instrutivo que os diferentes conceitos do eu façam sistema, pois são muito diferentes – indicam-no claramente os seus opostos, que de nenhum modo se coordenam: respectivamente não-consciente, exterioridade, cisão ou pluralidade interna da consciência, desposseção de si [ser-se um zombie ou uma máquina, com o mesmo comportamento, os mesmos desejos, as mesmas representações ou ideias que um eu, mas sem a consciência dissol, heteronomia da vontade, não sentir. Estes conceitos referem-se a figuras muito diversas; como explicar então que as figuras contrárias pareçam fundir-se num eu?.”

Vou procurar explicar de que modo os princípios da filosofia da mente segundo Dennett partem da razão que Ryle e Wittgenstein têm neste ponto, procurando no entanto superá-los. Eles são anti-interioristas, anti-cartesianos e o seu problema é o de Dennett – pensar como o pensamento acontece, sem que se possa evocar alguém a ver – mas eles não são naturalistas nem fazem epistemologia, e a sua posição acerca da relação entre níveis pessoal e sub-pessoal é uma defesa da não-relação.

Se não há um observador, deve haver algo de errado em considerar asserções de experiência interior (relatos fenomenológicos) como relatos cognitivamente infalíveis. Torna-se então bastante importante especificar em que é que redundam a nossa prática de introspecção consciente. Wittgenstein e Ryle falarão de asserções de experiência interior como comportamentos e exteriorizações. Mas voltando a Dennett, ele trata a questão do mental (cujos problemas centrais seriam o conteúdo e a consciência), basicamente em termos da Estratégia Intencional e dos Sistemas Intencionais. Os Sistemas Intencionais são sistemas físicos com *design*, entendendo por *design* uma organização forma/função, como a da nossa mão, ou de um artefacto ou do nosso cérebro. Todas as coisas com *design* são coisas-para, coisas finalistas, e como tal descritíveis em termos de razões (seja esse *design* resultante da selecção natural, como o nosso, das nossas partes, seja artefactual). Uma coisa-para é uma coisa previsível. Por exemplo um SI pode ser explicado e previsto através da atribuição de crenças e desejos sem qualquer referência à sua constituição física. Qualquer descrição mentalista ou finalista é um atalho em relação à descrição física: há sempre outras maneiras de explicar um sistema, que afinal é físico, como uma pessoa, um animal, mas a Estratégia Intencional é mais eficaz e rápida. Evidentemente, as explicações intencionais pressupõem o que podemos chamar o carácter apropriado da finalidade e da racionalidade e este não é um dado do funcionamento do sistema físico, no qual há sempre disrupções e ruído.

Descrever segundo estes parâmetros (atribuição de crenças e desejos) sistemas físicos é abordá-los segundo a Estratégia Intencional. É o que fazemos entre nós quando nos tratamos como pessoas uns aos outros e a nós próprios, por oposição ao que um médico faz connosco em cima de uma mesa de operações. Esta é o quadro das descrições do mundo em termos de mentes e comportamento segundo Dennett. Num certo sentido, a não ser à luz de um olhar finalista não podemos dizer que há no mundo coisas como mente, comportamento e significação.

É sob este pressuposto geral da teoria da mente que Dennett distingue, no tratamento de qualquer fenómeno mental, o nível pessoal do nível sub-pessoal, em que falarmos de neurofisiologia ou, num nível mais abstracto, de modelos de cognição. Sem o nível sub-pessoal o nível pessoal

não existe (a não ser que falemos de almas desencarnadas, o que não é o caso). No entanto as características do discurso acerca de pessoas e da sua consciência de experiências (por exemplo dor) e do discurso acerca de neurónios e neurotransmissores não são idênticas (evidentemente que têm que ser acerca do mesmo mundo).

Por exemplo especificamente na dor<sup>164</sup> como se relacionam o nível pessoal e o nível sub-pessoal? Se fizermos um diagrama de caixas, como os que se usam na psicologia cognitiva e na neurofisiologia para estas questões, e que pretendem figurar a decomposição funcional da existência dos circuitos da dor num corpo<sup>165</sup>, falaremos de nociceptores, sistema límbico, cortex somato-sensorial, análise perceptual de alto nível, centros de controlo, comandos efectores, mas não poderemos localizar a dor no dígrama da dor (i.e., não há mais uma caixa para a dor, por exemplo entre a análise perceptual de alto nível e os centros de controlo). A dor sentida ela própria não aparece no diagrama porque ele é um diagrama de causas e efeitos e trajectos nervosos, é essa a relação entre as caixas e não é nessa linguagem que podemos falar de dor. É nesse sentido que separamos teoria do nível sub-pessoal da teoria do nível pessoal: só existe dor ao nível pessoal.

Vamos descrever de novo a situação, um pouco à maneira de Wittgenstein: Temos o sistema A, descrito pelo diagrama e que diz: ‘Sinto dor’. Qual é o estatuto desse dizer de nível pessoal, dessa asserção de experiência interior? É uma descrição causal? É o relato de uma acção agenciada?<sup>166</sup> Tem autoridade epistémica? É incorrigível? Deve ser tratado como susceptível de verdade e falsidade? Alguém pergunta a A, o sistema descrito, uma pessoa: ‘Porque tiraste a mão do fogo?’ A responde: ‘Porque dói’. E, como diria Wittgenstein, a explicação pára<sup>167</sup>. Dizer “Sinto dor” é uma *Ausserung*, parte do comportamento de A e não descrição ou observação de factos por A. Que A sinta dor é um facto bruto. A não é o seu próprio neurofisiólogo, não é capaz de responder causalmente, em termos de eventos nervosos (não é capaz de dizer o que está no diagrama, não sabe nada disso a partir da experiência). Uma pessoa não sabe explicar como distingue uma sensação dolorosa duma que não é dolorosa: a qualidade de dor da dor, a *painfulness*, é in-analisável ou só circularmente definível<sup>168</sup>. Como se sabe, Wittgenstein abandona o verificacionismo e

<sup>164</sup> A dor é aliás o o exemplo que Dennett usa em Dennett 1969 para exemplificar a distinção dos níveis pessoal e subpessoal. Cf. p. 90.

<sup>165</sup> Cf. Dennett 1978a, p. 216.

<sup>166</sup> O sentir como agir de Humphrey também não era um agir central intencionado

<sup>167</sup> As Investigações Filosóficas estão cheias de exemplos deste tipo de análise. Talvez o mais interessante fosse comparar o caso da dor com casos do género: “Resolve este problema de aritmética” / [acção] / “Como chegaste a esta resposta?” / “Somei e dividi” / “E como somaste e dividiste?” / Não sei, somei e dividi, é tudo”..

<sup>168</sup> Cf. Dennett 1978 a, p. 216 e segs.

o solipsismo de posições anteriores às *Investigações* também por causa deste tipo de asserções de experiência interior que não são de observação: nós não temos nenhum acesso cognitivo às causas da nossa vida mental, apenas a experimentamos, não a observamos ou conhecemos. Somos capazes de a dizer, não sabemos o que dizemos. Sentir dor acontece ao nível pessoal mas não é uma discriminação agenciada, é algo que ocorre sem qualquer actividade pessoal. Eu não sei explicar porque detesto e fujo da dor, não posso pôr-me lá, no lugar do observador-condutor a fazer a dor doer, ou a mandar a mão retirar-se (a chave desse controlo comportamental não é minha). Aliás, não existe qualquer observador-condutor (embora ele seja frequentemente necessário em teoria cognitiva) e eu não vivo sequer ao micro nível de mim que fazer essa descrição implicaria. Eu não existo a esse nível. Ou: o eu não existe a esse nível. Ou também: a esse nível de mim existem outros eus, outros controladores.

É certo que para o eu a quem dói, a ‘descrição’ pessoal mentalista da dor pára, mas começa outra explicação: a do neurofisiólogo. A conclusão de Dennett, na linha de Wittgenstein e Ryle, é que falar de dor, ou de *qualia* ou descrever fenomenologias não é linguagem referencial (referencial é a linguagem das ciências físicas). A dor é um exemplo de que os níveis pessoal e sub-pessoal não podem ser confundidos na teoria da mente. Por isso não devemos postular episódios e estados para-mecânicos ao nível pessoal nem presumir que a relação entre sujeito e conteúdos da experiência interior é epistemicamente relevante (recorde-se: dizer ‘sinto dor’ não é saber o que quer que seja do diagrama).

Aceitando que ao nível pessoal não posso presumir fazer explicações ‘para-mecânicas’ ou observacionais de mim (não tenho nenhuma posse teórica da minha interioridade, apenas sou capaz de a dizer), Dennett não pensa no entanto que as estratégias fenomenológicas de Wittgenstein e de Ryle sejam suficientes em teoria da mente. Se queremos epistemologia da fenomenologia, é preciso passar além de ambos, pois Wittgenstein e Ryle querem, como se sabe, deixar cair o que não é deste nível pessoal como sendo sem interesse para a filosofia (o que exclui para ambos qualquer discurso científico natural, causal, da teoria filosófica da mente)<sup>169</sup>.

Dennett não quer deixar cair o interesse filosófico do nível sub-pessoal: ele está interessado em epistemologia e não apenas em descrição ou

<sup>169</sup> Isto explica a razão pela qual muitos filósofos, sob a influência de Wittgenstein (passando agora a outro ponto da interioridade natural, a agência) “rejeitaram por exemplo a pressuposição do debate clássico sobre vontade livre, nomeadamente a ideia de que a acção humana poderia ser coerentemente subsumida à lei causal. A linguagem da acção seria conceptual e logicamente distinta da linguagem da causação, e por isso uma coisa é explicar o movimento corporal, outra explicar acção humana. Esta última é explicada por razões e intenções e não por causas, e a explicação tem como finalidade tornar o comportamento inteligível em termos de racionalidade. A física poderia bem ser determinista que isso não teria influência alguma na teorização da acção humana” [cf. Kim & Sosa 1995, p.180]. A ideia de uma explicação determinista da acção humana seria um erro conceptual.



fenomenologia como Wittgenstein e Ryle, e por isso tem que conceber a relação entre um nível e outro. O seu fio condutor é o seguinte: embora não seja verdade que o nível pessoal seja o único que existe quando o assunto são mentes e acções humanas, a distinção entre sub-pessoal e pessoal é absolutamente pertinente. Dennett decide então<sup>170</sup> colocar entre o nível pessoal e o nível sub-pessoal, aquilo a que chama ‘a barreira da fusão’ e considerar que o tratamento dos ‘idiomas mentais’ do nível pessoal, só pode ser holista. A linguagem da mente (que descreve ao nível pessoal) não é referencial. Não podemos tomar individuações como ‘o pensamento p da pessoa A’, por isso não se põe a questão da identidade ou não identidade do mental com o físico individualmente. Por exemplo, não faz sentido a pergunta ‘O que é que existe no cérebro que é esta dor?’ (ou o que é que há que é este pensamento?). Aliás, pensamentos não existem ao nível sub-pessoal. Os cérebros não pensam, agem ou reagem, só podem ser considerados mecanicamente. A sua descrição ao nível sub-pessoal é referencial, por oposição à descrição intencional do nível pessoal que é holista imprecisa e aproximativa. Mais: ‘a história extensional total não nos diria o que o animal (ou pessoa) está a fazer’<sup>171</sup> (ou a pensar). Entra aqui a relação entre descrições intensional e extensional, que é de interpretação<sup>172</sup>. Tudo isto aplica à vida mental em geral e em particular ao componente da interioridade que vamos ver a seguir, o conceito de acção.

#### 4. Terceiro problema para a interioridade natural: o nível pessoal e as pessoas

“Podemos, com efeito, pensar, sentir, querer, lembrarmo-nos, agir em todas as acepções do termo, sem ter consciência de tudo isso (...) (então) para que serve a consciência se é supérflua para o essencial da existência? (...)

A força e a acuidade da consciência parecem estar em razão directa com a capacidade do homem em se exprimir, e esta mesma capacidade em proporção com a necessidade de comunicação (...). Em suma, o desenvolvimento da linguagem e o desenvolvimento da consciência (não o da razão, mas somente da razão que se torna consciente de si própria) andam a par”

Nietzsche, *Gaia Ciência*, 354, *Do génio da espécie*

Se uma descrição extensional nunca diria o que um animal ou pessoa está a fazer, o que queremos dizer quando atribuímos a génese causal das nossas “acções” a um eu? Quando podemos dizer que há aí um agente, que há aí uma acção? Isto é importante porque, como disse, ser e sentir-se

<sup>170</sup> Cf. Dennett 1969, pp. 14, 16, 19, 96, 113.

<sup>171</sup> Dennett 1969, p. 78.

<sup>172</sup> Dennett 1969, p. 10. *Interpretação é outra formulação possível para ‘Estratégia Intencional’.*

agente, é uma componente central do tipo de interioridade sofisticada que é a pessoalidade. Na terminologia do ponto anterior podemos dizer que apenas ao nível pessoal há actores, agentes (ou dor). Mas se já dissemos que toda a descrição intencional, ao nível pessoal, de um humano é aproximativa e interpretativa, o problema quanto à parte da interioridade que é ser agente, sentir-se agente, é saber como é que distinguimos o agir intencional consciente de funcionamento inintencionado e inconsciente de um corpo que está lá antes, durante e depois da “acção”.

Vou tratar este problema a partir de Dennett<sup>173</sup>, que o coloca da seguinte maneira: haverá condições necessárias e suficientes para a atribuição da pessoalidade ou será esta um conceito incoerente? O conceito de pessoa envolve por um lado auto-consciência e, por outro, um auto-intencionar-se, um agenciar-se do qual podem ser dadas razões. A ideia de Dennett é que há uma interdependência destes dois aspectos. Dennett dirá por isso que a pessoa é uma noção moral-metafísica.

Saber quando há pessoalidade plena é importante até legalmente para considerarmos alguém imputável ou imputável. E há, evidentemente, humanos aos quais negamos elementos cruciais (não todos) da pessoalidade sobretudo por razões legais, falamos então de inimputabilidade em casos nos quais não é possível afirmar que uma determinada acção foi uma acção deste agente, porque não é essa pessoa que pode dar as razões dos seus actos.

Nas célebres palavras de Locke<sup>174</sup>, há pessoa se e só se houver auto-consciência (“*we must consider what person stands for: this, I think, is a thinking intelligent being that has reason and reflection and can consider itself as itself, the same thinking thing in different times and places, which it does only by that consciousness which is inseparable from thinking, and it seems to me, essential to it*”<sup>175</sup> (...)) “*Person is a forensic term, appropriating actions and their merit, and so belongs only to intelligent agents, capable of law, happiness and misery*”<sup>176</sup>. A definição de Locke, que envolve auto-consciência e a relaciona com as razões para a acção é perfeitamente geral e não tem que se ater ao humano, mesmo que nos casos experimentados até agora só humanos tenham satisfeito esta descrição. Este é o quadro do problema e a relação entre uma parte da definição e outra é do teor do problema.

Dennett analisa seis condições de pessoalidade, todas elas correctas (i.e, são condições *necessárias* da pessoalidade), para saber qual é a sua

<sup>173</sup> Dennett 1978b, *Conditions of Personhood*.

<sup>174</sup> Cf. Locke 1690, Livro II, Cap XXVII.

<sup>175</sup> Locke 1690, Book II, Cap. XXVII, 9.

<sup>176</sup> Locke 1690, Book II, Cap. XXVII, 26.

dependência umas das outras e porque é que é difícil (impossível, aliás) dizer que são em conjunto condições *suficientes* para haver pessoalidade. Estas condições são as seguintes: (1) Racionalidade, (2) Ser sujeito de uma atribuição de predicados intencionais, (3) Dependência de uma atitude tomada em relação a si (a EI), (4) Capacidade de reciprocidade nesta estratégia intencional, (5) Capacidade de comunicação verbal, (6) Autoconsciência, ligada à ideia (moral) de agente.

A ordem destes critérios de pessoalidade não é arbitrária para Dennett: é a ordem da sua dependência, excepto pelo facto de as primeiras três condições (racionalidade, intencionalidade, estratégia intencional) serem interdependentes. Por definição, “Ser racional é ser intencional é ser objecto de uma certa estratégia”<sup>177</sup>. As condições 1-3 são os conceitos com que Dennett define e trata os Sistemas Intencionais, i.e, todos sistemas físicos cujo comportamento pode ser explicado e previsto através da adscrição de crenças e desejos, como foi dito acima. Há duas coisas a notar aqui: 1. Há inúmeros SIs sem consciência 2. A consciência de tipo humano só pode advir à intencionalidade de SIs, e não em qualquer outra parte do mundo físico. Isto quer dizer que para Dennett no não-*design* (na não-racionalidade) não pode existir a consciência, mas pode perfeitamente existir *design*, racionalidade, sem consciência.

Começa-se pelo uso da linguagem intencional. Em termos de ciências do cérebro e do comportamento a completa abstenção de linguagem intencional para falar de animais, pessoas e algumas máquinas torna a descrição dos fenómenos mentais e comportamentais impossível de manejar e é absurda, como o behaviorismo com o seu tabu do mental, mostrou na prática psicológica até à exaustão. Como Dennett nota, a maneira mais simples de embaraçar o behaviorista é montar uma experiência em que se engane os sujeitos (por exemplo ratos). Se eles são enganados, estão a ter crenças falsas, pelas quais o seu comportamento é previsível e não pelo ambiente actual.

Mas não basta que se possa e deva utilizar linguagem intencional acerca de um sistema para afirmarmos que há aí pessoalidade. Não basta dizer que as Pessoas são a subclasse dos Sistemas Intencionais (que satisfazem as condições 1-3) constituída pelos seres que *realmente* têm crenças e desejos (e não apenas funcionam como-se) porque são conscientes.

Será que é então a reciprocidade, a condição 4, que prevê que um Sistema Intencional adopte a Estratégia Intencional em relação a outro sistema que a adopta em relação a ele, que supõe necessariamente a consciência e marca o limiar da pessoalidade? Os seres humanos não são os

<sup>177</sup> Cf. Dennett 1978, p. 271.

únicos Sistemas Intencionais capazes de atribuição de intencionalidade ao outro. Vários animais são “psicólogos naturais” (a expressão é de N. Humphrey), i.e., são Sistemas Intencionais de segunda ordem neste sentido, e portanto pode-se-lhes adscrever crenças acerca de crenças. Vários animais são por exemplo capazes de comportamento de logro, logo mostram não apenas a capacidade de acreditar o falso como o desejo de induzir crenças falsas no outro (por exemplo, há pássaros que fingem ter uma asa partida para enganar o predador, que assim os considera presa fácil, e o afastar do ninho<sup>178</sup>). Se o falso não é descritível sem uma disparidade entre crença e aquilo que é, no logro está envolvido algo como: A acredita que q e deseja que B acredite que não-q. Há portanto crenças acerca de crenças. Os humanos não são os únicos a ter crenças acerca de crenças. É uma questão empírica saber que outros sistemas as têm (estes pássaros tê-las-iam). Mas o importante para Dennett é que isso mostra que a linguagem natural não é estritamente necessária para a existência de um Sistema Intencional de segunda ordem, com crenças acerca de crenças e que tem assim comportamentos que se fossem de um humano seriam inteligentes (mas não o são).

O que estamos a fazer então quando descrevemos comportamentos assim? Não é que não haja inteligência nestas situações e comportamentos, mas ela não é do “agente”, é no “agente”: quando descrevemos o comportamento do pássaro descrevemos “a ordem que está lá” mesmo que não corresponda a intenções conscientes. Há aqui um mundo proto-voluntário que não é pessoal: há razões para a acção no “agente” que não são razões para o agente.

A reciprocidade das crenças acerca de crenças é sem dúvida um grande passo em direcção ao nosso tipo de personalidade, mas o facto de comportamento que as envolvem poderem ser rigidamente fixados, instintivos em tantos animais, mostra-nos que não é verdade que as crenças e desejos do sistema intencional tenham que estar representadas explicitamente para o agente para a previsão do seu comportamento inteligente por adscrição dessas crenças ser possível. A reciprocidade (4) é independente de linguagem (5) e de auto-consciência (6).

Mas então o que é que a existência de linguagem faz a esta situação? Para Dennett, as condições griceanas<sup>179</sup> da significação servem de pista quanto à diferença que a existência de linguagem faz num sistema intencional. Nomeadamente, a célebre ideia griceana de *non natural meaning*, que Grice define pela existência de uma situação em que um enunciador quer dizer alguma coisa enunciando alguma (outra) coisa, devendo este

<sup>178</sup> Cf. Dennett 1978b, p. 276.

<sup>179</sup> Cf. Grice 1957 e Grice 1959.

querer-dizer ser analisado em termos das intenções do enunciador, é esclarecedora aqui. O *non natural meaning* faz a diferença do modo humano de significar: quando alguém faz uma asserção significando alguma coisa querendo com essa enunciação significar alguma coisa diferente, isso implica a existência de uma terceira ordem no funcionamento de Sistemas Intencionais. A deve pretender que B reconheça que A (in) tenciona que B produza uma dada resposta<sup>180</sup>. Para começar isto mostra, segundo Dennett que a condição 5 depende da condição 4 (reciprocidade) e não vice-versa. Em segundo lugar mostra que a segunda ordem não é suficiente para uma reciprocidade genuína ou usual de tipo humano. É preciso mais, é preciso o reconhecimento mútuo tácito (o “encontro”) que subjaz aos casos de Grice.

Grice mostra que a comunicação linguística humana, ao ser esta possibilidade de manipulação, depende da confiança. Deve existir aqui uma norma, e essa norma é a sinceridade, o dizer o verdadeiro (evidentemente não no sentido voluntarista de eu fazer de propósito para ser sincero quando falo). O sucesso da manipulação linguisticamente possível não depende do facto de o entendedor se dedicar conscientemente a estes raciocínios (como o pássaro da “falsa” asa partida também não o fazia): nem quem fala nem quem compreende precisa de estar a dedicar-se elaboradamente e conscientemente à decifração da intenções “encaixadas” que Grice analisa. Importante é esta presença da terceira ordem.

Mas mesmo sendo a pedra de toque do algo a mais que a presença de linguagem envolve na interioridade de tipo humano, o que é relevante é que as intenções griceanas são inconscientes. Se fossem voluntariamente intencionadas seriam inteligentes, mas não o são. Em suma, até aqui chega-se sem consciência, como Nietzsche muito bem viu.

Estamos a chegar ao que interessa mais: Dennett pensa que se existe uma dependência entre consciência e intencionalidade/racionalidade, aquilo que temos é uma dependência da consciência em relação a todos estes tipos de intencionalidade e inteligência que podem ainda ser inconscientes. Dennett pensa mesmo que as primeiras cinco condições podem ter, sozinhas, um papel na ética: isso é-lhe sugerido pela teoria rawlsiana da justiça. Como é sabido, da posição original<sup>181</sup> são derivados segundo Rawls, os princípios da justiça. Como se passa isto? A teoria da justiça é uma parte da teoria da escolha racional. A norma da interacção humana que serão os princípios da justiça acordados não é derivada intencionalmente,

<sup>180</sup> Por exemplo uma situação como esta só seria possível com linguagem verbal: eu digo “vou sair” para que B pense que o estou a querer enganar e venha a correr a minha casa ver-me. Assim eu posso sair.

<sup>181</sup> Cf. Rawls, *Uma Teoria da Justiça*, 1971, p.37, Capítulos II e III.

voluntariamente, não se presume qualquer moralidade ou bondade das partes. A solução óptima é uma solução racional para calculadores interessados no seu próprio bem e que têm que considerar as crenças e desejos dos outros. Não fazer assim, i.e., não procurar derivar o moral do não moral sem voluntariedade, ficar à espera de obter o moral ou dá-lo como garantido (como nas teorias dos direitos humanos) seria *naïf*.

Assim como “dizer o verdadeiro” é a norma da enunciação (e não é porque queiramos ser sinceros), o “ser justo, tratar o outro como pessoa” será a norma da interacção pessoal (e não é que queiramos ser justos).

Aliás, a interacção humana não tem que ser regida pelos princípios acordados: apenas interacção humana pessoal o será. De toda esta descrição Dennett conclui que o conceito de pessoa é inescapavelmente normativo e idealizado. Uma pessoa não é uma coisa natural no mundo.

Mas o que foi dito até aqui não explica porquê e quando podemos por vezes considerar pessoas individuais como responsáveis pelas suas acções. Aliás, poderia até ser utilizado no sentido de dizer que nunca o podemos fazer. Quem pode ajudar nessa tarefa, segundo Dennett, é Elizabeth Anscombe. Na sua análise da intenção (i.e., do intencional intencionado, da acção, portanto) Anscombe defendeu<sup>182</sup> que só chamamos acções a movimentos apercebidos ou praticados segundo descrições mentalistas (se eu caio de um segundo andar isso não é uma acção, por oposição ao caso em que eu me atiro). Aquilo que acontece pode ser num certo sentido o mesmo, mas só no segundo caso me podem ser pedidas razões.

Imaginemos o movimento do braço de uma pessoa que se move para espetar um punhal no peito de alguém. É um movimento, um evento físico, só é uma acção mediante um apercebimento consciente (de alguém). Imaginemos que estive hipnotizada e acordo no momento em que constato que acabei de fazer isso. Será assim mais fácil entender a caracterização das acções intencionais segundo Anscombe: (1) Acções não são simplesmente movimentos corporais mas movimentos sob descrições particulares; (2) Para uma acção ser intencional, o actor tem que aperceber a acção, i.e., tem que aperceber o movimento sob a particular descrição (se estou a bater nas teclas do computador ao ritmo de *Rule Britannia* – e me dizem que estou a fazer isso, e eu digo: ah! estou a fazer isso?- isso não é intencional), (3) Não é suficiente que eu me aperceba do que estou a fazer para que o que estou a fazer seja intencional. Tem que haver apercebimento *sem observação* (se eu apenas observei que estava a fazer isso – que era eu que estava a espetar o punhal,...isso não é intencional).

A acção é portanto “coisa mental”, da qual há razões, por oposição ao movimento, que tem causas. E este é o topo da construção de pessoa:

<sup>182</sup> Cf. Anscombe 1957.

*pessoa é uma fonte de ações conscientes e intencionadas das quais pode dar razões. O que se impõe então verificar é que um humano não é só isto nem é sempre isto, precisamente porque este design da interioridade não caracteriza uma imaterialidade não mundana, mas foi surgindo de descrições normativas.*

Evidentemente, só as ações cabem na ordem das razões explicitamente conscientes (no nível pessoal). Se eu não estou numa posição privilegiada, não há razão especial para me perguntar a mim a razão da ação. Mas nem por isso podemos afirmar que onde há razões não há causas (ao contrário do que Anscombe queria). A causação subsiste sempre, as razões pertencem ao nível pessoal, aquele cuja descrição nunca é mais do que imprecisa (do ponto anterior).

Que tipo de acesso e de domínio temos então das nossas razões para agir? Uma pessoa enquanto sistema capaz de auto monitorização e sondagem de si certamente não tem um acesso epistemicamente incorrigível ao nível pessoal. Aliás o verdadeiro ponto, se recordarmos a caracterização dos níveis pessoal /sub pessoal é que talvez elas não sejam absolutamente determinadas, como nada no mundo mentalmente descrito o é. Se a descrição intencional à qual as razões pertencem é sempre aproximativa, não há nenhuma resolução completa de movimentos em ações, de causação em razões, e por isso não há uma racionalidade sem resto das ações humanas. Pelo menos é o que pensa Dennett, para quem não há razão prática pura, nem máquinas semânticas, nem puros agentes, e que completa a sua descrição da personalidade assim: alguém só sabe o que está a fazer ou a tentar fazer se puder dar a razão da ação. É isso que faz desse alguém o agente da ação. Harry Frankfurt<sup>183</sup> apresenta um conceito útil para conceber a situação: à prática de um movimento como ação voluntária e consciente Harry Frankfurt chama uma 'volição de segunda ordem' (além de desejar, nós podemos querer querer ou não querer ter certos desejos e motivos, essas são volições de segunda ordem). As Pessoas são esse tipo de interior natural: a sub-classe dos Sistemas Intencionais capazes de volições de segunda ordem e de agir de acordo com elas. Esta seria uma condição essencial para ser uma pessoa. Segundo Dennett, só esta avaliação reflexiva é/possibilita a auto-consciência, que implicará a adoção por um humano, um sistema físico, causal, em relação a si próprio do papel de argumentador persuasor e exigidor de razões sobre si, como sobre outra parte do mundo, Para Dennett, só esta estratégia e este particular acesso a si é auto-consciência genuína. Esta está portanto ligada à ideia de agente. Só aqui 'a ordem que está lá' de Anscombe não pode estar a não ser que esteja em episódios de pensamento consciente.

<sup>183</sup> Cf. Frankfurt 1971.

A primeira conclusão é que por vezes temos pessoas e as suas acções, mas evidentemente não sempre, porque para isso não bastam corpos humanos e os seus movimentos e funcionamentos. Os conceitos de pessoa, agente, acção, razão são normativos por isso só podemos ter deles aproximações. Nenhuma entidade ou evento actual os preenche completamente e sempre, em todos os momentos. (isto, é claro, se ao contrário de Kant não admitirmos nenhuma descontinuidade entre desejar e querer). A inexistência dessa descontinuidade é uma das interpretações da ideia de Dennett segundo a qual “Uma máquina semântica é uma impossibilidade física”: nada no mundo se move ou acontece puramente devido a razões.

A segunda conclusão de Dennett é a continuidade entre os aspectos moral e metafísico do conceito de pessoa. Dar-se a si próprio as razões da acção nunca pode ser um “acto puro”, uma comunicação directa, pois razões são descrição normativas, *short cuts*. Ninguém é movido por razões em vez de causas, aliás é por isso que não podemos fazer uma distinção absoluta entre ser consciente de razões e guiar-se por elas e agir como se se fosse guiado por razões (*pace* Kant...). A razão pura prática é impossível para entidades físicas, mas a racionalidade consciente de um agente, é certamente possível: o verdadeiro problema é saber como, e é isso que Dennett pretende mostrar com as condições de pessoalidade. É claro que os casos polémicos são precisamente aqueles em que é abalada a nossa convicção de que um humano é ou foi uma pessoa, em que não sabemos se estamos a lidar com uma pessoa em todos estes sentidos (por exemplo quando uma criança de onze anos assassina outra ou quando sob influência de drogas ou hipnose ou dormindo alguém comete o que seria um crime). E evidentemente haverá casos em que não poderemos dizer de nós próprios se fomos ou somos pessoas.

#### **Quarto problema para a interioridade natural: A Identidade Pessoal e o alargamento ‘daquilo que importa’ (*what matters*)**

“Queremos saber de onde vem a adesão a si do sujeito - mau grado a dificuldade em se aferrar.”, Fernando Gil, Eu, *Modos da Evidência*

Como vimos, aquilo que fazemos é ameaçado de uma gradual erosão pela “subtracção daquilo que acontece”<sup>184</sup>. Esta subtracção dá-se mesmo na parte do mundo pela qual cada um de nós seria um agente. Mas exactamente como é que o nosso sentimento de agir está especialmente ligado a uma particular parte do mundo pelo qual agimos e que não é o mundo todo? A resposta a esta pergunta passa inevitavelmente, como vimos, pela

<sup>184</sup> Nagel, *Mortal Questions*, “Moral Luck”, p.38.



relação entre auto-consciência e controlo de si, mas também pela Identidade Pessoal. Esta pode ser formulada assim: o que é que eu penso que sou? Sou aquilo que penso ser? Essa crença é justificada?

Se mudarmos o nosso pensamento quanto àquilo que pensamos ser, mudaremos também o nosso pensamento quanto àquilo que pensamos ser racional e moral fazer. É por isso que a Visão Reduccionista da Identidade Pessoal que Parfit defende conduz a alterações nas teorias da racionalidade e da moralidade e portanto na ética. A minha intenção é dar alguns exemplos desta ligação, que são ao mesmo tempo exemplos de um alargamento daquilo que importa para além do indivíduo que somos. Esses exemplos provêm de *Reasons and Persons* (1984), um livro de Parfit considerado por muitos como um marco incontornável da filosofia moral da segunda metade do século XX. Como disse no início, a questão da interioridade no seu aspecto ético envolve não apenas o que chamei a partir de Dennett *condições de pessoalidade*—estas são muito importantes para defendermos a possibilidade de existência de agentes auto-conscientes e racionais e portanto a significatividade de acções por oposição a movimentos—mas também a Identidade Pessoal. Um pensamento daquilo que sou acompanha sempre aquilo que eu penso que deve ser a minha acção.

No capítulo XXVII do segundo livro do *Essay Concerning Human Understanding*, que citei atrás, Locke lançou a maneira reduccionista contemporânea de pensar a Identidade Pessoal, ao definir pessoa através da autoconsciência e da memória de actos passados<sup>185</sup>, separando assim a mesmidade pessoal da mesmidade de uma alma possivelmente imortal ou da mesmidade de um corpo orgânico<sup>186</sup>. Segundo Locke: “*as far as this consciousness [aquilo pelo qual um ser pensante e inteligente se considera o mesmo consigo próprio] can be extended backwards to any past action or thought, so far reaches the identity of that person*”. Sem este prolongamento no tempo de um agente capaz de ‘*law, happiness and misery*’, não pode haver acção significativa. Locke viu que esta ideia de mesmidade de uma pessoa não tinha que especialmente ater-se ao humano, e que ela era num certo sentido jurídica (“*pessoa é um termo forense*”, disse Locke<sup>187</sup>), ao permitir a autoria de actos. Viu também que sem a explicitação desta noção questões tão básicas na filosofia moral e política como a culpa e o

<sup>184</sup> Nagel, *Mortal Questions*, “*Moral Luck*”, p.38.

<sup>185</sup> Segundo Parfit deveríamos acrescentar intenções de futuro, planos de vida em geral.

<sup>186</sup> Enquanto corpos orgânicos, a nossa mesmidade seria estabelecida do mesmo modo que a de vegetais ou animais, segundo Locke. Para a posição de Locke quanto a identidade pessoal e relação desta com o tema mais geral da identidade, cf. John Locke, *Essay Concerning Human Understanding*, Book II, Cap XXVII, *Of Identity and Diversity*, especialmente os pontos 8. *Same Man*, 9. *Personal Identity*, 10. *Consciousness Makes Personal Identity*, 11. *Personal Identity in change of substances*, 16. *Consciousness makes the same person*, 17. *Self depends on consciousness*. Para a explicitação da posição de raiz lockiana de Parfit, cf. Engel 1996, p. 165.

<sup>187</sup> Locke 1690.

merecimento, a manutenção de compromissos<sup>188</sup>, a compensação, a punição não poderiam ser racionalmente tratadas.

O teste de Parfit para a Identidade Pessoal (e ela só pode ser verificada testando-a, uma vez que aquilo que ela é não se distingue dos critérios para o seu reconhecimento) é o seguinte: Quando, em que circunstâncias, afirmaríamos que fomos/somos/seremos o mesmo eu? Através de casos reais e de experiências de pensamento, mas de qualquer modo, testando, saberemos exactamente o que é que importa para podermos dizer que sobrevivemos como a mesma pessoa.

Resumindo o resultado das inúmeras e minuciosas experiências de pensamento que constituem este teste<sup>189</sup>, Parfit conclui que nós não somos aquilo que usualmente pensamos ser. Não somos aquilo que pensamos ser porque a Identidade Pessoal não é um facto a mais (um *deep further fact*). A nossa Identidade Pessoal é a continuidade psicológica e o que quer que fisicamente a garanta (no caso normal um mesmo cérebro biológico, mas não necessariamente). O que é importante aqui é que esta continuidade psicológica é intransitiva, é uma questão de grau, logo nós estamos muito mais fracamente ligados a partes das nossas vidas (muito passadas, muito futuras) do que a outras (imediatamente passadas, imediatamente futuras)<sup>190</sup>. É por isso que somos naturalmente temporalmente preconceituosos em relação ao que importa para nós. Além disso, para Parfit, se a Identidade Pessoal não é um facto a mais além da continuidade, devemos deixar de pensar que existe sempre uma resposta determinada à pergunta, feita quanto a uma determinada experiência, “essa experiência é/foi/será minha, ou não?”. Isto aplica-se às minhas duas unidades, ou às duas unidades que são o eu: o meu eu agora todo – pense-se no exemplo de Dennett quanto à dor com que iniciei este texto – e a minha vida consciente ao longo do tempo. Para Parfit há margens não determinadas de cada uma destas unidades.

As principais objecções ao reducionismo foram postas a Locke: se substituímos identidade por continuidade psicológica, e consideramos

<sup>188</sup> Cf. Parfit 1984, p. 327, o caso do russo do século XIX. Um jovem nobre com ideais socialistas tem a intenção de doar aos camponeses as vastas propriedades que herdará. Sabendo que pode vir a deixar de ter esses ideais, prepara e assina um documento oficial que automaticamente efectiva a intenção e que só pode ser revogado com o consentimento da mulher. Ela promete que nunca dará esse consentimento, mesmo que ele mude de ideias e lho peça. Ele diz: ‘Encaro os meus ideais como essenciais àquilo que eu sou. Se alguma vez os perder, deixarei de existir. Quero que encares o teu marido de então não como sendo eu, aquele que te está a pedir esta promessa, mas como um eu posterior corrompido. Promete-me que nunca farás o que ele então pedir’. Na meia idade, o nobre russo pede à mulher que consinta na revogação do documento. Pode ela considerar-se liberta do compromisso anterior?

<sup>189</sup> Sem irmos aos casos da fissão e fusão, transplantes de cérebros e partes de cérebros, teletransporte e implante de memórias, em que Parfit se deleita, pense-se apenas em estados posteriores a AVCs graves ou comas ou amnésias.

<sup>190</sup> Temos aliás preconceitos naturais a esses propósitos, queremos por exemplo que uma dor a ter (nossa) seja muito futura.

(como Parfit faz) que os factos particulares que constituem a identidade de uma pessoa através do tempo podem ser descritos impessoalmente, mais como eventos de *Es denkt* lichtenbergiano do que como modos de um cogito, o que fazemos com as quebras, as perdas de consciência, o sono, a amnésia, o esquecimento? Se somos assim descontinuados e indeterminados, seremos sempre novos, outras pessoas em relação a eus anteriores<sup>191</sup>? E não será circular explicar a identidade pela memória se nós só temos precisamente as nossas memórias<sup>192</sup> e não as de outrem?

Parfit responde a todas estas objecções, mas atente-se no ponto central da resposta: o que Parfit pensa é que ‘ser ou não ser’ (eu, uma experiência minha) não é uma coisa assim tão determinada. É por isso que pode ser por exemplo racional considerar que uma pessoa deixou de existir antes de todos os seus órgãos deixarem de funcionar (é o que faz a medicina hoje), ou considerar que apenas matar pessoas (e não seres humanos) é errado<sup>193</sup>.

Como disse no início, o que Parfit pretende, ao ligar os temas ‘Razões’ e ‘Pessoas’ é fazer nascer uma ética (e não mais Meta-Ética<sup>194</sup>) da análise da ligação entre concepções de racionalidade na acção e teorias da identidade Pessoal. Podemos considerar que o fundo desta ligação é o facto de para muitas pessoas ser muito pior ser acusado de irracionalidade do que de imoralidade<sup>195</sup>. O cepticismo moral é bastante natural e para Parfit a demonstração de que não devemos ser cépticos morais só pode ser conseguida através de uma teoria da racionalidade, i.e., provando a alguém que tentar prosseguir os fins dados por uma determinada teoria da moralidade/racionalidade (os objectos de análise de Parfit são o egoísmo racional, o consequencialismo, e a moralidade de senso comum<sup>196</sup>), é pior nos seus resultados do que não ser assim guiado, mostrando que se está a ser irracional, e que há razões para a acção que não estão a ser reconhecidas. Parfit afirma que três teorias da moralidade/racionalidade são auto-refutantes,

<sup>191</sup> Esta objecção é de Reid.

<sup>192</sup> Esta objecção é de Butler.

<sup>193</sup> Penso que esta discussão ganha em partir de uma distinção kantiana entre sujeito transcendental (autoconsciência como condição do pensamento qualquer) e sujeito empírico (identidade pessoal). As conclusões de Parfit acerca de identidade pessoal não são acerca desse sujeito transcendental. Evidentemente as questões tendem a fundir-se devido à reticência em considerar o sujeito transcendental como separado do mundo físico. O anticartesianismo necessário aqui como primeiro passo é também o movimento de Kant nos Paralogismos [cf. *Crítica da Razão Pura*]: não há intuição do eu enquanto objecto, embora a ilusão natural consista na transformação da unidade da consciência na intuição do sujeito enquanto objecto. Mas há uma outra unidade da autoconsciência, que é forçada pelo reconhecimento de uma actividade (e num sentido de actividade em que o puro pensamento não se opõe a acção). Há uma agência no acontecer do pensamento (o espírito sabe que é um porque tem debaixo dos olhos a unidade do seu acto...) que talvez não possa ser confundida com a questão da identidade pessoal. Cf. Blackburn, *Has Kant refuted Parfit?* in Dancy 1997.

<sup>194</sup> Parfit quer conclusões substantivas quanto ao que seria bom e mau nos resultados de acções e não fazer uma análise do nosso vocabulário moral.

<sup>195</sup> Parfit 1984, *Concluding Chapter*, p.450.

<sup>196</sup> A moralidade de senso comum (M) é menos explicitamente teórica do que as outras duas, evidentemente, mas envolve nomeadamente, como traço central, a parcialidade em relação aos próximos.

que se auto-anulam e utiliza inúmeros exemplos práticos que supostamente o provam. É evidente que há no método de Parfit uma aposta intelectualista, i.e. supõe-se que aquilo que somos conduzimos a pensar e acreditar é influente no modo como regemos a nossa vida moral, e que uma vez tendo verificado que temos crenças indefensáveis porque irracionais abdicaremos delas. Mas como em qualquer intelectualismo, a resposta às objecções é: e qual é a alternativa?<sup>197</sup>

Um ponto central em *Reasons and Persons* é que a Visão Reducionista da Identidade Pessoal permite sustentar argumentos contra o egoísmo racional como teoria da racionalidade na acção, tornando mais plausível a impessoalidade utilitarista. Penso que o aspecto mais importante desta discussão é o facto de exatamente esta teoria da racionalidade, o egoísmo racional e um certo utilitarismo como ‘moderação’ deste via a benevolência racional desejável, serem normalmente os instrumentos filosóficos *prêt-à-porter* quando se trata de justificar publicamente decisões colectivas, decisões de *public policy*, económicas e sociais, nas sociedades em que a questão da justificação pública se põe. Como diz A. Sen, como disse Adam Smith, a economia é filosofia moral<sup>198</sup>, seja explícita ou implicitamente, ao ser a área em que decisões com peso moral são tomadas em nome de uma teoria da racionalidade. As decisões morais de peso colectivo só poderiam ser tomadas em nome da racionalidade na medida em que o ponto de vista moral que desejamos tem que ser nitidamente para-individual (pense-se nas limitações de um pensamento moral como o de Lévinas quando se trata de filosofia moral pública). Algum cálculo racional tem que ter direito de cidadania teórica, ou nunca haverá mais do que duas pessoas na situação moral, e nunca haverá justiça ou imparcialidade. Historicamente, a tradição utilitarista foi muito importante para este fim (daí também a sua importância para o pensamento jurídico)<sup>199</sup>.

Mas o problema do utilitarismo, maugrado a sua boa atenção à quantidade de prazer e sofrimento<sup>200</sup>, a sua visão da necessidade de uma consideração global, supra-pessoal daquilo que importa (que é afinal atenção à situação moral como envolvendo muitas e não uma ou duas pessoas) é, como bem viu H. Sidgwick<sup>201</sup>, *the separateness of persons*, o carácter separado das pessoas, o facto de pessoas serem entidades individuadas.

<sup>197</sup> Como Parfit bem sabe (cf. p. 120) há aqui um objecção ‘humiana’ que não pode ser afastada com uma profissão de fé intelectualista: a afirmação de que desejos não podem ser racionais ou irracionais.

<sup>198</sup> Cf. Sen 1993, *Comportement économique et sentiments moraux*, p.6.

<sup>199</sup> Cf. Goodin 1995, acerca do utilitarismo como sendo por vocação uma ‘filosofia pública’.

<sup>200</sup> Que evidentemente o conduz à justificação do sacrifício do indivíduo. Um autor que tem acentuado o carácter sacrificial do utilitarismo é Jean-Pierre Dupuy.

<sup>201</sup> Cf. H. Sidgwick, *Methods of Ethics*, London, McMillan 1907.

Isso torna a benevolência racional muito pouco natural. Pelo contrário, o que é natural é o egoísmo, que é obviamente a forma bruta, o aparecimento psicológico, da teoria da racionalidade como interesse próprio. Para vermos o quão pouco somos benevolentes racionais basta pensarmos em casos de manual de ética: por exemplo o caso em alguém tem que decidir salvar o filho ou dez mil pessoas, quando o filho pilota o avião que vai lançar a bomba sobre a cidade, ou o caso que alguém deve decidir torturar a filha inocente do terrorista na sua frente, para obter informações que evitarão a morte de milhares de pessoas. Penso que isto torna clara a importância do problema das pessoas e da sua identidade para as decisões quanto àquilo que importa: sermos ou não sermos separados uns dos outros e determinados como pessoas, termos uma inclinação favorável aos nossos próximos, tem realmente a ver com a benevolência, imparcialidade e motivação racional de que somos capazes na acção. Mas se a Visão Reduccionista é verdadeira, a impessoalidade do utilitarismo é muito menos implausível do que teríamos tendência a pensar.

Recapitulando, a impessoalidade é o ponto comum das avaliações que Parfit faz de teorias da racionalidade e Identidade Pessoal. Devemos segundo Parfit procurar impessoalidade nas razões para agir. Ora, um campo de eleição para a sua discussão desta tese é a parte da ética (central para Parfit) que diz respeito às Gerações Futuras. Para que a teoria do Fim-Presente<sup>202</sup> que Parfit defende venha a ser uma filosofia prática elaborada, teria que passar a prova da sua utilidade na parte da ética que diz respeito às Gerações Futuras, na escolha racional daquilo que importa quando se trata da distribuição de dor, prazer, sofrimento, por pessoas futuras incluindo aquela a que chamo 'eu'.

O primeiro passo na abordagem desta questão é o Problema da Não Identidade, que se enuncia assim: Nós poderíamos não ter existido. As pessoas futuras estão nas mesmas circunstâncias e dependentes das decisões presentes. Segundo Parfit, três séculos daqui para a frente, não haverá ninguém a viver que viveria se as decisões presentes tivessem sido diferentes. E se ninguém é tão infeliz e miserável que seria melhor nunca ter existido, segue-se que não haverá ninguém a viver para quem teria sido melhor se uma diferente política tivesse sido seguida<sup>203</sup>. Portanto, não é por afectarmos os interesses de pessoas determinadas que podemos avaliar a bondade ou não bondade das decisões (se elas não serão as mesmas pes-

<sup>202</sup> Que abdica da neutralidade quanto ao tempo do egoísmo racional mas não considera a separação das pessoas como básica. Cf. Parfit 1984, o caso global das Partes I, II, III.

<sup>203</sup> Parfit 1984, p. 361.

soas numa alternativa e noutra, *quem* estaria a ser afectado?). Em suma, é prévio ao problema da escolha de um bem saber se esse bem é bem para alguém. Ora frequentemente será o caso que esse alguém só existirá no caso de a decisão—que está precisamente em causa—ser de uma determinada maneira e não de outra. Um exemplo de Parfit são os Dois Programas Médicos<sup>204</sup>.

Estas considerações mostram segundo Parfit que uma Teoria da Beneficência não pode consistir em Princípios de Afecção de Pessoas, mas em princípios relativos à qualidade de vidas. Para saber o que beneficia alguém temos não apenas que saber o que é esse alguém como também se esse alguém existirá de qualquer maneira (na linguagem de Parfit são diferentes as “Same People Choices” das “Different People Choices”<sup>205</sup>). Dizer que precisamos de uma teoria do que é beneficência significa que não queremos estar, como Parfit pensa que usualmente estamos, completamente enganados quanto àquilo em que consiste beneficiar e prejudicar outras pessoas.

Um exemplo muito importante é sumariado pelo caso dos Torturadores Inofensivos. Pensamos frequentemente que a avaliação moral se pode reportar apenas a efeitos de acções isoladas. Mas grande parte das situações humanas que segundo Parfit é importante avaliar moralmente caem sob seguinte descrição:

“Nos Maus Velhos Tempos, cada torturador infligia uma dor terrível a uma vítima. As coisas mudaram. Cada um dos mil torturadores pressionou um botão accionando o interruptor de cada um dos mil instrumentos de tortura uma vez. A vítima sofre a mesma dor terrível. Mas nenhum dos torturadores torna a dor da vítima perceptivelmente pior”<sup>206</sup>. De novo, apenas um ponto de vista impessoal tornará moralmente avaliáveis estas situações.

Um problema da Impessoalidade desejada das razões para agir é em geral conduzir à Conclusão Repugnante: Para qualquer população possível

<sup>204</sup> Cf. Parfit 1984, p.367, *Os Programas Médicos*: Há duas condições raras, J e K, que não podem ser detectadas sem testes especiais. Se uma mulher grávida tem a condição J, isto fará com que a criança em causa venha a ser portadora de uma certa deficiência. Um simples tratamento impediria este efeito. Se uma mulher tem a condição K quando concebe a criança, isto fará com que a criança seja portadora da mesma particular deficiência. A condição K não pode ser tratada, mas desaparece sempre ao fim de dois meses. Suponhamos que há dois programas de investigação médica planeados, mas que só há fundos para um, portanto o outro tem que ser cancelado. No primeiro programa, milhões de mulheres seriam testadas durante a gravidez. Aquelas em quem se verificasse a condição J seriam tratadas. No segundo programa milhões de mulheres seriam testadas quando tencionassem engravidar. Aquelas em quem se verificasse a condição K seriam aconselhadas a adiar a sua intenção pelo menos dois meses, depois dos quais esta condição intratável teria desaparecido. Suponhamos, finalmente, que podemos prever que estes dois programas alcançariam resultados num número idêntico de casos. Se se escolher o Teste na Gravidez, 1000 crianças por ano nascerão normais e não com uma deficiência. Se se escolher o Teste anterior à concepção, 1000 crianças normais nascerão por ano em lugar das 1000 (diferentes) crianças portadoras da deficiência.

<sup>205</sup> Parfit 1984, p.356.

<sup>206</sup> Parfit 1984, p.80.

de pelo menos 10 bilhões de pessoas, todas com uma muito alta qualidade de vida, deve haver alguma população imaginável muito maior, cuja existência, se todas as outras coisas forem iguais, seria melhor, mesmo apesar dos seus membros terem vidas que mal merecem ser vividas<sup>207</sup>.

A intuição que Parfit quer defender é que devemos rejeitar essa conclusão. Não se segue do Problema da Não Identidade e da sua ligação à Conclusão Repugnante que não temos obrigação de pensar nas decisões sociais e políticas respeitantes a pessoas futuras incapazes de reciprocidade. A questão é que se queremos evitar a Conclusão Repugnante não podemos resolver o Problema da Não Identidade através de um Princípio de Afecção de Pessoas. A conclusão “não é bem vinda” pois abre problemas que Parfit considera não conseguir resolver.

Mas outras conclusões de *Reasons and Persons* são “benvindas”, como o reducionismo quanto à Identidade Pessoal: Parfit pensa que devemos ser reducionistas quanto à identidade pessoal e que isso será eticamente melhor. Quanto ao efeito nas emoções da Visão Reducionista ele afirma que vê a verdade como ‘libertadora e consoladora’ (‘faz-me preocupar-me menos com o meu próprio futuro e a minha morte e mais com os outros e dou as boas vindas a este alargamento na minha preocupação’<sup>208</sup>). Mas Parfit admite que este efeito nas emoções (da sua conclusão) pode ser diferente para diferentes pessoas. Em geral, a Pretensão Extrema a que somos conduzidos pela Visão Reducionista, é que não temos qualquer razão para nos importarmos com o nosso futuro. Porque é que uma parte da série de sentimentos deveria estar mais preocupada com outra parte da mesma série do que com qualquer outra série?<sup>209</sup> O que está em causa na “adesão a si”?<sup>210</sup>

Filósofos religiosos contemporâneos como R. Swinburne afirmaram que se Parfit tem razão deveríamos ser indiferentes quanto a vivermos ou morreremos<sup>211</sup>, pois a continuidade por si não tem valor. Outro filósofo, John Perry, que ao contrário de Swinburne subscreve o reducionismo, argumenta que se Parfit tem razão, o facto de uma dor futura ser minha não me daria mais razão para preveni-la do que o facto de ser de outro<sup>212</sup>. Não há mais razões para impedir a dor de um estranho do que uma dor minha futura. Eu não deveria ser mais central para mim próprio na consideração de razões para agir do que outra pessoa qualquer.

<sup>207</sup> Parfit 1984, p.381.

<sup>208</sup> Parfit 1984, p. 347.

<sup>209</sup> É a questão de Sidgwick, cit. in Parfit 1984, p. 307.

<sup>210</sup> Cf. Gil 1998.

<sup>211</sup> Cit. in Parfit 1984, p. 307.

<sup>212</sup> Argumento de Perry, cf. Parfit 1984, p.308. Cf. também J. Perry, “The Importance of Being Identical” in Rorty 1976, p.67.

A adesão a si envolve escolhas religiosas, e o próprio Parfit relaciona o seu pensamento com o budismo. No entanto, como salienta Ribert Merrihew Adams, é interessante notar que embora de facto “os textos budistas utilizem argumentos metafísicos reducionistas acerca da identidade num esforço para enfraquecer atitudes como a autocentração e a ansiedade acerca da morte (...) essa não é a única resposta religiosa possível a conclusões reducionistas”. Kierkegaard, por exemplo, não estava certamente comprometido com um reducionismo metafísico acerca da identidade pessoal. Mas certamente concordaria com Parfit quanto ao facto de que ser o mesmo eu ao longo do tempo de uma maneira moralmente e humanamente significativa requer a conectividade psicológica. De facto, ele pensava que isso requer uma tomada de responsabilidade “pelo nosso passado, seguida da afirmação repetida de um (mesmo) compromisso”<sup>213</sup>. Parfit repete sempre que uma coisa é o efeito da visão reducionista nas nossas emoções, outra coisa é o facto de as emoções e atitudes serem justificadas. Claro que a natureza nos fez animais mais preocupados com o nosso futuro do que com o futuro de outros ou não sobreviveríamos. Mas o facto de haver uma explicação evolucionista desta disposição não a torna neutra, nem insusceptível de avaliação, uma vez racionalmente considerada<sup>214</sup>.

Como já acentuei, o ponto comum das conclusões de Parfit acerca de teorias da racionalidade, teorias da identidade pessoal e ética é a defesa da impessoalidade<sup>215</sup>. E o que se ganha segundo Parfit nesta conjunção temática da filosofia prática em torno da impessoalidade é uma maior racionalidade na apreciação daquilo que importa, além de um afastamento justificado do cepticismo moral. Para Parfit, as razões para agir existem; o problema da ética, uma vez explicitado, é saber se algumas dessas razões são razões morais<sup>216</sup>. Parfit defende que há razões para agir que são de facto razões morais e associa o pensamento acerca dessas razões ao progresso na ética. A impessoalidade a que somos racionalmente conduzidos na filosofia prática representa um alargamento da interioridade e daquilo que importa para além do seu lugar de origem, os seres individuados. A intuição de Parfit é que a abordagem moral das situações humanas num mun-

<sup>213</sup> R. M. Adams in Dancy 1997, p. 267.

<sup>214</sup> Cf. Parfit, p. 308.

<sup>215</sup> A impessoalidade do utilitarismo seria assim menos implausível do que a maioria de nós crê. Parfit vê em Buda e Wittgenstein dois precursores da impessoalidade na apreciação moral (Cf. Parfit 1984, Wittgenstein and Buddha, p. 273, e Buddha's view, p.502) e em Buda um reducionista quanto à identidade pessoal. No entanto quanto ao budismo note-se que, se bem que a visão da Identidade Pessoal possa ser de facto concordante com a de Parfit, a relação do reducionismo com a teoria do Fim-Presente é problemática. Como John Merrihew Adams nota bem (cf. Should Ethics Be More Impersonal? in Dancy, 1997, p. 287) o que Parfit valoriza através da teoria do fim presente é a escolha do prazer no presente, e esta escolha é precisamente aquilo que o budismo quer eliminar, ao considerar que a cessação do sofrimento virá com a cessação do desejo.

<sup>216</sup> Parfit 1984, Concluding Chapter.



do com bilhões de pessoas só pode progredir através desta impessoalidade uma vez que a situação que constitui o dado bruto do do pensamento ético se apresenta da seguinte maneira, descrita por T. Nagel: ‘Há tantas pessoas que nem conseguimos bem imaginar quantas e a cada uma a sua vida e o seu bem importam’<sup>217</sup>.

Pretendi, neste artigo, evidenciar alguns pontos que uma teoria da interioridade deve tratar. Eles apareceram sob a forma de uma história unificada ‘daquilo que importa’. Espero que tenha sido claro que várias das teses trazidas à discussão não se conjugam sem mais problemas e que, assim sendo, há um trabalho a fazer no sentido de uma mais perfeita adequação entre elas. Utilizei a expressão de Parfit ‘*what matters*’ ao longo de todo o artigo para unificar os temas da interioridade natural. Estes vão desde a sensibilidade, ao sentimento de agir, à auto-consciência, aos dilemas que a concepção de Identidade Pessoal provoca quando estão em causa a racionalidade e benevolência na acção. Penso ter assim tornado mais visíveis alguns dos problemas filosóficos gerados pelo tratamento teórico da dor e do sofrimento.

### Referências

- Adams, Robert M., *Should Ethics Be More Impersonal?* In Dancy 1997.
- Anscombe, G.E.M, *Intención*, Barcelona, Gedisa, 1991 (ed.orig. 1957).
- Blackburn, Simon, *Has Kant refuted Parfit?* in Dancy 1997.
- Dancy, Jonathan, *Reading Parfit*, Oxford, Blackwell, 1997.
- Dennett, Daniel, *Content and Consciousness*, London, Routledge and Kegan Paul, 1969.
- Dennett, Daniel, *Consciousness Explained*, Boston, Little Brown, 1991.
- Dennett, Daniel, *Kinds of Minds*, New York, Basic Books, 1996.
- Dennett, Daniel, *Brainstorms*, Cambridge, Mass, The MIT Press, 1981.
- Dennett, Daniel, *Why you Can't Make a Computer that Feels Pain* in Dennett 1978.
- Dennett, Daniel, *Conditions of Personhood*, in Dennett, 1978.
- Gil, Fernando, Eu, *Análise* 20, 1998.

<sup>217</sup> Cf. Nagel 1991, ‘Two Standpoints’: Para Nagel, o mais difícil em ética (ou mesmo, finalmente, inexequível) seria a integração dos pontos de vista pessoal e impessoal (necessário para, por exemplo, conceber questões como a imparcialidade, a igualdade e a justiça, que são afinal os outros nomes da impessoalidade de Parfit).

Goodin, Robert, *Utilitarianism as a Public Philosophy*, Cambridge, Cambridge University Press, 1995.

Goodin, Robert, *Utility and the Good*, in Singer 1991.

Guttenplan, Samuel, (ed.), *A Companion to the Philosophy of Mind*, Oxford, Blackwell, 1994.

Humphrey, Nicholas, *Una historia de la mente*, Barcelona, Gedisa, 1995 (edição original *A History of the Mind*, London, Chatto and Windus, 1992).

Kant, Immanuel, *Crítica da Razão Pura* (trad. A. Morujão), Lisboa, Fundação Calouste Gulbenkian, 1985.

Kant, Immanuel, *Crítica da Razão Prática* (trad. Artur Morão), Lisboa, Edições 70, 1994.

Kim, Jaegwon & Sosa, Ernest, *A Companion to Metaphysics*, Oxford, Blackwell, 1995 (artigos “Free Will”, “Persons and Personal Identity”).

Locke, John (1690), *An Essay Concerning Human Understanding*, Oxford, Clarendon, 1979.

Martin, John H., *Neuroanatomía*, Madrid, Prentice Hall, 1998.

Nagel, Thomas, *Mortal Questions*, Cambridge, Cambridge University Press, 1979.

Nagel, *Equality and Partiality*, Oxford, Oxford University Press, 1991.

Nietzche, Friedrich, *A Gaia Ciência*, Guimarães Editores, 1987.

O’Neill, Onora, “*Kantian Ethics*”, in Singer 1991.

Parfit, Derek, *Reasons and Persons*, Oxford, Clarendon Press, 1984.

Petit, Philip, “*Consequentialism*”, in Singer 1991.

Pinto, João Alberto, *Sobre o capítulo ‘Le Sentir’ da ‘Phénoménologie de la Perception’ de M. Merleau-Ponty*, Revista da Faculdade de Letras, Série de Filosofia, nº14, 1997.

Pinto, João Alberto, *Materialismo, Superveniência e Consciência*, Porto, Campo das Letras, 2007. (Originalmente Pinto 1999, Dissertação de Mestrado).

Rawls, John, *Uma Teoria da Justiça*, Lisboa, Presença, 1993.

Rorty, Amélie, *The Identities of Persons*, Berkeley, University of California Press, 1976

Ryle, Gilbert, *The Concept of Mind*, (1949), London, Penguin, 1966

Sass, Louis, Schizophrenia, Self-Consciousness and the Modern Mind, *Journal of Consciousness Studies*, vol. 5, nº5/6, 1998

Searle, John, *Intentionality*, Cambridge, Cambridge University Press, 1983,

Sen, Amartya, *Éthique et Economie*, Paris, PUF, 1993 (ed. original, *On Ethics and Economics*, 1987)

Sen, Amartya & Nussbaum, Martha (eds.), *The Quality of Life*, Oxford, Oxford University Press, 1993.

Singer, Peter (ed.), *A Companion to Ethics*, Oxford, Blackwell, 1991.

Stocker, Michael, *Parfit and the Time of Value*, in Dancy 1997.

Wittgenstein, Ludwig, *Investigações Filosóficas*, Lisboa, Fundação Calouste Gulbenkian, 1987.

Young, R., *The implications of determinism*, in Singer 1991.

RECENSÃO A ANTÓNIO DAMÁSIO, *THE FEELING OF WHAT HAPPENS* (1999)\*

Em *The Feeling of What Happens* (um belo título, que se poderia traduzir por ‘o sentimento daquilo que acontece’), o neurocientista António Damásio apresenta a sua teoria da natureza da consciência. Noutras palavras, Damásio procura imaginar e explicar como é possível que os nossos pensamentos e sentimentos sejam do mesmo mundo que as células e moléculas que constituem o nosso cérebro e como é possível que não apenas pensemos mas também, simultaneamente, nos sintamos pensar.

No início da obra (Chapter 1, *Stepping into the Light*) Damásio utiliza a metáfora da subida ao palco, para debaixo das luzes, de modo a delinear o seu objecto teórico. É essa subida ao palco para debaixo das luzes, a transposição de um limiar, que distingue a consciência relativamente a outros estados internos de informação e que faz dela a ‘revelação da existência’ (p.315) que se pretende explicar em *The Feeling of What Happens*. *The Feeling of What Happens* é portanto uma obra dedicada ao problema da consciência, um problema que constitui, neste momento, um foco da atenção de cientistas cognitivos e de filósofos. No entanto nem sempre foi assim.

Na introdução de uma antologia recente dedicada à questão<sup>218</sup>, os autores fazem notar a estranheza da história da investigação científica e filosófica da consciência nos últimos cem anos: «Talvez não exista outro fenómeno em relação ao qual o pêndulo da credibilidade intelectual tenha oscilado a tais extremos (...) no espaço de aproximadamente um século, fazendo-o passar por tempos de exaltação e por tempos em que foi encarado como um tabu»<sup>219</sup>. Nem o afastamento do behaviorismo na filosofia e na psicologia dos anos 60, um afastamento que produziu o cognitivismo

\* A. Damásio, *The Feeling of What Happens - Body and Emotion in the Making of Consciousness*, New York, Harcourt Brace, 1999.

<sup>218</sup> Block, N., Flanagan, O & Güzeldere, G. (eds), *The Nature of Consciousness*, Cambridge MA, MIT Press, 1997.

<sup>219</sup> G. Guzeldere, Preface, in Block, Flanagan & Guzeldere 1997: xi.

psicológico<sup>220</sup> e a sua justificação filosófica, o funcionalismo<sup>221</sup>, reintroduziu a consciência na teoria da mente. De facto, apenas nos anos 90 se deu o definitivo retorno à respeitabilidade do problema da consciência. A prova desta recente respeitabilidade é a atenção recebida pelo problema nas neurociências - atenção da qual o livro de Damásio é um exemplo - na neuropsicologia, na psicologia cognitiva e na filosofia da mente. Nesta última, a consciência parece aliás ter-se tornado o tópico da década de noventa<sup>222</sup>.

Entre os filósofos, a tentativa de abordagem da consciência a partir de um ponto de vista naturalista conduz frequentemente à proclamação de uma última fronteira. A consciência ou experiência fenomenal seria o problema limite para uma metafísica materialista devido ao carácter especial dos *qualia*<sup>223</sup>. O carácter especial dos *qualia* está precisamente associado ao sentimento de si tratado por Damásio em *The Feeling of What Happens*, i.e. ao facto de - pelo menos para algumas entidades, as entidades conscientes - ser como alguma coisa ser. O problema da consciência é por isso considerado por muitos filósofos contemporâneos como um problema filosófico especial, um *hard problem*<sup>224</sup> (problema difícil) relativamente aos restantes problemas da cognição e um problema responsável pela existência de um *explanatory gap*<sup>225</sup> (abismo explicativo).

Noutras disciplinas, uma grande diversidade (ou mesmo divergência) caracteriza as abordagens empírico-especulativas recentes. De um

<sup>220</sup> De um modo esquemático, pode-se considerar que o cognitivismo é o estudo dos fenómenos mentais à imagem dos programas de computador. De acordo com a ideia cognitivista, a mente está para o cérebro como o software para o hardware num computador, i.e. a mente é um programa ou um conjunto de programas que correm no cérebro.

<sup>221</sup> De novo de um modo esquemático, o funcionalismo pode ser considerado como uma oposição à identificação estrita entre estados mentais e estados físicos. De acordo com o funcionalismo os mesmos estados mentais (por exemplo 'dor', ou um cálculo) podem ser implementados ou realizados em diferentes substratos físicos. O funcionalismo é assim de certo modo um dualismo materialista, uma aposta na independência da cognição relativamente à física. Embora as mentes tenham que ser incorporadas não têm necessariamente que ser incorporadas da mesma maneira.

<sup>222</sup> Alguns exemplos de obras marcantes, sobre a consciência, escritas por filósofos, nos anos 90 são: Daniel Dennett 1991, *Consciousness Explained*, Colin McGinn 1991, *The Problem of Consciousness*, William Seager 1991, *The Metaphysics of Consciousness*, Owen Flanagan 1992, *Consciousness Reconsidered*, John Searle 1992, *The Rediscovery of Mind*, Fred Dretske 1993, *Conscious Experience*, Paul Churchland 1995, *The Engine of Reason*, David Chalmers, 1996, *The Conscious Mind*, John Searle 1997, *The Mystery of Consciousness*. A própria antologia referida, *The Nature of Consciousness* (Block, Flanagan & Guzeldere 1997), na qual aparecem vários textos provenientes ou derivados das obras referidas, constitui uma referência marcante.

<sup>223</sup> 'Qualia' é o termo utilizado na filosofia da mente para nomear estados mentais sentidos, como cores vistas e sons ouvidos. O termo é frequentemente utilizado para marcar a oposição entre estes estados mentais sentidos e os inúmeros processos cognitivos que podem ser inconscientes.

<sup>224</sup> A expressão é de David Chalmers (cf. Chalmers 1996) e pretende contrastar o problema da consciência (ou experiência, ou subjectividade sentida) com os restantes problemas da cognição, que são problemas relativos a habilidades e funções cognitivas, como a discriminação e categorização, o processamento e integração de informação, o controlo do comportamento, a capacidade de reportar verbalmente estados internos, a distinção entre a vigília e o sono, etc. De acordo com Chalmers, a distinção entre os problemas fáceis e o problema difícil é (uma distinção) conceptual, e não uma pretensão segundo a qual a consciência e a cognição não teriam nada em comum.

<sup>225</sup> A expressão foi introduzida pelo filósofo Joseph Levine no artigo "Materialism and Qualia: The explanatory gap", *Pacific Philosophical Quarterly*, 64, 1983. Após a nomeação do abismo, resta saber se a existência deste representa um problema epistemológico ou um problema ontológico. O próprio J. Levine pensa que a consciência não coloca um problema ontológico ao materialismo mas apenas um problema epistemológico.

ponto de vista neurobiológico, F. Crick e C. Koch<sup>226</sup>, por exemplo, consideram que o correlato neuronal da consciência são oscilações de 40Hz que têm um papel importante na ligação (*binding*) num todo dos vários tipos de informação processados no cérebro<sup>227</sup>. De um ponto de vista funcional, o psicólogo cognitivo Bernard Baars<sup>228</sup> define a consciência como o espaço de trabalho global num sistema de processadores de informação inteligentes e distribuídos. Quando os processadores acedem ao espaço de trabalho global eles difundem ou emitem (*broadcast*) para o sistema na sua totalidade os conteúdos processados. Aquilo de que o sistema é consciente será assim aquilo que é globalmente acessível, sendo os conteúdos da consciência os *conteúdos do espaço de trabalho global*<sup>229</sup>. O efeito de difusão faz com que a informação deixe de estar compartimentada, sendo assim susceptível de um uso generalizado e não apenas específico pelo sistema cognitivo. Sem esta acessibilidade mútua dos conteúdos no espaço de trabalho global a informação processada não seria nada para o (seu) sujeito.

Vindo de fora das áreas biológicas e psicológicas, o físico e matemático R. Penrose<sup>230</sup> defende que a teoria requerida para a explicação da consciência (o seu nível e o seu tipo) é totalmente diferentes dos até agora mencionados. Quanto ao tipo de teoria, a explicação da consciência envolverá de acordo com Penrose a mecânica quântica<sup>231</sup> e a lógica matemática

<sup>226</sup> F. Crick & C. Koch, *Towards a Neurobiological Theory of Consciousness, Seminars in the Neurosciences*, 2, 1990.

<sup>227</sup> Antes de mais, o correlato neuronal da consciência aparece como solução para um determinado problema, o *binding problem*. Este problema consiste no seguinte: nós temos uma experiência visual unificada de objectos, e, no entanto, os traços dos objectos vistos - traços como a cor, a forma e o movimento - são tratados por regiões diferentes e separadas entre si do córtex visual. O problema que se coloca consiste em saber como é possível que do tratamento desunificado se passe a uma experiência consciente unificada do objecto. A estratégia de Crick consiste em perguntar como é que os neurónios envolvidos numa particular percepção são temporariamente activados como unidade. A hipótese dos 40 Hz sugere como solução o disparo sincronizado de neurónios numa determinada frequência. Este disparo sincronizado seria assim o procurado correlato neuronal da consciência. Como o tálamo - o tálamo é uma estrutura sub-cortical, do diencéfalo, essencial para a [re] transmissão de informação para os hemisférios cerebrais - parece ter um papel central na consciência [esta parece estar dependente de circuitos ligando o tálamo ao córtex] a definição de consciência [ou melhor do correlato neuronal da consciência] sugerida é a seguinte: o correlato neuronal da consciência é o disparo sincronizado, numa determinada frequência, de neurónios dos circuitos que ligam o tálamo ao córtex.

<sup>228</sup> B. Baars, *A Cognitive Theory of Consciousness*, Cambridge, Cambridge University Press, 1988.

<sup>229</sup> O espaço de trabalho é global não apenas no sentido funcional -i.e. é um 'espaço' onde tudo pode entrar em contacto com tudo- como também no sentido anatómico de ser distribuído pelo córtex, envolvendo mesmo outras regiões do cérebro, o que significa por exemplo a utilização do mesmo hardware utilizado pela memória. Baars sugere que o ERTAS é o 'correlato neuronal' do seu modelo funcional. ERTAS é a sigla de *extended reticular-thalamic activating system*. A sugestão de Baars apoia-se no facto de a estimulação numa determinada área do tronco cerebral se estender ao córtex [o animal fica alerta, orienta-se, desperta se estiver a dormir]. Essa área do tronco cerebral [a formação reticular, sistema de activação reticular] veio a ser considerada apenas parte de um sistema ascendente mais amplo que integra o tálamo e o córtex. A sigla sublinha portanto que o sistema, embora se estenda, está centrado no tálamo.

<sup>230</sup> Cf. Penrose 1989, *The Emperor's New Mind - Concerning Computers, Minds and the Laws of Physics*, Oxford, Oxford University Press e Penrose 1994, *Shadows of the Mind: A Search for the Missing Science of Consciousness*, Oxford, Oxford University Press.

<sup>231</sup> A relação entre a física quântica e a consciência estabelece-se devido à importância que a consciência pode ter nas próprias fundações da teoria quântica: o colapso da função de onda seria provocado, de acordo com as interpretações mais comuns da teoria quântica, pela medição, que supõe a consciência. A consciência determinaria indeterminações quânticas.

(nomeadamente o teorema de Gödel), e não a neuroanatomia, a neurofisiologia ou a psicologia cognitiva. Quanto ao nível ou escala da teoria, o nível dos neurónios não é o nível apropriado para a explicação da consciência. Segundo Penrose os neurónios apenas amplificam o nível mais baixo no qual reside a explicação física última da mente consciente. De acordo com Penrose os seres humanos conscientes são capazes de habilidades cognitivas que os computadores não conseguem simular. Segue-se que o entendimento humano não pode ser uma actividade algorítmica. O teorema de Gödel mostra que existem verdades em sistemas matemáticos que não podem ser demonstradas nesses sistemas. O pensamento humano consciente nem poderia assim ser simulado em computador. Ora, os neurónios são computáveis (i.e. as suas características podem ser simuladas em computador). Como a consciência tem características não computáveis, os neurónios não podem explicar a consciência. Para explicar a consciência é necessária qualquer coisa não computável, efeitos que se encontram ao nível sub-neuronal, nomeadamente no citoesqueleto das células, nos microtúbulos dos neurónios. Compreender estes efeitos requererá uma revolução na física e Penrose alude a uma teoria da gravidade quântica. A sua ideia acerca da consciência é portanto que esta poderia ser uma *manifestação dos estados quânticos internos (do citoesqueleto) dos neurónios e do envolvimento destes estados na interacção entre os níveis quântico e clássico de actividade*.

É neste contexto, de um grande interesse mas ao mesmo tempo de uma grande divergência relativamente ao problema da consciência, que se insere o trabalho de António Damásio, um trabalho que obteve uma considerável repercussão exterior aos meios científicos e académicos devido à obra *O Erro de Descartes*.

A teoria defendida por Damásio em *The Feeling of What Happens* pretende acrescentar algo de novo a abordagens como as de Crick e Kock, Baars ou Penrose. De acordo com Damásio, o problema da consciência é um problema de dupla face: ele é por um lado o problema do filme dentro da cabeça (o fluxo fenomenológico<sup>232</sup>, o decorrer temporal da consciência e dos seus conteúdos) e por outro o problema da pertença do filme (o sentido do *self*, a reportação da fenomenologia a uma entidade que se sente ser). É a este último problema que Damásio se dedica em *The Feeling of What Happens*. O aspecto do problema da consciência visado é portanto a subjectividade (do *self*<sup>233</sup>, do eu), um aspecto diferentes dos visados por

<sup>232</sup> Em filosofia é frequente o uso desta expressão para nomear a corrente da consciência.

<sup>233</sup> Utiliza-se aqui o termo *self*, como o próprio Damásio faz, como uma forma de nomear a distinção entre si e não-si que não é ainda consciente ou linguística. O *self* não se identifica portanto com o eu: ele existe por exemplo em seres vivos não humanos, como uma representação contínua e constantemente modificada do corpo próprio, sem que se possa afirmar que nesses seres existe um eu. Existe no entanto alguma coisa à qual algum saber pode ser atribuído.

Crick e Kock, Baars ou Penrose. A hipótese explorada na obra é que a subjectividade tem raízes corpóreas estratificadas e diversificadas, i.e. que o cérebro tem o corpo em mente de diversas maneiras e a diferentes níveis<sup>234</sup>.

Como o sub-título do livro indicia, o trabalho de Damásio é em grande parte dedicado a explorar a ligação entre a consciência, as emoções e o corpo próprio, simultaneamente ao nível neurobiológico, ao nível das arquitecturas cognitivas e ao nível fenomenológico.

Já em *O Erro de Descartes* a busca do enraizamento corpóreo da subjectividade conduziu Damásio a uma teoria da emocionalização da mente. Se um dos sentidos do erro cometido por Descartes é o dualismo mente/corpo, o outro sentido é a intelectualização do mental, a separação entre uma razão que seria pura e as emoções. Pelo contrário, Damásio sublinha a importância das emoções nos fenómenos de racionalidade prática como o planeamento das acções próprias e as decisões, analisando nomeadamente as características de patologias neuropsicológicas (como as provocadas por lesões nos lobos frontais<sup>235</sup>) ligadas à ausência do funcional normal das emoções. Uma vez que na ausência do normal funcionamento das emoções o mecanismo da racionalidade prática é perturbado, Damásio defende que a natureza construiu o 'dispositivo cerebral para a racionalidade' não em cima e por cima dos mecanismos biológicos de regulação do corpo, aos quais as emoções se associam, mas a partir deles e com eles. Este é um dos aspectos da corporeidade da subjectividade<sup>236</sup>.

A corporeidade da subjectividade conduz ainda Damásio a uma teoria neurobiológica mais ou menos localizacionista do *self* e do eu. Existem determinadas regiões cerebrais ligadas à constituição do *self* e do eu neuronais (i.e. à representação do *self* e do eu no cérebro do organismo), nomeadamente núcleos do tronco cerebral, o hipotálamo, estruturas basais, o córtex da ínsula e o córtex somato-sensorial<sup>237</sup>. Estas estruturas constituem

<sup>234</sup> O primeiro aspecto do problema da consciência poderia ser segundo o próprio Damásio tratado por meio de modelos de consciência como os apresentados por B. Baars (o modelo do espaço de trabalho global atrás referido) ou por D. Dennett (o Modelo dos Esboços Múltiplos apresentado em Dennett 1991, *Consciousness Explained*).

<sup>235</sup> Cf. o caso de Phineas Gage, exemplar em Damásio 1994, *O Erro de Descartes* (Lisboa, Europa-América), Capítulo 1. Segundo Damásio, a lesão cerebral de Phineas Gage provoca uma alteração do papel regulador das emoções na racionalidade prática. A racionalidade prática é a capacidade que cada um de nós tem de planear e decidir as suas próprias acções. Ora, se Phineas Gage perde a capacidade de planear racionalmente o seu próprio futuro é porque sofre uma importante alteração de carácter, tornando-se nomeadamente caprichoso e obsceno quando antes fora equilibrado e ponderado. Não se pode no entanto afirmar que a lesão tenha provocado uma diminuição da sua destreza física ou da sua capacidade intelectual: o que foi alterado foi a dimensão pessoal e social do raciocínio. Existe assim uma dissociação entre a inteligência (por hipótese intacta) e outros mecanismos não estritamente intelectuais influentes na racionalidade prática (por hipótese as emoções).

<sup>236</sup> Este tema do 'body-minded brain', i.e. do cérebro que tem o corpo em mente, é o tema central de Damásio 1999. No entanto começara já a ser explorado no Capítulo 10 de Damásio 1994. A ideia central é que o corpo contribui com um conteúdo essencial para o funcionamento da mente normal.

<sup>237</sup> Os núcleos do tronco cerebral regulam estados do corpo e mapeiam sinais corporais. O hipotálamo intervém na representação 'actual' do corpo mantendo um registo do estado de várias dimensões do meio interno (como a circulação de nutrientes, a concentração iónica, o PH, etc). O córtex da ínsula e o córtex somato-sensorial mantêm representações do corpo. Damásio liga estas regiões ao que chama proto-self. Cf Damásio 1999: 155.



a base neuroanatómica de uma distinção entre si e não-si. Tal distinção não é no entanto necessariamente consciente ou pessoal em todos os níveis envolvidos.

Uma das intenções de Damásio em *The Feeling of What Happens* é aliás distinguir as fundações neurobiológicas do *self*<sup>238</sup> das 'localizações' do eu autobiográfico da identidade pessoal<sup>239</sup>. O sentimento de si tem assim em cada um de nós uma versão nuclear e uma versão mais sofisticada, ligada à consciência alargada (*extended consciousness*) e ao eu. O eu está associado a uma reunião das memórias daquilo que acontece. Só esta reunião de memórias permite que um eu se mantenha e que se reconheça como o mesmo ao longo do tempo, e a uma consciência de si propriamente pessoal. O cérebro tem, assim, como uma das suas funções, manter uma narrativa constante e contínua na qual o eu é a personagem principal. Damásio considera que o eu da identidade pessoal é, relativamente ao *self*, uma realidade de nível funcional e cortical mais elevado. É no entanto sobre o *self* que vem estabelecer-se o eu e também a consciência alargada. É a essa transição entre representações de si ainda não conscientes e a consciência propriamente dita, i.e. a consciência como aparição, que é dedicada toda a parte III do livro (Chapter 5, *The Organism and the Object*, Chapter 6, *The Making of Core Consciousness*, Chapter 7, *Extended Consciousness*, Chapter 8, *The Neurology of Consciousness*). A transição não é no entanto uma definitiva passagem para o outro lado, para uma consciência pura, que seria um saber de segunda ordem, um saber que se sabe *neutro*. Pelo contrário a consciência não se destaca de um fundo que acompanha a vida consciente e que é constituído pelas representações de si a vários níveis mantidas pelo cérebro. Damásio sugere portanto que o cérebro é uma audiência cativa de representações de si (i.e. do corpo próprio, da história biográfica deste), representações que ele próprio constantemente mantém, e que esse facto é essencial para a aparição da aparição, i.e. para o surgimento da consciência como um sentido de si próprio no acto de conhecer.

A exploração da importância do corpo (mais especificamente da representação do corpo no e pelo cérebro desse corpo) na constituição da consciência conduz Damásio finalmente a uma tese central, explicitada no último capítulo (Chapter 11, *Under the Light*). Segundo Damásio a consciência inicia-se com um sentimento, um sentimento que é um sentimento de saber e um sentimento de si. O segredo da constituição da consciência

<sup>238</sup> De facto Damásio distingue *proto-self*, *self* e eu autobiográfico. Não é no entanto demasiado violento do ponto de vista da interpretação sublinhar apenas a distinção entre um *self* que não tem que ser consciente nem linguístico e o eu.

<sup>239</sup> Determinadas estruturas – como córtices sensoriais e temporais, o hipocampo e os córtices pré-frontais (cf. p.157) – não são, segundo Damásio, necessárias para implementar o *self* não consciente nem linguístico, estando no entanto associadas ao eu autobiográfico.

reside, nas palavras de Damásio, no facto de «o estabelecimento de uma relação entre qualquer objecto e o organismo se tornar no sentimento de um sentimento. A misteriosa perspectiva em primeira pessoa da consciência consiste em (...) informação expressa como sentimento» (p.313). Este sentimento é segundo Damásio gerado pela maneira como o cérebro sente a carne do corpo próprio depois de o cérebro ter agido sobre esta. É com a perda da inocência acerca da existência que se paga a possibilidade de uma existência melhor ou mais rica: «O sentimento daquilo que acontece é a resposta a uma questão que nunca pusemos, e é também a moeda num trato faustiano que nunca poderíamos ter negociado. A Natureza fê-lo por nós» (p.316).

Escrita por um neurocientista, esta é no entanto também uma obra filosófica que encara directamente o problema da subjectividade. Damásio defende que uma subjectividade com raízes corpóreas acompanha todo o nosso pensamento consciente e se traduz no facto de a nossa vida mental não ser apenas cognição e cálculo mas também sentimento de si e revelação da existência.



## *MODELS OF UNDERSTANDING – MINDS AND MACHINES*

**Abstract:** This article is about minds and machines, or, to be more precise, about cognitivism as a model for understanding the mind. The sub-title is a quotation from Hilary Putnam's famous 1960 paper "Minds and Machines", a sort of manifesto for functionalism in the philosophy of mind. Although Putnam himself came to reject functionalism, to a large extent because he now thinks that reason cannot be naturalized, functionalism is central to what I will call 'cognitivism'. I will be considering cognitivism in a historical, rather than argumentative, way. Basically I want to consider two aspects: (i) some ideas of authors (such as H. Putnam, J. Searle, A. Turing, H. Simon and D. Dennett) which I think may help us take a stand in an ongoing discussion concerning the relevancy of cognitivism in thinking about the nature of mind, (ii) a noteworthy consequence of cognitivism: the fact that although cognitivism uses computational machines as a model, or metaphor, for understanding the mind, this results in weakening, and eventually doing away with, the natural-artificial dichotomy, which one could assume was presupposed by that model.

### **1. Putnam's functionalism plus Fodor's Language of Thought**

Hilary Putnam's formulation of functionalism, in the 60s, in papers such as "Minds and Machines" and "The Nature of Mental States", is a historical landmark in the philosophy of mind. Putnam's basic idea is that mental states (contrary to what Identity Theory materialists defended) are not brain states, but rather functional states implemented in the brain (although they could be implemented in another hardware – the idea of multiple realizability of mental states goes along with functionalism). In fact, it is precisely because mental states can have diverse physical realizations that they should not be identified with brain states. What Putnam suggests then, is that mental states of beings such as ourselves, are in a similar relation to neurophysiological states as logical (functional) states of computational machines are to the physical states of those machines ('the mind is to the brain, as software to hardware'). Putnam intends, with this position, to 'dissolve' the mind-body problem. In other words, he wants to show that the mind-body problem is not a genuine theoretical problem, but merely linguistic and logical in nature. According to Putnam, this same problem would arise for any cognitive system capable of self-monitoring

and of producing self-descriptions, if there were, in such a system, as there is in us, an asymmetry between the access to the logical level of the system, the program level, and the access to the physical level of the system. The program level is the level of the system in respect to which the system is incorrigible – as we can see by the status of statements such as ‘I know I feel pain’, ‘I know I think that p’. There is no such incorrigibility in the system’s access to its physical level (each one of us must learn about his or her own brain – the fact that our brain causes our mind does not by itself turn a person into a neuroscientist). Putnam’s functionalism makes us look at the (supposed) mind-body problem as related to this asymmetry, and not to some unique nature of human subjective experience.

Functionalism, together with the idea according to which the functional level is an autonomous, symbolic, representational-computational level (elaborated for instance by Jerry Fodor in his 1975 book in terms of a Language of Thought Hypothesis, which may be summarized as ‘No representations, no computations, no mind’) is what constitutes what I am calling cognitivism.

## **2. Strengths and weaknesses of the model**

What is wrong and what is right with this model for understanding the mind? How important is it today in the philosophy of cognitive science? How committed is the model – which makes us look at the nature of minds through the lens of computational machines – with a natural-artificial dichotomy? The first obvious comment is that this picture of the nature of mind does not immediately call our attention to something that may seem essential to our specific kind of mind: consciousness. On the contrary, it can very easily allow for the mental character of that which has never been, and will never be, conscious. From a neo-cartesian point of view such as that of John Searle, for instance, this is almost heretic, and anyway, an utterly unjustified position. Since the 60s much has been going on in the philosophy of mind and in cognitive science, and I will try to describe some of that history here in order to address the questions I have formulated above. John Searle’s work from the 80s and 90s is very helpful here: not only does he point directly at what is absent in the cognitivist model of the mind (namely, consciousness) but also he clearly sees the close relation between cognitivism and a specific discipline within cognitive science, Artificial Intelligence, and the significance of this relation. In fact, with his celebrated Chinese Room thought experiment, Searle intends not only to criticize cognitivism but also to argue in favour of the impossibility of what he calls Strong AI (Strong AI, in Searle’s own definition,

is the idea according to which AI is concerned not only with simulation of cognition but eventually with the real thing – any physical system that runs the right program will have a mind; a system's being intelligent, or even conscious, depends only on the right kind of functional organization, and so on programming, and not on hardware). As far as that 'condition of mentality' may be formulated independently of the system's physical build-up and biological origins, it is conceivable that physical systems other than humans, namely artificial systems, will be, for the exact same reasons as humans, intelligent and conscious.

But what does AI, as a discipline, have to do with cognitivism? Since the beginnings of AI, philosophers have taken an interest in it, a very natural interest, given the fact that philosophers have always been interested in the nature of thought, and in its relation with the physical world. This interest produced extreme positions about AI, ranging from the proclamation of an impossibility in principle of non-natural intelligence and consciousness, to the conviction that through AI a more general and abstract conception of the nature of intelligence would be reached, one which would allow us to see human beings and all intelligent beings – natural or non-natural – as examples of one same general phenomenon.

In contemporary philosophy of mind John Searle is a well known critic of both cognitivism and Strong AI (whereas for instance philosophers such as David Chalmers and Daniel Dennett defend both). The reason Searle's Chinese Room thought experiment is always brought in as a reference, is that it can be seen as cooling down the excitement about cognitivism and AI. So, it is especially interesting to notice that in this thought experiment the critique of cognitivism takes the form of an intuitive test of the difference between genuine mentality and simulated, merely attributed, mentality.

Before coming back to the Chinese Room, which is, as I said, an inevitable reference for thinking about cognitivism as a model of the mind, and because I pointed out a close connection between cognitivism and Strong AI, I want to make a few remarks about the contributions of a philosopher, who has, in a way, preceded Searle, as the official philosophical critic of AI. That philosopher is Hubert Dreyfus, and some of his ideas were passed on more or less explicitly to Searle (although Dreyfus is generally closer to European phenomenology than Searle). Dreyfus noticed quite early the overlap of interests of AI and philosophy and the result were some controversial works such as *Alchemy And Artificial Intelligence* (1965) and *What Computers Can't Do* (1972). There he criticized the claims of AI, and insisted on the relation, which seemed obvious to him, between the work then being developed in AI, based on a conception of mind as a symbolic

system and an idea of intelligence as problem-solving and the rationalist and intellectualist tradition in philosophy. Dreyfus point was that AI was repeating the intellectualist errors in the conception of mind and intelligence which had been pointed out by philosophers such as Heidegger, Wittgenstein and Merleau-Ponty. His rejection of a conception of mind as symbolic representations of the world, and of a conception of intelligence as rule-governed problem-solving was based on the conviction that these conceptions excluded important and basic parts of the mental. Those 'excluded parts' were for instance, body movements and pattern recognition, which, according to Dreyfus, underlie the possibility of the explicit skills involved in representing and problem solving.

Dreyfus' writings presented a sober perspective on some exaggerated forecasts that had accompanied the beginnings of AI. Although some of Dreyfus's own forecasts have been simply proved wrong (he was, for instance, convinced that a computer would never win a chess game with a human), what Dreyfus wanted to criticize was the aprioristic assumption according to which mind consisted in symbolic representations and rule-governed manipulation of such representations. According to this intellectualist view, even a cognitive skill such as perception would ultimately consist in problem-solving by means of application of rules. Now this conception was totally oblivious of something which, in Dreyfus's eyes, played a fundamental role in cognition: background knowledge or common sense. Background knowledge is not knowledge of facts but, in the case of people, what they know without knowing that they know it, something that they never learned but know how to act upon (as, for example, that people move more easily forwards than backwards, or that if one spills water on a towel on a table it will eventually get to the legs underneath the table). Other characteristics of human cognition which were absent from these early attempts at simulation, and which Dreyfus pointed out, were fringe consciousness, the tolerance of ambiguity, a proper body in the world that organizes and unifies the experience of objects and subjective impressions, the capacity for boredom, fatigue and loss of motivation, intentions and interests that guide the way human subjects confront situations in the world, making it the case that not all the facts in the world are equally relevant at a given instant, and so differentiating 'the world', etc. Some of Dreyfus's criticisms have been since then simply incorporated in the development of AI and Dreyfus himself was quick to admit the proximity between the principles of connexionism and the anti-intellectualist tradition in philosophy.

Let's then move on to Searle, who is perhaps the most well-known critic of the limits of the cognitivist model of the mind. Searle likes to say

that cognitive science is an exciting field of research based on a conceptual error concerning the nature of mind – cognitivism is the general name for that error and the Chinese Room thought experiment was the first attack Searle launched against it. “Minds, Brains and Programs”, the paper where the argument was first presented, appeared in the journal *Behavioral and Brain Sciences* in 1980 and the Chinese Room thought experiment hasn’t moved out of the philosopher of mind’s tool-kit since then. The Chinese Room consists of the following: somebody, who doesn’t speak any Chinese, is locked inside a room, where there are Chinese symbols in boxes. This person also has a book of instructions in English, which tells her how to combine and transform symbol sequences, in order to send them out of the room when other Chinese symbols are introduced in the room through a small window. The person inside the room knows nothing about this, but the people outside call the symbols which go in ‘questions’ and the symbols coming out ‘answers’. Therefore, from the perspective of those people outside the room, verbally interacting with it, the system speaks Chinese. Thus, the system behaves intelligently, passes the “Turing Test” – although the person inside knows very well she does not understand a word of Chinese. Searle claims that the Chinese Rooms makes the possibility of a system that has ‘attributed intentionality’ but no ‘intrinsic intentionality’ or ‘genuine semantics’ obvious.

It is not easy to say what exactly the argument is supposed to prove. In fact, as Searle admits, the Chinese Room is more of a parable than an argument. If it was presented as an argument it would go from the premisses ‘Programs are syntactic’ ‘Syntax is not sufficient for semantics’ ‘Minds have semantic content’ to the conclusions ‘Implementing a program is not sufficient for mind’, ‘Strong AI is unjustified’. Anyway, the Chinese Room is supposed to show that mind is not a program and therefore appropriate programming could never give mind to a system, since formal properties do not constitute genuine intentionality. Searle always stresses the fact that his argument has nothing to do with a specific stage of evolution of technology, but concerns only conceptual principles: cognitivism is wrong in considering that formal properties *per se* would be sufficient for mind (this position would go with the defense of Strong AI). For Searle, the essence of mind is consciousness, and the existence of consciousness is a biological fact. Although in the initial formulation of the Chinese Room, Searle evokes semantics and not consciousness (‘syntax is not sufficient for semantics’, he says), the reason why Searle thinks one cannot speak of syntactic mental processes without speaking of semantics is the fact that he thinks that semantics and consciousness are intimately connected. Basically, Searle thinks that one cannot legitimately consider



anything as mental if not for its relation (current or potential) with consciousness. This is what Searle calls the Connection Principle, which in fact he uses in another argument against cognitivism, formulated in the 90s (in a 1992 book called *The Rediscovery of Mind*). The line of argument goes – roughly – like this: Syntax is not a physical property; Cognitivism assumes that physical events are syntactic; cognitivism is based on a fallacy. Searle calls this fallacy the homunculus fallacy – what he means is that cognitivist explanations of mental phenomena treat the brain as if there were some agent inside it, doing symbol manipulation and computations. According to Searle, symbols and computations are not intrinsic features of the world. Yet cognitivism, evoking syntactic properties to explain mental phenomena, totally overlooks this, and treats properties which are there only for an observer as if they were natural properties. But if syntax is not a feature of the natural world but an interpretation of physical events, dependent on an observer, syntactic descriptions of cognitive systems, assigned relative to observers, simply cannot do any explanatory work. And so cognitivism is flawed.

Several aspects of Searle's 'Critique of Cognitive Reason', as he calls it, may be unified by noticing that for Searle, the Connection Principle is a basic principle which should be used in thinking about the mind. It states that one can conceive something which is currently unconscious as mental only as far as one can think of it as a possible content of consciousness. Only this – being potentially conscious – distinguishes intrinsic intentionality from *ersatz* intentionality (if not, how could we even start distinguishing a neuron from a non-evoked memory, calling one mental and the other not?). So, like Descartes, although from a materialist point of view, Searle believes that consciousness is the essence of the mind. Now, for Searle, consciousness is a physical property of the brain, characterized by ontological subjectivity. Something is ontologically subjective if we cannot describe it from a third person point of view (which is what we try to do when we investigate for instance the neurophysiology of consciousness). The problem that the existence of mind in the natural world poses is connected with its ontological subjectivity. It is because of ontological subjectivity that consciousness, although a physical property of the brain, is irreducible to any other physical feature. This ontological sense of subjectivity (the *idea* according to which this world is such that there are irreducibly subjective elements in it) cannot be mistaken for an epistemological sense of subjectivity (which concerns preconceptions that are supposed to be eliminated in the pursuit of objectivity which is part of the spirit of science). If we accept this distinction, we will have to know next how we can have an objective conception of the ontologically

subjective facts of consciousness. Searle's answer is biological naturalism: the idea that consciousness is a biological feature of the brain, human and of other animals, and an emergent property (as liquidity). Searle formulates the question in these terms because he thinks that many materialists are wrong when they think that without admitting reduction (something Searle himself does not admit) one necessarily accepts dualism (and for Searle that does not have that to be the case).

Let's go back to the Chinese Room, which, after all, opens the door to Searle's discovery of the irreducibly subjective element of this world. Searle uses it as an instrument to criticize the artificiality of simulations of cognition and the whole critique falls on the distinction between original, genuine intentionality on the one hand, and attributed intentionality on the other. But exactly how are we supposed to distinguish between genuine intentionality and merely attributed intentionality? Searle evokes the causal power some physical systems (namely human brains) have, while others don't, as well as the Connection Principle. But from a rhetorical point of view what Searle wants from us, is that we identify with the human being inside the room, manipulating the symbols and lacking any understanding. The basic reply to Searle is the so called systems reply: it is wrong to attribute understanding to the executor of the program (incidentally a human). Understanding is supposed to characterize the system as a whole, and that includes the pieces of paper with the rules and symbols. By the way, it is worth considering here that our neurons don't have any understanding of the language - Portuguese or English, or Swedish,... - that we speak, either. Yet, we wouldn't doubt that we, the global system, do in fact understand Portuguese or English or Swedish - the lack of genuine intentionality in our neurons is no proof of a lack of genuine understanding in us. It isn't fair, though, to take all this to mean that Searle thinks that the distinction between genuine intentionality and attributed intentionality corresponds to the distinction between 'natural' and 'artificial' - Searle is no simplistic critic of AI. He does have a problem here, though: his appeal to intuition, on which in fact the whole critique of cognitivism rests. I will now look for another approach to these same issues (cognitivism, strong AI, intentionality, the meaning of 'natural' and 'artificial') which does not share that problem.

### **3. Is there only one way for intelligence to be artificial?**

Until now I have been considering only philosophers, now I want to make a detour that will take us outside the discipline. Notice that the model of mind I have been considering makes it very clear how intelligence

could be artificial: through the lens of the materialist and dualist conception of cognitivism, we see mind as implemented software, a software which can be implemented in a hardware other than the biological. That's how intelligence could be artificial. Now I want to bring in another idea concerning the nature of the natural / artificial relation, an idea coming from Herbert Simon, one of the founders of AI as a discipline, and put forward for instance in his book *The Sciences of the Artificial* (1969). This is an idea that is shared by D. Dennett, who, as I said before, unlike Searle, defends cognitivism and Strong AI. Simon thinks that the sciences of the artificial by no means involve a move back, or forth – away, anyway – from natural sciences. The artificial and the natural are not, according to Simon, two kingdoms but two points of view, which do not stand opposed to one another. Everything that is artificial (and that, to Simon, is everything that is a functional/adaptive device, and should be assessed through norms of rational functioning) is also natural (that is, it is ultimately an object for physical explanation). According to H. Simon, what distinguishes the artificial, then, is something other than the existence of a distinct realm of entities. What characterizes the point of view of the artificial is the fact that it aims at systems in their status of interfaces of an interior and an exterior, thus creating the (new, in the scope of natural sciences) question of the rationality or adaptation of these interfaces to the environment. In other words, for Simon a science of the artificial is a science of the artificial because it deals with teleology, with the global behavior of systems and the purposes of that behavior. Purposes relate the interior and the exterior, independently of the material build-up of the systems. Both the interior of the system and its exterior continue to belong to the natural sciences it is the interface that is specifically artificial. In fact, Simon also defends that in some way certain natural organizations are in this sense artificial, at least as far as a natural system (biological) can be analyzed according to these parameters. And in *The Sciences of the Artificial* H. Simon does include psychology (defined as the science of behaving systems) in the sciences of the artificial: the psychology of a system is a science of the artificial in contrast, namely, with the neurophysiology of the same system, which aims exclusively at the physical interior of the system and not at the interior/exterior interface and at the purposes of the global behavior. Simon's 'artificial' could be called functional in the sense of teleological - what interests us here is the fact that what makes the functional 'functional' isn't the fact that characterizes artifacts but the facts that it characterizes adaptive devices. Now, an adaptive device may be, or not, artificial in the usual sense of artificial (being made or constructed by human beings). By the way, we too are adaptive devices put together by evolution by natural selection

– and this is exactly what Dennett uses against Searle, to fight Searle’s appeal to ‘intuitions’ and to what’s ‘original’ and genuine about our type of intentionality.

#### **4. Functionalism, cognitivism and the natural/artificial dichotomy. Conclusion: D. Dennett on natural intelligence.**

I will conclude taking Dennett as a reference. It is a fact that his defense of cognitivist functionalism and Strong AI involves, and this is essential, a representational theory of the nature of consciousness that is very different from Searle’s theory of consciousness, and also the reformulation of functionalism as an hypothesis concerning not only the brain but the whole body. What I’m interested in, is how Dennett’s brand of cognitivism makes us look at our own mind and intelligence. To start with, for Dennett it is precisely the fact that functionalist cognitivism does away with the distinction between natural and artificial that unifies cognitive science as a field, in that it allows us (i) to consider humans, animals and machines together as cognitive systems (ii) to assume that cognitive performances of actual, biologically based, minds are situated, with other possible types of minds, in the same design space (in fact, not only cognitive performances but also the systems to which they are due, as well as the artifacts these systems produce, for example human instruments). From here we must think in terms of thresholds, conditions of possibility for human minds, as a specific kind of minds (as opposed to something ‘original’). These thresholds are, according to Dennett, connected with architectures for communication and language, which render the speechact nature of human thought possible. This, allowing namely for mental acts such as endorsing and affirming one’s own beliefs through language and making voluntary decisions, is what makes human minds so much more powerful and sophisticated than those of other animals, so much so that, frequently, the work of a philosopher approaching, for instance, the problem of mind in cognitive ethology will consist in deflating interpretations of animal behavior, namely overattributions of consciousness and communication to other animals. Dennett’s cognitivist functionalism, with its deconstructionist stance on ‘genuine intentionality’ coexists, thus, both with a deflationary conception of animal mentality and with the idea that the distance between human minds and the minds of other species, even the most intelligent, is enormous. It is big enough, namely, to make all the difference in moral terms. Despite this, one should not claim that there is any difference in kind, since it is still a matter of cognitive architecture only, and so a difference in degree. All this goes against Searle’s positions – but why?

I want to answer this question by evoking, as Dennett himself does, Alan Turing, the english logician and mathematician. I'm interested in Turing as a philosopher – the author of the paper “Computing Machinery and Intelligence” published in the philosophical journal *Mind* in 1950 – rather than in the work for which he is considered one of the fathers of the computer, the creator of the concepts of Turing Machine and Universal Turing Machine. *Computing Machinery and Intelligence* was, already then, written against the critics of AI, and it is there that Turing proposes the Turing Test, the test that Searle's Chinese Room is supposed to have 'refuted'. The initial question is: can machines think? and the first thing Turing says is that a conceptual quarrel around the questions what is a machine? what is thinking? would lead to complications without end. The Turing Test is then suggested as a practical substitute for approaching the nature of intelligence. The Turing Test is an imitation game, one in which there is, in the original situation, an interrogator and two people, of whose gender the interrogator is unaware. The goal is to deceive the interrogator (through verbal interaction, the only one allowed). In the setting which interests us, there's no man and woman but rather a human and a machine. Same rules, same type of interaction. The goal is again to deceive the interrogator, this time concerning the status of machine of the machine. As we know, there's only verbal interaction, so all that can be done to find out which is which is to formulate questions (questions about mathematical calculations, interpretation of poems, ironic comments or deciphering of metaphors, for instance). Turing's point, the intention behind the Test, is that we should think about thought in a neutral, unbiased way – the Test assumes that what behaves intelligently is intelligent. The implicit suggestion is, I would say, holistic and pragmatic: intelligence is intelligent behavior and not a special substance (for example neural) or an extra ingredient (for example a soul – or genuine intentionality).

In the remaining of the paper Turing answers objections against the possibility of AI, all of which are still around today: he calls them the theological objection (according to which intelligence is due to a soul, only possessed by human beings), the 'heads in the sand' objection (one expects that AI does not come into being or it would be terrible), the mathematical objection, evoking Godel's theorem (according to which the human beings have mental capacities which exceed what is computable), the feelings and consciousness objection (according to which a machine could not have states such as depression, love, emotion, etc), the incapacities objection (according to which a machine would never be capable of mood, learning, morality, passion), etc. He also considers the argument according to which a machine does only what it is programmed to do, never originates anything

new, as well as the argument of the continuity of the nervous system. Turing analyzes these lines of objection one by one and continues to recommend the test against any apriorist verdict.

The points I want to make as a conclusion of this brief tour of cognitivism grow out of Dennett's endorsement of Turing's strategy. If we side with Turing and Dennett against Searle's appeal to intuition and to something 'original' and unique about our minds, what we have is the following: (i) we should not be apriorists in defining intelligence, (ii) the frontiers between machines and non-machines, between thinking and non-thinking, may be fuzzier than what we would intuitively think, (iii) even though we are mental beings, and so, in a way, incorrigible in accessing our own minds, our intuition is not necessarily an infallible guide in thinking about the nature of the mental.

### References

Andler, Daniel, 1992, *Introduction aux sciences cognitives*, Paris, Gallimard.

Boden, Margaret (ed.), 1996, *The Philosophy of Artificial Intelligence*, Oxford, Oxford University Press.

Chalmers, David, 1996, *The Conscious Mind*, Oxford, Oxford University Press.

Descombes, Vincent, 1995, *La dernière mentale*, Paris, Minuit, 1995.

Dreyfus, Hubert, 1965, *Alchemy and Artificial Intelligence*, Rand Corporation Report, Santa Monica, California.

Dreyfus, Hubert, 1972, *What Computers Can't Do*, New York, Harper & Row.

Fodor, Jerry, 1975, *The Language of Thought*, Cambridge MA, Harvard University Press.

Haugeland, John (ed), 1981, *Mind Design*, Cambridge MA, MIT Press.

Minsky, Marvin, 1985, *The Society of Mind*, New York, Simon & Schuster.

Newell, Allen, 1990, *Unified Theories of Cognition*, Cambridge MA, Harvard University Press.

Putnam, Hilary, 1960, "Minds and Machines", in *Philosophical Papers*, vol. 2, Cambridge, Cambridge University Press, 1975.

Putnam, Hilary, 1967, "The Nature of Mental States", in *Philosophical Papers*, vol. 2, Cambridge, Cambridge University Press, 1975 (first published as Psychological Predicates).

Searle, John, 1980, "Minds, Brains and Programs", *Behavioral and*

*Brain Sciences*, 3 (3).

Searle, John, 1992, *The Rediscovery of the Mind*, Cambridge MA, MIT Press.

Simon, Herbert, 1969, *The Sciences of the Artificial*, Cambridge Mass., MIT Press.

Turing, Alan, "Computing Machinery and Intelligence" (1950), in Dennett & Hofstadter (eds.), 1981, *The Mind's I*, New York, Bantam Books.

## HETEROFENOMENOLOGIA

Abstract: Este artigo é um comentário a uma proposta do filósofo da mente Daniel Dennett: desde há muito Dennett defende um método para introduzir dados de primeira pessoa em ciência cognitiva a que chama “heterofenomenologia”. São aqui feitas algumas críticas a esse método, não porque o problema que Dennett enfrenta ao propô-lo, o problema das relações aparência-realidade quando se trata da consciência, não pareça relevante e desafiador, mas porque a particular forma de Dennett o enfrentar incorre em suposições metafísicas duvidosas.

### **1. Consciência, fenomenologia e o problema das relações aparência-realidade**

O filósofo americano Daniel Dennett é um conhecido teórico da consciência e a consciência é um dos problemas centrais da filosofia da mente entendida como filosofia da ciência cognitiva. A teoria dennettiana da consciência inclui uma crítica do Teatro Cartesiano (uma concepção do mental como centro e apresentação a um sujeito que Dennett considera que continua presente mesmo em tempos em que o materialismo é dominante nos estudos da mente-cérebro) e a proposta de um modelo alternativo, o Modelo dos Esboços Múltiplos. Não vou falar aqui da desmontagem que Dennett faz do Teatro Cartesiano, nem da proposta que avança para o substituir mas apenas considerar a forma como o problema das relações aparência-realidade se lhe coloca quando discute filosoficamente a consciência.

O problema das relações aparência-realidade está presente na filosofia desde os inícios desta – pense-se em Parménides, em Platão, em Aristóteles – e continua hoje a colocar desafios, nomeadamente quando se trata da consciência e em concreto da forma como a vida mental própria aparece a cada um de nós, com os seus conteúdos específicos e a sua riqueza de detalhes. Dennett enfrenta este problema ao propôr um método para introduzir dados de primeira pessoa (i.e. relatos que os sujeitos fazem do seu fluxo de consciência) na ciência cognitiva. Chama a tal método ‘heterofenomenologia’.



Vejam os que isto significa. O termo ‘fenomenologia’ refere num primeiro momento a vida mental, a experiência consciente em cada um de nós. Este sentido de fenomenologia é um sentido mais básico do que o sentido de ‘fenomenologia’ como movimento filosófico específico, com raiz no filósofo alemão E. Husserl (1869-1938) e que se propunha descrever filosoficamente a experiência da consciência. O termo que Dennett utiliza, o termo ‘heterofenomenologia’, significa qualquer coisa como ‘fenomenologia em terceira pessoa’. Mas será que essa ideia de uma fenomenologia em terceira pessoa é uma boa ideia? Será que é possível fazer tal coisa?

Responder a estas perguntas conduz-nos a considerar a estranha situação em que nos encontramos quando se trata da nossa experiência consciente. A situação é mais ou menos a seguinte. Parece não poder ser o caso que a existência de experiência consciente seja uma aparência ilusória para o sujeito dessa experiência consciente (para mim, neste momento, por exemplo): essa é uma espécie de evidência cartesiana. Não é concebível que eu me engane. Mas o caso torna-se mais complicado, quando se pergunta se os conteúdos da experiência consciente podem ser ilusórios para o sujeito: por exemplo, eu olho e vejo um pássaro azul a voar baixinho, só que afinal não era um pássaro mas um pequeno helicóptero de brinquedo. É perfeitamente concebível que a minha experiência consciente me engane relativamente à forma como as coisas são no mundo. Isto mostra que estão em jogo quando se descreve a experiência consciente questões subtilmente distintas umas das outras, nomeadamente: Será possível ilusão acerca do facto de as coisas me aparecerem? Será possível ilusão acerca da forma como as coisas me aparecem? Será que posso confiar em mim própria relativamente à forma como as coisas me aparecem? Será que não posso estar enganada acerca da forma como as coisas me parecem? Que estas questões sejam diferentes umas das outras não é de todo indiferente para falar de fenomenologia e para considerar a introdução de dados fenomenológicos na ciência cognitiva.

Mas voltemos à fenomenologia no primeiro sentido de vida mental ou experiência consciente. Thomas Nagel, um outro filósofo da mente americano, afirmava no seu célebre artigo *What is it like to be a bat?*<sup>240</sup> que é uma constatação incontornável quando se trata da mente que os factos fenomenológicos são perfeitamente objectivos. Isto simplesmente quer dizer que o mundo, o nosso mundo, é tal que contém subjectividade, da mesma forma que o mundo é espaço-temporal. Nem a subjectividade nem o espaço-tempo são explicáveis em função de qualquer outra coisa. Esta existência de subjectividade (i.e. consciência), no mundo tal como este é

<sup>240</sup> NAGEL 1974.

seria então um facto bruto no mundo e um facto bruto no sistema do ponto de vista do próprio sistema. É assim que filósofos da mente como Nagel, mas também J. Searle, D. Chalmers e muitos outros começam a falar de consciência: há aparecer subjectivo, isso é um facto bruto, e esse facto parece dever ter repercussões no nosso quadro metafísico do mundo. No entanto, nem todos os filósofos analíticos da mente consideram a subjectividade assim fundamental e D. Dennett é o caso mais radical. Ele propõe, como parte da sua teoria da consciência, algo que parece uma eliminação da subjectividade da nossa concepção de mundo e curiosamente faz isso no próprio momento em que propõe um método fenomenológico para a teoria da mente, a ‘heterofenomenologia’. Vou procurar explicar como é que isto acontece e em que é que me parece criticável.

## 2. Heterofenomenologia

Em termos práticos, quando se investiga a mente, deparamo-nos com o problema de saber como devemos tomar – sejamos psiquiatras, antropólogos, filósofos, ou teóricos da mente de qualquer outro tipo – o ser do mundo de alguma maneira para alguém. O que devemos pensar daquilo que alguém diz acerca da forma como o mundo é para si? O que devemos pensar quando alguém descreve, descrevendo a sua experiência consciente, o mundo tal como o encontra (abro os olhos e vejo: “vejo os campos, Neera, / campos campos e soffro /...<sup>241</sup>)? Mais especificamente: devemos pensar no que é descrito como tratando-se de aparências (tal como exprimiríamos pela palavra inglesa *seeming*, querendo dizer pareceres ou aparências possivelmente ilusórios)? Ou esse espaço entre real e ilusório não deve ser demasiado prontamente introduzido na descrição, tratando-se antes de mais apenas do aparecer das coisas a alguém, a que deve ser feita justiça? Tratar-se-ia então de ‘haver aparecer’, expresso pelo grego *Phainomenon* ou pelo alemão *Erscheinung* – em português poderíamos dizer ‘Fenómeno’ enquanto dar-se, aparecer, ainda sem conotação de aparência. Saber como deve ser tomado esse aparecer do mundo (para mim, para um eu qualquer, de uma dada forma) é uma questão metodológica. No entanto, não é possível tratá-la sem considerar questões metafísicas acerca de pensamento e mundo.

Vejamos o que Dennett faz. Ao contrário do que acontece com outros filósofos da mente (atrás referi Thomas Nagel) ‘fenomenologia’, tal como Dennett usa o termo, não nomeia exactamente o facto bruto de haver experiência consciente. Na verdade ele usa o termo à imagem do uso

<sup>241</sup> Fernando Pessoa [Ricardo Reis].

científico-natural deste, em que ‘fenomenologia’ significa sinais, sintomas (por exemplo de um paciente), aquilo que se observa. A ideia é então aplicada à consciência. Temos então os conteúdos da mente de um indivíduo, conteúdos subjectivos, acedidos em primeira pessoa e inacessíveis em terceira pessoa: ‘estou a ouvir uma sonata de Brahms’, ‘vejo o papel de parede cheio de Marylins cobrindo as paredes do quarto’<sup>242</sup>. O conhecimento desses conteúdos em primeira pessoa é considerado como sendo introspectivo, sendo esse conhecimento intropectivo tomado (pelo sujeito da experiência) como infalível. Isto é ‘auto-fenomenologia’ e não é ainda ‘utilizável’ pelo cientista cognitivo. Há aqui um pormenor muito importante: falando de fenomenologia Dennett fala sempre de aparências (*‘Conscious experiences are seemings’*, diz). A heterofenomenologia tal como Dennett a propõe pretende ser uma ponte entre estas aparências, estes *seemings*, e as ciências naturais – a ideia é um observador exterior (o heterofenomenólogo) tomar pelo seu valor facial aquilo que cada sujeito diz acerca da forma como o mundo é, passando depois ao contraste entre aquilo que é assim ‘recolhido’ com aquilo que acontece no sujeito do ponto de vista cognitivo/físico, nomeadamente no cérebro do sujeito que descreve a sua experiência consciente.

Estes usos do termo ‘fenomenologia’ por Dennett (‘auto-fenomenologia’, ‘hetero-fenomenologia’) quase não têm pontos comuns com o entendimento que Husserl teve da fenomenologia como método filosófico e do propósito deste<sup>243</sup>. Husserl estava interessado em desenvolver a fenomenologia como método para a investigação filosófica em geral, não, certamente, como método para a ciência cognitiva (no seu tempo falar-se-ia de psicologia). Quando avança com a proposta da fenomenologia, trata-se para Husserl de responder à pergunta ‘O que se faz quando se faz filosofia?’ (com toda a latitude de tópicos desta, da filosofia da matemática à ética) e não de responder à pergunta ‘Como se deve tomar em teoria da mente a forma como as coisas são para um particular sujeito?’ (é disto que se trata para Dennett). A fenomenologia husserliana inicial tinha um propósito epistemológico e mesmo ontológico geral: tratava-se para Husserl de compreender as relações da subjectividade com a objectividade, daí o interesse por tópicos como ciência, lógica, conhecimento, significação, evidência, verdade (são estes os tópicos das *Investigações Lógicas* (1900-1901), por exemplo); nisso a fenomenologia de Husserl tinha muito em comum com a filosofia analítica inicial, nomeadamente com G. Frege. Em contraste, Dennett, como outros filósofos da mente, quer sobretudo compreender como é que a consciência e os seus conteúdos se relacionam com o mundo

<sup>242</sup> Dennett 1991 : 354.

físico, assumindo portanto desde logo que o mundo existe tal como é posto pela atitude natural, que é um e que é físico, que se sabe o que se está a fazer quando se faz filosofia. Aliás, Dennett frequentemente identifica fenomenologia no sentido ingénuo com a tradição iniciada por Husserl, o que não é muito justo. É de qualquer forma à auto-fenomenologia assim concebida (conhecimento em primeira pessoa, introspecção tomada como infalível), que Dennett opõe como alternativa a heterofenomenologia – esta é suposta ser uma abordagem da consciência em terceira pessoa que toma a primeira pessoa ‘de forma absolutamente séria’. Isto é o que afirma Dennett. Mas o que é tomar a primeira pessoa ‘de forma absolutamente séria’? Disse atrás que Dennett pensa que experiências conscientes são aparências (*seemings*). É preciso acrescentar que ele pensa que esses *seemings* são crenças, crenças acerca da forma como as coisas me parecem, de um ponto de vista subjectivo, o meu ponto de vista. Isto é para Dennett, a subjectividade; é isso que é para mim ser eu, este eu.

Esta identificação da minha experiência com crenças acerca da forma como as coisas parecem permite a Dennett o traço mais característico da sua heterofenomenologia: a ideia de acordo com a qual é possível, e mesmo inevitável que numa ciência da consciência aconteça qualquer coisa como uma ‘refutação da fenomenologia’ (das crenças acerca da forma como as coisas parecem) por outros dados (dados relativos ao mesmo sistema cognitivo, recolhidos através dos métodos de terceira pessoa das ciências naturais). Sem entrarmos na discussão específica das posições de Dennett (elas são bastante subtis) vejamos o que está aqui em jogo: por exemplo um sujeito tem a experiência consciente de xxxxxx (1); no seu cérebro é registado yyyyyy (2). O ponto de Dennett é que (2) pode desmentir (1). Esta ideia de que o ponto de vista de terceira pessoa pode ‘desmentir’ a experiência consciente é alias sustentada ao mesmo tempo que se defende, com a crítica aos *qualia*, uma inexistência de espaço de si a si num sujeito.

É importante salientar que a identificação da subjectividade com um conjunto de crenças acerca da forma como as coisas parecem é necessária para entender a célebre posição de Dennett acerca de *qualia* (o eliminativismo dos *qualia* é talvez a parte mais conhecida da sua teoria da consciência). *Qualia* significa para a maioria dos filósofos da mente o sentir qualitativo da minha mente, o como é ser-se um dado sujeito. É costume inclusive contrastar *qualia* com crenças. Como teórico da consciência, Dennett é conhecido por eliminar os *qualia*. O que significará isto? Significará que ele pensa que somos todos zombies, seres sem interior, seres que não se sentem ser? Não é exactamente isso - vejamos brevemente o que é esta eliminação dos *qualia*.

No artigo *Quining Qualia*<sup>244</sup> Dennett apresenta várias *intuition-pumps* (histórias para provocar as nossas intuições) para pensar sobre *qualia* e conjuntamente com os casos mais clássicos de inversão inter e intrapessoal do espectro visual, encontra-se o exemplo de dois provadores de café, Chase e Sanborn. O trabalho de Chase e Sanborn consiste em assegurar que o sabor do café que provam se mantém o mesmo ao longo do tempo. Continuando a ser eficazes na tarefa, ambos perderam no entanto o prazer que costumavam sentir quando provavam este café, que costumava ser o seu favorito acima de qualquer outro café. Um deles pensa que isso aconteceu porque se tornou um provador de café muito mais sofisticado, embora o sabor daquele café permaneça idêntico. O outro pensa que isso aconteceu porque as suas papilas gustativas se alteraram, embora o sabor daquele café inicial se mantenha idêntico. Qual será realmente o caso com cada um deles? Segundo Dennett, nenhum deles é capaz de estabelecer intrasubjectivamente a resposta. Ora, se os sujeitos não apercebem explicitamente uma diferença, não existe diferença. Não existem *qualia* inefáveis em cada um. Dennett chama a este princípio verificacionismo de primeira pessoa. De acordo com o verificacionismo de primeira pessoa o apercebimento de x (um conteúdo específico na consciência de um agente cognitivo, o estar consciente de um dado conteúdo, por exemplo *este* sabor de café) não pode existir sem uma crença do agente. Segundo Dennett, “There is no phenomenon as really seeming over and above the phenomenon of judging in one way or another that something is the case”<sup>245</sup>. Assim sendo, é por principio impossível, para o próprio sujeito, decidir, no que diz respeito ao seu apercebimento consciente de x entre aparência e realidade. Quando se trata daquilo de que os agentes são conscientes não faz sentido pensar que há algum *real seem*, algum parecer real, que existe nos sujeitos sem que estes tenham crenças acerca disso. Diferenças que não fazem diferença não existem.

Em termos da história da filosofia este verificacionismo de primeira pessoa liga Dennett ao filósofo austríaco L. Wittgenstein. Ele foi pela primeira vez formulado em *Are Dreams Experiences?* (1978) um artigo no qual Dennett parte das propostas do filósofo Norman Malcolm acerca dos sonhos no livro *Dreaming* (Malcolm negara, de forma radical, que investigações sobre o sono REM, um estágio do sono, pudessem ser sobre o que na linguagem comum chamamos ‘sonhar’). As propostas de Malcolm por sua vez tinham tido origem numa observação de Wittgenstein nas *Investigações Filosóficas* (Part II, 11, 213), em que Wittgenstein diz “A questão

<sup>244</sup> Dennett 1988.

<sup>245</sup> Dennett 1991: 364.

de saber se a memória engana a pessoa que sonhou quando ela, depois de acordar, narra o sonho, não se pode levantar a não ser que se introduza um critério completamente novo para determinar a concordância da narrativa com o sonho, um critério que distinga a verdade da sinceridade”.

O ponto de Dennett é que quando se trata da mente aquilo que se tem na mão são sempre relações entre memória de conteúdos e expressão: não faz sentido evocar uma pura apresentação fenomenal independente de memória e da expressão pelo sujeito, pois qual seria a evidência da existência dessa ‘pura apresentação’? Nos termos introduzidos por outro filósofo da mente americano, Ned Block, Dennett não aceita a ideia de consciência fenomenal sem consciência de acesso (*phenomenality without reflexivity*). Pensa que aqueles que querem defender essa diferença estão amarrados ao ‘centralismo cartesiano’, à presunção de que o mental envolve um centro a que tudo se apresenta. Mas essa é uma má ideia. É isso a eliminação dos *qualia*.

Os princípios da heterofenomenologia vêm juntar-se a esta eliminação dos *qualia*. Consideremos como exemplo um caso concreto muito estudado, os fenómenos de cegueira de inatensão (*inattention blindness*). Dennett proclama aliás, em *Sweet Dreams*<sup>246</sup>, ter sido ele a sugerir este tipo de experiências. Numa das mais discutidas experiências, os sujeitos vêm um video em que vários jogadores jogam basketball passando a bola entre si. Os sujeitos receberam instruções para contar os passes realizados e para se concentrarem nisso. Uma pessoa vestida de gorila passa entre os jogadores. Quando no fim da experiência se pergunta aos sujeitos de viram algo de pouco usual, 50% respondem negativamente<sup>247</sup>. Qual será realmente o caso com cada um dos sujeitos que respondem negativamente (i.e. que respondem que não viram nada de anormal)? Eles viram ou não viram o gorila? De acordo com Dennett nenhum deles é capaz de estabelecer intrasubjectivamente uma resposta a esta pergunta e nós não temos o direito de presumir que existiu uma apresentação fenomenal no sujeito não apercebida pelo sujeito (‘ver o gorila’, sem saber que o vê). Mas e se se fosse ‘observar’ o cérebro dos sujeitos? É concebível que se pudesse ‘desmentir’ o que eles relatam (há fixações de conteúdo que poderiam, se ‘sondadas’, ter dado origem a relatos sobre o gorila). Mas isso continuaria a nada dizer sobre a consciência dos sujeitos – aí a última palavra é o verificacionismo de primeira pessoa.

<sup>246</sup> Dennett 2005.

<sup>247</sup> Cf. Simons & Chabris 1999. Existem vários videos disponíveis on-line cf. por exemplo o site do Visual Cognition Lab da Universidade de Illinois ([http://viscog.beckman.uiuc.edu/djs\\_lab/demos.html](http://viscog.beckman.uiuc.edu/djs_lab/demos.html))

### 3. A forma como a minha vida mental me aparece: conclusão

O que há a dizer acerca disto? É uma complexa e interessante posição sobre a nossa experiência consciente, mas pelo menos uma coisa não está bem. Formulei no início do ensaio uma distinção entre aparecer e parecer, para poder identificar as questões subtilmente distintas que o problema das relações aparência-realidade traz quando se trata de consciência. Ora, uma das constações perante as posições de Dennett quanto a heterofenomenologia e *qualia* é que falando de consciência ele fala sempre de *seemings*, pareceres como aparências. Ele é cego à existência de um aparecer da consciência, daí que a intuição comum a muitos filósofos da mente – que a consciência é um aparecer – lhe esteja fechada. Só isto explica que ele elimine a subjectividade ao mesmo tempo que propõe um método fenomenológico: os pareceres ou aparências são identificados com crenças (i.e. pareceres judicativos), essas crenças (constitutivas da subjectividade) podem ser demonstradas ser falsas a partir de um ponto de vista externo. Elas são um mero parecer, os pareceres podem ser ilusórios e portanto podem ser corrigidos (a partir de fora, pelo heterofenomenólogo). O que fica é apenas mundo tal como as ciências naturais dizem que este é – não existe subjectividade, apenas a (suposta) ‘*view from nowhere*’ da ciência natural. Se a consciência fosse vista como um aparecer –é para aí que o *What is it like to be?* nageliano aponta – esta eliminação da subjectividade não faria sentido. Apenas identificando a subjectividade com aparências Dennett pode ‘eliminar-la’.

Mas este é um ponto cego da concepção dennettiana do mental. No seu livro maior sobre o problema da consciência, *Consciousness Explained*, Dennett insistia em que quando se trata de consciência não se pode eliminar o observador já que o que está em causa é o próprio observador. Procurei até aqui mostrar que nas suas posições sobre heterofenomenologia e *qualia* dá precisamente esse último passo: eliminar o observador, eliminar a subjectividade. Do ponto de vista da história da filosofia isso pode explicar-se por Dennett ser um quineano<sup>248</sup>. A concepção quineana de ontologia é muito espalhada e aparentemente neutra. Segundo Quine, a resposta à pergunta ontológica ‘o que é que há?’ deve ser procurada nos compromissos ontológicos das nossas melhores teorias: o que há é o que as nossas melhores teorias nos dizem que há. Isto parece neutro mas na verdade tem reflexos directos na forma como se concebe a vida mental e particularmente a consciência. Ao contrário do que se passa por exemplo nas filosofias de inspiração kantiana, a questão das relações aparecer-realidade

<sup>248</sup> Um seguidor de W. V. Quine (1908-2000), um importante lógico e filósofo americano.

nem sequer se coloca para Quine quando se trata de considerar questões ontológicas. Perante a pergunta ‘O que é que há?’ a resposta é ‘o que há é o que as nossas melhores teorias científicas dizem que há’. Noutras palavras, estamos já num nível em que há pensamento, há teorias, é em função dessas que vamos dizer ‘aquilo que há’. Isso fará com que um quineano como Dennett trabalhe sobre a consciência munido de um ponto cego específico: se a questão dos *qualia* e da heterofenomenologia diz respeito ao aparecer (do pensamento no mundo) ela não vai caber nesta malha. Aliás, repare-se: a questão em jogo não é uma questão qualquer, não é apenas uma questão standard de filosofia da mente. De certa forma a questão das relações aparecer-realidade é o problema central da filosofia: há mundo e há pensamento no mundo, e os problemas filosóficos específicos decorrem daí. Mas os compromissos quineanos de Dennett não lhe permitem ver este alcance maior do que está em causa quando se trata de consciência. Em suma, o que é criticável na ‘heterofenomenologia’ de Dennett não é sequer sobretudo o que autores como J. Dokic e E. Pacherie chamam o seu ‘compromisso doxológico’ (a ideia de que na consciência não há espaço para a distinção entre aparência e realidade, o que não permitiria conceber a natureza de certas patologias da consciência como por exemplo a visão cega)<sup>249</sup> mas sobretudo a eliminação do ineliminável – a eliminação da realidade do aparecer. O aparecer das coisas à consciência não pode ser simplesmente identificado com aparências ilusórias, corrigíveis de algum outro ponto de vista, o ponto ‘de vista de lugar nenhum’ que seria o ponto de vista das ciências naturais. Aquilo que está em causa é a existência de pensamento no mundo, e esse facto tem que ter um lugar na nossa teoria da mente.

### Referências

Dennett, D. 1988, Quining Qualia, in Marcel and Bisiach eds, *Consciousness in Contemporary Science*, Cambridge, Cambridge University Press.

Dennett, D. 1991, *Consciousness Explained*, New York, Little Brown and Co.

Dennett, D. 2005, *Sweet Dreams*, Cambridge MA, MIT Press.

Dennett, D. 2007, Heterophenomenology reconsidered, in *Phenomenology and the Cognitive Sciences* 6.

Dokic, J & Pacherie, E., 2007, Too much ado about belief, *Phenomenology and the Cognitive Sciences* 6.

<sup>249</sup> Dokic & Pacherie 2007.



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

Miguens, S., 2008, *Será que a minha mente está dentro da minha cabeça?*, Porto, Campo das Letras.

Nagel, T., 1974, What is it like to be a bat?, in *Mortal Questions*, Cambridge, Cambridge University Press.

Nagel, T., 1986, *The View from Nowhere*, Oxford, Oxford University Press.

Simons, D. & Chabris, C. 1999, Gorillas in our midst : sustained inattention blindness for dynamic events, *Perception* 28,1059-1074.

# IDENTIDADE PESSOAL E ACÇÃO



## PROBLEMAS DE IDENTIDADE PESSOAL<sup>250</sup>

Resumo: O objectivo deste artigo é mostrar que debaixo da expressão ‘problemas de Identidade Pessoal’ caem questões de ordem muito diferente: questões *cognitivas*, questões *metafísicas* e outras questões a que se chamará questões *narrativas* (psicológicas, sociais, artísticas, etc). Pretende-se distinguir o mais claramente possível o âmbito de cada uma destas questões, uma vez que o facto de elas serem frequentemente agregadas produz mais um problema (desnecessário) da Identidade Pessoal.

### **1. De que falamos quando falamos de Identidade Pessoal: a natureza do problema da Identidade Pessoal, a importância do problema da Identidade Pessoal.**

Para que serve falar de identidade?<sup>251</sup> Antes de mais, falar de identidade serve para saber exactamente de que se fala. Falar-se-á aqui de *Identidade Pessoal* (IP), i.e. daquilo que faz com que uma particular pessoa seja *essa* pessoa e a *mesma* pessoa *ao longo do tempo*. Quando se fala de Identidade Pessoal não se está portanto a falar de identidade no sentido formal, lógico-matemático, caso em que a identidade é uma relação pela qual por exemplo  $A=A$ , mas, antes de mais, de um processo cognitivo, um processo de reidentificação do mesmo como o mesmo.

Saber que mesmidade é essa que é reidentificada e o que é isso que a reidentifica é o núcleo do problema, já que tudo leva a crer que não existe

<sup>250</sup> O presente artigo tem uma tripla origem: em primeiro lugar e antes de mais, ele tem origem no trabalho relativo à cadeira de Filosofia do Conhecimento leccionada na FLUP. Entre 1996/97 e 1999/2000 o programa de Filosofia do Conhecimento teve como um dos temas orientadores o problema mente-corpo. Nesse contexto, um tema específico tratado era a questão da Identidade Pessoal, quer na filosofia da mente (em T. Nagel, D. Dennett, etc) quer em autores como Descartes, Leibniz, Hume e Kant. Em segundo lugar aparece a necessidade de ordenar materiais recolhidos e tratados por ocasião da frequência do seminário intitulado *Metaphysics - Personal Identity* do Professor Derek Parfit, na New York University, durante o Semestre de Outono de 2000. Não há aqui qualquer pretensão de exaustividade, uma vez que o Professor Parfit procurava nesse seminário refazer totalmente vários argumentos do seu livro *Reasons and Persons* (Parfit 1984). Uma terceira origem deste texto, mais próxima e mais directa, foi a participação na Mesa Redonda de Primavera da SPAE (Sociedade Portuguesa de Antropologia e Etnologia), dedicada ao tema *Identidade/Identities*, em Março de 2001, a convite do Professor Doutor Vítor Oliveira Jorge (Departamento de Ciências e Técnicas do Património/FLUP).

<sup>251</sup> Este foi o primeiro desafio da mesa redonda referida na nota anterior.

aí um ‘simples’ à espera de ser identificado (ou de identificar). Usualmente a reidentificação efectuada é reportada a uma *continuidade* existente: quando se fala de Identidade Pessoal fala-se de *critérios de Identidade Pessoal* e a continuidade é habitualmente o primeiro critério avançado, seja a continuidade física (a mesmidade de um corpo físico ou de uma parte suficiente deste, por exemplo, de um cérebro) seja a continuidade psicológica (a continuidade da vida mental a partir de dentro, para um determinado sujeito que recorda os sucessivos estados).

Mas desta forma apenas se dispõe os dados do problema. Como se verá, há tudo a ganhar na distinção clara de tipos e níveis de questões que se colocam em torno da IP. De facto, as questões da IP incluem:

- (1) *questões cognitivas* relativas ao auto-apercebimento (*self-awareness*), envolvendo aspectos como o auto-controlo, a memória, a introspecção, a propriocepção, etc, e relativas à auto-referência, nomeadamente a partir da instituição de uma auto-representação linguística explícita (o Eu) que possibilita um certo tipo de centralização e de unificação em sistemas cognitivos nos quais ao nível do *hardware*, nenhuma unificação ou centralização ‘existe’,
- (2) *questões metafísicas* relativas à natureza de pessoas, ao tipo de entidades no mundo que são as pessoas,
- (3) *questões narrativas* relativas ao relato autobiográfico de si ao longo do tempo, à constituição sempre mais ou menos ficcionalizada que é uma parte importante da psicologia das pessoas bem como às fontes e materiais deste relato.

Outras questões ainda, como as relativas ao sentimento de si, à consciência fenomenal e à racionalidade auto-consciente a que se pode chamar ‘pensamento’ são difíceis de situar: não é por exemplo simples decidir se se trata de questões do primeiro tipo, do segundo tipo, ou de ambos. No entanto, todas estas questões têm a sua parte no problema da natureza da IP, frequentemente aparecendo ‘cruzadas’ na literatura relevante.

A distinção aqui proposta (entre sub-questões cognitivas, metafísicas e narrativas da questão da IP) tem como objectivo o esclarecimento conceptual, embora em muitos casos (ou em todos) este apenas seja possível mediante investigações empíricas. A grande importância que a questão da IP assume decorre, no entanto, de uma outra ordem de razões que não o mero desejo de esclarecimento conceptual. De facto, a concepção da IP tem fortes repercussões práticas (morais, legais). Em concreto e de forma esquemática, considerando por exemplo a grande divisão de opiniões quanto àquilo que, do ponto de vista metafísico, é uma pessoa, temos que de um lado existem posições gradualistas e reducionistas segundo as quais uma pessoa não é algo desde logo determinado, um facto a mais – posições

como as de John Locke, William James ou, hoje, Derek Parfit – e do outro lado posições absolutistas, segundo as quais ser uma pessoa é algo de determinado, específico, absoluto, um facto a mais. É o caso das posições defendidas pelos filósofos escoceses Joseph Butler e Thomas Reid, que no século XVIII se opuseram à concepção lockeana de Identidade Pessoal, da posição defendida por Kant, na sua filosofia moral, e até de certo modo da posição defendida hoje por Donald Davidson<sup>252</sup>. É também, como é evidente, o caso das posições de matriz religiosa que prevêm a existência de uma alma imaterial individual. De acordo com as posições a que se está a chamar ‘absolutistas’ e que usualmente ligam de forma estreita as noções de ‘humano’ e de ‘pessoa’, uma entidade que é uma pessoa é sempre e em todas as circunstâncias uma pessoa, havendo algo de essencial na qualificação. Em contrapartida de acordo com posições gradualistas e reducionistas, é possível por exemplo, no limite, que existam indivíduos humanos que já não são ou ainda não são pessoas, indivíduos que não são humanos mas são pessoas, indivíduos que começam por ser biologicamente humanos, têm os seus materiais substituídos progressivamente e continuam a ser a mesma pessoa. Como é óbvio, a opção por uma das posições constituirá fundamento para decisões práticas muito diferentes.

Uma palavra quanto a método. O tratamento da questão da IP na filosofia da mente nas últimas décadas faz frequentemente apelo, como já Locke fazia no *Ensaio sobre o Entendimento Humano*, a casos imaginários variados. Dada a frequências dos pressupostos funcionalistas (de acordo com os quais a matéria de que um efeito é feito – no caso o efeito é *ser uma pessoa* e o material é por exemplo uma determinada organização físico-biológica – é secundária no que diz respeito à natureza da entidade resultante) não são raras situações de quase ficção científica, em que se experimenta por exemplo sobre a ideia de padrões de mentalidade pessoal armazenáveis e replicáveis, sobre transplantes de cérebros ou partes de cérebros, sobre casos de ‘espalhamento da implementação’ dos padrões mentais que constituem uma pessoa (por exemplo a outras escalas que não as correspondentes aos corpos e órgãos humanos). Encontram-se casos de reimplementação dando lugar a uma espécie de imortalidade materialista, casos de indistinção entre cópias e originais dos padrões da mentalidade que mostram a irrelevância da distinção entre natural e artificial, e muitas outras, frequentemente fantásticas, variações. Um lugar por excelência para encontrar experiências semelhantes é a obra *The Mind's I*, de D. Dennett e D. Hofstadter, um outro o livro *Reasons and Persons* de Parfit<sup>253</sup>. Os

<sup>252</sup> Cf. Davidson 1982, *Rational Animals*.

<sup>253</sup> Hofstadter & Dennett 1981. Parfit 1984.

exemplos poderiam no entanto multiplicar-se. Evidentemente, as críticas à relevância do método dos casos imaginários são também frequentes na literatura. Mas, precisamente, admitir ou não admitir à partida o método dos casos imaginários traduz já compromissos teóricos quanto àquilo que no mundo são mentes e pessoas e quanto à forma como as questões da IP devem ser pensadas.

## 2. Questões cognitivas

À medida que vamos sabendo mais acerca das condições de auto-apercebimento, auto-referência, auto-representação e auto-consciência em sistemas cognitivos físicos, é muito provável que concluamos que não é nada simples ligar aquilo que nos sentimos ser (do que faz parte, de forma importante, para cada um de nós, por exemplo ser-um) como aquilo que sabemos cientificamente acerca do *hardware* e da organização funcional da região espaço-temporal causalmente responsável por isso que é fenomenologicamente, pra nós, ‘sentir-se ser-um’. A própria característica de ser-um, embora central para o que cada nós acredita ser, é muito pouco garantida por várias razões. Não apenas porque não existe aí um centro físico real da actividade cognitiva (por exemplo um ‘centro do cérebro’<sup>254</sup>), sendo o Eu que possibilita a unidade e centralidade da nossa auto-referência uma *representação* e um *efeito* de actividade neuronal acentrada<sup>255</sup>, como também porque a nossa vida mental é *intervalada* (pense-se no sono, no sonho, na amnésia, na hipnose, na anestesia), a cada instante ela própria *acentrada* (‘incluindo’ apercebimentos não conscientes ou não unificados) e sempre mais ou menos *cindida ou cindível* (pense-se em esquizofrenias, paranóias, na desordem de personalidade múltipla, em casos de cérebro dividido, de *akrasia*, de auto-engano e muitas outras patologias da IP). Derek Parfit imagina mesmo, em *Reasons and Persons*, situações de *fissão definitiva*, i.e. situações em que aquilo que é agora uma vida mental pode, no limite, vir a ser bifurcado (i.e. que aquilo que é hoje *alguém* seria de alguma forma psicologicamente continuado por duas pessoas diferentes). A experiência mental do ‘teletransporte que não resulta completamente’ porque a pessoa inicial – aquela que é ‘copiada’, transcrita e enviada para outro ponto do espaço – apesar de tudo permanece, paralelamente, é frequentemente utilizada como ‘alavanca’ para a imaginação desta situação de fissão (fazem-no por exemplo Dennett em *The Mind’s I* e Parfit em *Reasons and Persons*<sup>256</sup>).

<sup>254</sup> Esta é a razão pela qual por exemplo D. Dennett constantemente insiste, nas suas críticas à concepção da mente como Teatro Cartesiano numa crítica à noção de qualia, concebidos como propriedades incorrigivelmente conhecidas da experiência consciente própria [cf. Dennett 1991].

<sup>255</sup> Cf. Dennett 1989, Dennett & Humphrey 1989, Dennett 1991, Dennett 1992.

<sup>256</sup> Dennett, in Hofstadter & Dennett 1981:3-7 e Parfit 1984: 199-200, 75, *Simple Teletransportation and the Branch-Line Case* e em geral o Capítulo 10, *What We Believe Ourselves to Be*.

Tudo isto para dizer que, antes de abordar aspectos mais específicos das questões cognitivas da IP, é conveniente estarmos preparados para o facto de não estar garantida a coincidência entre a perspectiva de nós próprios a partir de dentro e a perspectiva que vamos progressivamente tendo, em terceira pessoa, acerca dos suportes físicos da cognição e da unidade e centração que neles e por eles se realiza, que pode deixar de se realizar ou sofrer transformações estranhas. T. Nagel, por exemplo, há muito insiste que nada garante à partida uma tal coincidência<sup>257</sup>.

Mas exactamente o que podemos saber ‘do ponto de vista da terceira pessoa’ sobre as condições da IP? Em primeiro lugar, que a questão da unificação e da centração do auto-apercebimento (*self-awareness*) é uma questão relativa à organização e às funções de um sistema cognitivo, uma questão que deve ser tratada abaixo do nível da consciência do indivíduo ou sistema em causa. É em grande parte deste modo, i.e. como questão sub-pessoal, que cientistas cognitivos como D. Hofstadter, M. Minsky ou A. Damásio e filósofos como D. Dennett ou O. Flanagan consideram a questão da IP<sup>258</sup>. As funções em causa são responsáveis pelo estabelecimento não apenas de um *sentido de unidade do sistema para o próprio sistema* como também pelo estabelecimento de *um sentido de controlo do sistema por si, na sua globalidade* e não apenas das/nas suas partes, o qual permite a acção global finalista e o comportamento dirigido por fins (*goal directed behavior*)<sup>259</sup>. Tomando como referência o caso humano, o sentido de unidade, de controlo e a acção global finalista envolvem nomeadamente as seguintes condições:

- (1) Um certo tipo de auto-referência no sistema considerado como globalidade, possibilitado por uma representação<sup>260</sup> (o Eu). Um ponto central da teoria cognitiva da IP será portanto ‘explicar o Eu’ e os vários estratos deste (é o que faz cada um dos autores acima referidos).
- (2) Se o Eu permite a auto-referência num sistema cognitivo, convém não esquecer que ‘Eu’ é apenas um indexical desconteúdo. Ora, há alguma coisa mais no auto-apercebimento do sistema, alguma coisa que tem a ver com os ‘materiais’ (mentais ou outros) que são referidos ao longo do tempo pela ocorrência do indexical. Por exemplo nos humanos a situação em foco assume, a partir da

<sup>257</sup> Nagel 1979, *Brain Bisection and the Unity of Consciousness*. O exemplo de Nagel é o cérebro dividido (*split brain*).

<sup>258</sup> Cf. Hofstadter 1979, Minsky 1985, Damásio 1994, Damásio 1999, Dennett 1991, Flanagan 1992.

<sup>259</sup> Cf. Minsky 1985.

<sup>260</sup> Representação no sentido da ciência cognitiva, i.e. estrutura no cérebro ou outro hardware que de uma forma ou outra segue (*tracks*) alguma coisa no mundo.

<sup>261</sup> Dennett 1991.



representação do Eu, a forma de uma narrativa centrada no Eu (na metáfora que Dennett vai buscar à física, o Eu é o ‘centro de gravidade dessa narrativa’<sup>261</sup>). Isto não significa que a narrativa seja o único material que o indexical Eu refere (a propriocepção, o sentir de sensações, o iniciar de acções são outros materiais)<sup>262</sup>.

- (3) A existência de uma auto-representação, de um símbolo-de-si, permite a unificação, a centralização e a dotação global de finalidades pelo sistema a si próprio (aquilo que o sistema total deve fazer), a assim o sentido de alguma entidade ao comando dos comportamentos do sistema<sup>263</sup>. Está-se assim perante uma unificação e uma centração de sistemas físicos que não são à partida ou desde logo unificados e centrados (enquanto corpos biológicos, nomeadamente, os corpos humanos são conjuntos de funcionamentos relativamente autónomos e independentes, e aqui incluem-se mesmo os funcionamentos cognitivos do sistema) cujo estatuto é preciso esclarecer. Uma proposta seria considerar que elas são virtuais (o que não significa irreais). Apenas essa unificação e essa centração permitem a existência das finalidades globais ligadas ao comportamento-dirigido-a-fins (*goal-directed*) e ao auto-controlo centralizado caracteristicamente humano.
- (4) Na medida em que a representação de unificação, que possibilita a auto-referência, embora sendo uma auto-representação, uma representação de globalidade e de unificação, é apenas *mais uma representação* no sistema cognitivo, na vida mental deste, está definitivamente instaurada nos sistemas nos quais existe essa representação uma cisão entre um ponto de vista subjectivo e um ponto de vista objectivador sobre si<sup>264</sup>.

A auto-referência provoca frequentemente paradoxos (pense-se

<sup>262</sup> Cf. Damásio 1999, onde se faz uma interessante distinção entre *proto-self*, *core self* e *extended self*. Fernando Gil (cf. Gil 2000) propõe uma distinção análoga. Uma das intenções de Damásio (Damásio 1999) é distinguir as fundações neurobiológicas do *self* das fundações neurobiológicas do eu autobiográfico. A representação de si que cada um de nós tem teria assim uma versão totalmente inconsciente, ligada nomeadamente à representação do corpo e das várias partes deste no cérebro, uma versão-núcleo, o *core-self*, que inclui já aquilo a que Damásio chama (e acerca do qual avança uma explicação) um *feeling of knowing* e uma versão sofisticada, linguística, o *extended self*. Este último está ligado à reunião das memórias daquilo que acontece, permitindo uma consciência de si propriamente pessoal. Esta IP narrativa é de nível funcional e cortical mais elevado do que o *proto-self* e o *core-self*, no entanto estabelece-se sobre eles.

<sup>263</sup> Como defende Minsky, essas são as funções que o Eu cumpre (Minsky 1985, Chapter 4). O ponto comum às funções é permitir centração num sistema acentrado. Minsky propõe que talvez seja porque não há ninguém dentro da nossa cabeça que nos faça fazer as coisas que queremos, porque constantemente há desunidades, tensões internas, dissensões, que construímos o mito (cognitivamente fundamental) segundo o qual nós ‘estamos dentro de nós’. Noutras palavras, uma das funções do Eu é permitir um (e um só) Ideal do Eu (Minsky fala de *self-images*, *self-ideals*), um auto-controlo global, uma moldagem de si estabelecendo uma mediação entre agentes cognitivos especializados. Outros aspectos da função do Eu seriam impedir-nos de mudar demasiado rapidamente e esconder de nós próprios a natureza dos nossos Ideais do Eu (Cf. Minsky 1985, 4.4. *The Conservative Self*).

<sup>264</sup> Hofstadter 1979, Dennett & Hofstadter 1981.

em paradoxos célebres, como o paradoxo de Russell ou o paradoxo do barbeiro). Esses paradoxos provocam ciclos de auto-referência, anéis estranhos (*strange loops*, como diz Hofstadter<sup>265</sup>). Uma solução possível perante problemas gerados pela auto-referência é a instauração de uma distinção entre meta-linguagem e linguagem-objecto. No entanto, como é óbvio, a solução funciona apenas para linguagens que estão ‘fora de nós’ e que são de alguma forma controláveis, nomeadamente as linguagens formais. Não nos podemos dividir a nós próprios dessa maneira, temos que ‘viver em todos os níveis ao mesmo tempo’ e portanto numa situação ‘paradoxal’ de auto-referência. De acordo com vários autores (por exemplo D. Hofstadter, M. Minsky, D. Dennett) a auto-referência e a interacção entre níveis no cérebro são em grande medida responsáveis por aquilo a que chamamos ‘consciência’ (pelo menos enquanto auto-apercebimento).

- (5) O auto-apercebimento não assume desde logo a forma do uso de um Eu em qualquer sistema cognitivo e nomeadamente nos sistemas cognitivos biológicos. Nestes, o Eu instala-se sobre uma propriedade de si-corpórea, tácita, que não é apenas humana e a que se pode chamar, como em psicologia e em biologia, *self* e que está ligada a algum tipo de representação não linguística do corpo. Como diria A. Damásio, a nossa mente é tal que tem o corpo próprio sempre em mente. É quando passa a haver o Eu que passa a ser possível uma auto-referência não tácita mas explícita, com apoio na linguagem<sup>266</sup>. É no entanto importante notar que tal não significa que acontece uma passagem entre o *self* e o Eu que seja uma passagem para um saber que se sabe neutro, um Eu puro, um Eu-qualquer, uma consciência-de, que de um ponto de vista epistemológico seria neutra. Pelo contrário, a auto-referência possível quando o ‘Eu’ tem usos auto-conscientes, não se separa nunca de um fundo de representação de si. Por exemplo segundo Damásio esse fundo é essencial para a ‘aparição da aparição’<sup>267</sup> (esta é de resto uma tese original, perante o problema da consciência, uma resposta à questão ‘o que é que *subjectiviza* a mentalidade?’, que deve ser comparada com posições funcionalistas mais ortodoxas, como por exemplo as de D. Dennett ou de B. Baars<sup>268</sup>).

<sup>265</sup> Hofstadter 1979.

<sup>266</sup> Cf. Damásio 1999, a análise das diferenças entre *core consciousness* e *extended consciousness*, e entre *proto-self*, *core self* e *autobiographical self*.

<sup>267</sup> Damásio 1999.

<sup>268</sup> Cf. Dennett 1991 e Baars 1988.

Até aqui referiu-se alguns dos problemas cognitivos sub-pessoais relativos à condições de identidade de um dado sistema cognitivo, nomeadamente o auto-apercebimento, a auto-referência, a unificação e a céntracção, a dotação global de finalidades, o relato narrativo de si, a relação entre esse relato e outros materiais que o indexical Eu refere, a cisão entre ponto de vista subjectivo e objectivo sobre si a partir do momento em que uma das representações do sistema é uma auto-representação. Estes problemas são problemas sub-pessoais e dizem respeito a uma ‘forma’ da vida mental humana que não é ela própria objecto de consciência<sup>269</sup>. Ora até certo ponto uma descrição fenomenológica (uma descrição pessoal, cujo objecto se considerará aqui consistir em algo como uma primeira pessoa generalizável) tem uma função análoga ao esboço de descrição cognitiva apresentado. De facto também ela, embora feita em primeira pessoa, ao nível do apercebimento consciente do sistema, pode pretender ser neutra relativamente aos conteúdos de uma vida mental específica. De modo a não incorrer em compromissos de escola procurarei as linhas de orientação de uma tal descrição fenomenológica em William James. O princípio é naturalista: conhecemos tão pouco e tão indirectamente a natureza da nossa mente como conhecemos pouco tudo o resto que através dela se nos apresenta. Por essa razão o inquérito acerca do mental, como todo o inquérito, inicia-se com uma recolha de fenomenologias, no sentido científico corrente. É aqui que se situa o problema a que vou chamar o problema de W. James<sup>270</sup>: o que é ter uma vida consciente normal? Será que há uma resposta para essa pergunta? Será que *isso* é o mesmo em todos nós? A hipótese colocada pelo próprio James é que há alguma coisa que é ter uma vida consciente normal, uma estrutura fenomenológica universal (no sentido de ser comum a todos os humanos adultos conscientes, feita a abstracção das particularidades relativas a cada história mental). James chegou a algumas características dessa estrutura, nomeadamente as seguintes<sup>271</sup>:

- (1) existe um fluxo de consciência, uma sucessividade de estados mentais (o que não implica a existência de um eu como uma coisa a mais, como o autor desse fluxo de estados nem a percepção de um objecto específico que seria o eu);
- (2) há uma adopção do passado (pelo menos imediato) pelo presente do fluxo, que estabelece uma continuidade. Esta adopção do

<sup>269</sup> De resto, e como áparte histórico, as questões dos níveis, aspectos e estratos da representação de si que são hoje tão importantes na filosofia da mente são questões com uma longa história na tradição psicanalítica, que as relaciona de forma importante com os afectos. Essa é uma dimensão que se perde quando a psicanálise é hoje (quase generalizadamente) posta de lado como pouco científica nas discussões da filosofia da mente e da ciência cognitiva. A questão do afecto é normalmente reintroduzida na teoria da mente por aqueles que nunca puderam deixar de a considerar (por exemplo os psiquiatras), ou via a neurociência (pense-se no trabalho de A. Damásio sobre o sentimento de si) e a neuropsicologia.

<sup>270</sup> James 1950 e James 1961.

<sup>271</sup> As descrições que se seguem são livres, não é minha intenção seguir exactamente os textos de James.

passado pelo presente está ligada a um rastro de memória, i.e. há uma memória imediata que ‘faz parte’ do estar consciente, a que os psicólogos chamam memória de curto prazo. Esta memória de curto prazo é suficiente para estabelecer a continuidade do fluxo, mesmo que os conteúdos da memória de longo prazo se percam.

- (3) mesmo que seja possível uma análise lógica dos conteúdos do fluxo, estes não se dão (pelo menos não são experienciados como tal) de forma atómica e composicional, ao contrário do que por exemplo o atomismo empirista clássico poderia pensar acerca das ‘ideias’;
- (4) ocorrem interrupções de pensamento e de sentimento. No entanto, é difícil saber se alguma vez ficamos/somos totalmente inscientes, e também se aquilo que recomeça depois da interrupção é o mesmo fluxo;
- (5) isso que ocorre (o fluxo, os estados de consciência) ocorre em eus pessoais, particulares, separados, é ‘possuído’ pessoalmente, em isolamento e privacidade relativamente a outras consciências ou eus;
- (6) isso que ocorre (os estados de consciência) ocorre sob o modo de uma mudança constante de conteúdos;
- (7) isso que ocorre aparece como contínuo e continuado, i.e. mesmo quando há intervalo a união parece ser retomada, e as mudanças qualitativas nunca são totalmente abruptas (nas palavras de James «*it feels unbroken*»<sup>272</sup> e não «*jointed*» ou «*chopped up in bits*»);
- (8) isso que ocorre envolve não apenas conteúdos explícitos mas também margens (os termos usados por James são *fringe*, *psychic overtone*). Noutras palavras, há um halo de relações que envolve os conteúdos, i.e. ‘ressonâncias inarticuladas’ são parte integrante de cada vida mental;
- (9) o Eu é um modelo da vida mental na própria vida mental. Isto significa que não há aí um eu ‘anterior’ à vida mental (James propõe que os pensamentos são eles próprios os pensadores<sup>273</sup>);

Mudando de perspectiva e passando para a filosofia da mente contemporânea de modo a estabelecer o ponto que aqui se pretende, é fácil verificar que a intenção de um modelo da consciência como por exemplo o Modelo dos Esboços Múltiplos de D. Dennett<sup>274</sup> é fazer a articulação entre as funções, os mecanismos cognitivos sub-pessoais, considerados de

<sup>272</sup> James 1950: 238.

<sup>273</sup> James 1961: 83.

<sup>274</sup> Dennett 1991.

um ponto de vista funcionalista e a fenomenologia, mostrando assim que o abismo nageliano entre a fisiologia e a fenomenologia<sup>275</sup> pode e dever ser preenchido. Esta pretensão é totalmente independente por exemplo do tratamento dos *qualia* por Dennett (um tratamento objectável por inúmeras razões e sobretudo por assentar numa petição de princípio quanto à consciência fenomenal). Mas de resto, este modelo, ou outro com intenção semelhante, diz respeito a um coeficiente específico de ‘conteúdo’ da forma humana de ser consciente que é devido a condições cognitivas sub-pessoais. Evidentemente, ao mesmo tempo modelos deste tipo ilustram as articulações onde podem instalar-se patologias da consciência.

Note-se que relacionando o primeiro ponto (a descrição sub-pessoal de uma mecânica do auto-apercebimento, da auto-referência, do controlo centrado de si, num sistema que funciona como se fosse unificado e centrado) com o segundo ponto (a caracterização da fenomenologia) destas questões cognitivas obtém-se por exemplo que o problema do estatuto do Eu no fluxo da consciência é um problema específico dentro do problema geral da vida mental. Por exemplo de acordo com o Modelo dos Esboços Múltiplos de Denett a representação de unidade centração e controlo tem um estatuto *virtual*. Realidade virtual não é irrealidade, e a realidade virtual do Eu resulta de um auto-apercebimento (o que mostra que ao contrário do que é normalmente dito, inclusive pelo próprio Dennett, o Modelo dos Esboços Múltiplos não põe totalmente de lado o Teatro Cartesiano na concepção da mente: o Modelos dos Esboços Múltiplos não pode abdicar do auto-apercebimento (*self-awareness* com uma forma específica), o que não significa de modo algum incorrigibilidade epistémica no acesso aos conteúdos conscientes. Esta ideia acerca de uma unidade virtual que apoia o sentido de controlo centrado, o sentido da agência, e que pode claramente não ocorrer de ‘forma normal’ dando origem às várias desordens da mesmidade pessoal, com as desposseções e des-unidades, observações de si como outro, etc, que estas envolvem, é uma ideia de ciência cognitiva e que deve ser avaliada como tal.

### **3. Questões metafísicas: Locke, Descartes, Hume, Kant, Parfit e as pessoas.**

Vou partir do princípio de que a questão metafísica da IP não é uma questão sub-pessoal acerca de modelos cognitivos nem directamente uma questão sobre como é, fenomenologicamente, sentir-se ser. O que está em causa na questão metafísica é saber que tipo de entidades no mundo são as

<sup>275</sup> Nagel 1979c.

peçoas. Por vezes a questão assume a forma de procura da marca distintiva desse particular tipo de entidades, dos traços pelos quais elas podem ser reconhecidas. Têm sido propostos traços tais como a auto-consciência, a vontade livre, a existência de estados mentais que são acerca de outros estados mentais<sup>276</sup>, a existência de volições de segunda ordem<sup>277</sup> (i.e. a possibilidade de se determinar a querer ou não querer aquilo que se deseja). Sempre subjacente à discussão está evidentemente o confronto entre as concepções a que no início deste artigo se chamou concepções absolutistas e concepções não absolutistas de pessoa. Vale a pena fazer uma digressão histórica de modo a procurar compreender a partir de J. Locke o que se entende por pessoa de um ponto de vista metafísico numa concepção não absolutista, já que Locke é uma referência (absoluta...) nesta discussão.

Essencialmente Locke pensava que a Identidade de uma pessoa não consistia nem na identidade de uma entidade imaterial (uma alma) nem na identidade de um corpo animal (humano), mas sim na mesmidade de uma consciência ao longo do tempo, i.e. na continuidade psicológica, na ligação e sucessividade do acontecer mental através da memória. Isso permitiu-lhe fazer uma distinção, muito influente de então para cá, entre ‘humano’, uma classificação de espécie biológica aplicada a determinados indivíduos, que não são chimpanzés ou babuínos, e ‘pessoa’, termo que nomeia algo de mental. Segundo Locke, o termo ‘pessoa’ nomeia um ser pensante e inteligente que possui razão e reflexão e que é capaz de se considerar a si próprio como si próprio e como uma mesma entidade pensante em diferentes tempos e lugares (*Ensaio Sobre o Entendimento Humano*, Livro II, xxvii, 9)<sup>278</sup>. A isto Locke acrescentava que ‘pessoa’, ao contrário de (ou por contraste) com ‘humano’ é um termo forense, que avalia acções e o mérito destas e que pertence ou se aplica apenas a agentes capazes de lei, felicidade e infelicidade (as palavras de Locke são ‘*law, happiness and misery*’).

Como se notou no início do presente artigo, é claro que a preocupação especial com o que são pessoas é em grande parte devida ao envolvimento da noção em práticas jurídicas e morais (por exemplo apenas faz sentido considerar uma pessoa responsável por um crime se ela recorda tê-lo cometido). Aliás, no caso de Locke havia também um agenda teológica, relativa a responsabilidade e ressurreição. Sublinhe-se que Locke não rejeitava a existência de uma substância imaterial ‘em cada pessoa’, discernida por ser ‘o-que-pensa’ quando existe pensamento. O que Locke

<sup>276</sup> Dennett 1981.

<sup>277</sup> Frankfurt 1971.

<sup>278</sup> Locke 1999.

rejeitava era que ‘ser uma mesma pessoa ao longo do tempo’ requeresse a identidade de uma substância imaterial, i.e. que fossem necessárias almas para explicar a Identidade Pessoal. Se a consciência fosse identificada com a alma e com a nossa essência, essa essência poderia estar ausente (por exemplo durante o sono) e de acordo com Locke essa concepção era indefensável, pois a ausência da essência de uma entidade tornaria impossível que essa entidade continuasse a ser a mesma entidade.

Os críticos contemporâneos de Locke, por exemplo J. Butler e T. Reid criticaram na teoria lockeana por um lado aquilo que entenderam ser uma circularidade (se o critério da IP é a continuidade psicológica, a memória, esse critério é circular, pois cada indivíduo recorda apenas as suas próprias memórias), a possibilidade de transferência de consciência de um corpo para outro corpo, e a possibilidade de interrupção da identidade quando a memória é perdida (uma pessoa que não se recorda de episódios biográficos seus não é a mesma pessoa).

É verdade que a concepção lockeana de pessoa tem algo de nominalista: por exemplo e ao contrário da noção lockeana de pessoa, o cogito cartesiano garante a cada indivíduo consciente imediatamente a ‘essência de pessoa’. Em contrapartida, a essência consciente do cogito cartesiano não parece deixar grande espaço para distinguir os indivíduos uns dos outros – afinal, se a essência do ser pensante é a consciência, o que nos separaria fundamentalmente uns dos outros enquanto pensantes, o que individualizaria a nossa essência pensante?<sup>279</sup> A definição cartesiana de mente como consciência, i.e. como substância distinta do mundo físico, não espacial, que tem uma relação especial com as ideias (definidas como algo que está de tal modo na consciência que é imediatamente apercebido) no seio de uma metafísica trissubstancialista é origem de um problema grave para a identidade pessoal, um problema que terá que ser resolvido na antropologia cartesiana, dando razões para a individuação de algo que não teria por que, aparentemente, por essência, ser individualizado. Por definição, substância é o que pode ser pensado separadamente, pensado por si, enquanto sendo de uma natureza específica. Ora, o facto de a natureza de pensamento poder ser pensada separadamente não faz dela pensamento individualizado. O resultado do cartesianismo é portanto não apenas a desnaturalização da consciência e a instauração de uma descontinuidade abrupta entre consciência e não consciência, como também uma certa desindividuação do mental enquanto consciência. A partir da definição de mental como consciência imaterial basta um passo até à ideia segundo a qual o espírito poderia existir não apenas sem o corpo como também sem as pai-

<sup>281</sup> *Descartes encontra-se perante um ‘problema averroísta’.*

xões derivadas da união com o corpo, podendo em última análise e idealmente o espírito ser consciente sem sentir (a intelectualização da noção de mente é além do dualismo um dos erros de Descartes<sup>280</sup>). O problema seria saber se esse espírito seria ainda individual, pessoal. De acordo com Descartes, apenas as ideias e as vontades, que não são provocadas por nada de exterior à consciência, são modos puros da *res cogitans*. E do ponto de vista metafísico está dada a última palavra quanto à natureza do espírito. Alguma coisa tem que ser feita no entanto para reintroduzir a individuação do pensamento e essa é uma tarefa para a antropologia cartesiana, em torno da ligação de uma alma a um corpo próprio, da teoria das paixões (nas *Meditações sobre a Filosofia Primeira* o ‘mergulho no mundo’ da 6ª Meditação, após a descoberta da essência do espírito e da essência da matéria, constitui um pontos focais para a procura da individuação do espírito). No entanto, e apesar da antropologia da mistura corpo/alma e da justificação da natureza composta do humanos por um propósito prático<sup>281</sup>, de acordo com os princípios da metafísica cartesiana a individuação do espírito não pode senão ser muito pouco essencial (dir-se-ia aliás que toda a antropologia não pode senão ser muito pouco essencial para Descartes). Essencial é unicamente a característica de ser pensante do ser pensante, ligada à imaterialidade, que nada individua necessariamente, do *cogito*.

Se Locke é a referência das discussões sobre identidade pessoal, Descartes é o alvo da maior parte dos ataques, que se dirigem não à particular solução cartesiana do problema da individuação das pessoas na sua antropologia mas sim aos ‘Egos Cartesianos’. Os Egos Cartesianos, puros e imateriais, são o inimigo imaginário das discussões filosófica contemporâneas da identidade pessoal. Se Egos cartesianos existissem, eles seriam: (1) distintos por essência de um corpo no mundo mesmo que a ele estivessem ligados (2) separados e individuados como objectos físicos, embora de natureza totalmente mental (3) suficientes só por si para explicar a co-personalidade das várias experiências e a unidade da consciência. Se eles não existem há todo um programa de explicação que se abre e que tem por finalidade explicar o que um ego cartesiano explicaria se existisse.

Farei apenas mais duas referências a textos clássicos, que comprovam nomeadamente o cruzamento da questão metafísica da natureza das entidades com a questão epistemológica (entendida como uma questão acerca do que é conhecido quando é conhecido um eu, uma estabilidade individual pessoal). Por contraste com o essencialismo imaterialista (e

<sup>279</sup> Damásio 1994.

<sup>280</sup> De facto, no contexto da ‘reinstalação’ do espírito no mundo na 6ª Meditação, Descartes evocará a fiabilidade da informação perceptiva e um sistema de manutenção do corpo que assegura a sobrevivência. Tudo isso funciona bem mas não é fonte de saber. Não se trata de fiabilidade epistemológica: nenhuma ideia corporeamente baseada é essencial à mente enquanto mente.



em última análise anti-individualista) cartesiano, e de forma semelhante a Locke, também D. Hume desenvolveu uma teoria tendencialmente reducionista da identidade pessoal, afirmando de forma célebre que por mais que procurasse dentro de si não encontrava coisa alguma a que chamar um *self*, mas apenas percepções, sempre várias, variadas e mutáveis (*Treatise of Human Nature*, Book I, Part IV, VI, *Of personal identity*). De acordo com a visão humeana do Eu evasivo, nunca encontrado, uma pessoa é apenas uma série de experiências de alguma forma ligadas. Hume atribui, como se sabe, a estabilidade da identidade de cada ser pensante (como de resto a estabilidade de outras identidades no mundo) ao trabalho da mente, i.e. da imaginação. Uma pessoa pensada por si própria como entidade individual é assim, para Hume, uma ficção sobre um feixe de impressões.

A primeira coisa a notar relativamente à agenda epistemológica do problema é que o estatuto cartesiano de privilégio do conhecimento de si (ao conhecer o espírito e no espírito conhece-se *essências*) desaparece imediatamente nestas circunstâncias humeanas. Sei tão pouco o que eu sou, o que a minha mente é, como sei pouco acerca do que o resto do mundo é. Hume merece nesse aspecto ser considerado como um precursor da abordagem a que hoje se chama abordagem naturalista do mental. Do ponto de vista do conhecimento, é isto que Hume tem a dizer sobre a identidade pessoal: as ideias básicas são o feixe de impressões e a ficção de unidade. Do ponto de vista da acção, das paixões e dos sentimentos morais, a teoria assume outro aspecto, que não se desenvolverá aqui<sup>282</sup>.

Da mesma forma que Hume, Kant defendeu que a razão humana nada pode demonstrar acerca da substancialidade da alma individual. Não é possível conhecer a alma como objecto. É no entanto uma fundamental e natural ilusão pensar que sim: tomar a simplicidade lógica do pensamento do eu por conhecimento de uma alma simples, imaterial, imortal, é para Kant, um exemplo de paralogismo, de incorrecção de raciocínio, de raciocínio que vai para além da experiência possível (*Crítica da Razão Pura*, Dialéctica Transcendental, Livro Segundo, Dos raciocínios dialécticos da razão pura, Cap.I, Dos paralogismos da razão pura, Primeiro paralogismo (da substancialidade)). Sendo assim, evidentemente, o eu-penso, que de

<sup>282</sup> A ideia de F. Gil segundo a qual a primeira evidência alucinada é precisamente a da nossa identidade [Gil 2000] é de certo modo humeana, no sentido de atribuir à identidade um estatuto de ficção natural. As análises de F. Gil têm ainda a vantagem de visar o aspecto afectivo da relação do eu com o eu, nomeadamente com a vida mental própria. Repare-se que devido à relação da identidade pessoal com o tempo [a identidade pessoal não é uma relação formal mas uma continuidade ou persistência ao longo do tempo] a identidade pessoal, com toda a sua ficcionalidade, virtualidade e indeterminação funciona como uma barreira contra o fim, o que é evidentemente importante considerar quando se analisa a sua função afectiva, aquilo a que F. Gil chama a adesão a si. De facto, ao contrário do que acontece com a identidade formal, na identidade pessoal as questões do princípio e do fim importam. A história a que chamo minha começa e acaba, de uma forma que não tem paralelo com o destino de  $7+5=12$ .

acordo com a teoria do conhecimento kantiana acompanha todas as representações, não é (à partida, desde logo, alguma vez ou tanto quanto a estrita teoria do conhecimento permite afirmar) uma alma-substância. A chegada à natureza de alma imortal e de liberdade que Kant supõe corresponderem a cada pessoa só pode portanto fazer-se por outro caminho, o caminho do pensamento moral, que parte da experiência do dever.

Transpondo para o nosso tempo a discussão acerca da metafísica da identidade pessoal, tomar-se-á em seguida como referência uma das mais conhecidas e discutidas concepções reducionistas actuais da identidade pessoal, a de Derek Parfit<sup>283</sup>. Parfit defende o seguinte acerca de pessoas e de identidade pessoal. Embora não sejam Egos Cartesianos, as Pessoas não são (no sentido *identificativo* de ser) corpos nem tão pouco séries de experiências (como um seguidor de Hume ou de Buda poderia defender). As pessoas *têm* corpos, pensamentos e experiências. Simplesmente, a existência de pessoas *consiste* na existência de um corpo e na ocorrência de eventos mentais e físicos inter-relacionados. Este é o núcleo do *reduccionismo constitutivo* de Parfit. O reduccionismo constitutivo acerca de pessoas opõe-se ao reduccionismo identificativo e ao reduccionismo eliminativo.

De acordo com o reduccionismo constitutivo acerca de pessoas (a posição de Parfit), *Embora Xs (pessoas) sejam diferentes de Ys (corpos, eventos mentais e físicos), a existência de Xs simplesmente consiste na existência de Ys*. Pelo contrário, de acordo com o reduccionismo identificativo, o ‘é’ de identidade ou instanciação entre X e Y assinala que X e Y são uma e a mesma coisa, que tudo o que é verdadeiro acerca de X é verdadeiro acerca de Y, não havendo qualquer necessidade de distingui-los (por exemplo ‘O cometa Haley é um pedaço de gelo’). Quanto ao reduccionismo eliminativo, quando este é justificado, uma determinada entidade nomeada não existe: um exemplo frequentemente utilizado por Parfit é o do Americano Médio (*average american*), por exemplo em ‘O americano médio tem 2 filhos’. O Americano Médio não existe: as asserções acerca do americano médio são asserções acerca de americanos particulares reais. Mas os casos em que a eliminação é a melhor forma de exprimir o reduccionismo são raros. Por exemplo em relação à questão do reduccionismo relativo a pessoas a possibilidade de eliminação coloca-se de forma muito diferente do que acontece com o Americano Médio. Parfit perguntaria por exemplo a alguém que falasse de constituintes últimos do mundo, em termos directamente ontológicos, à maneira de Quine, como relativos àquilo com que uma teoria se compromete ‘E nesse caso você pensa que não existe?’. Parfit pensa que não tem sentido responder ‘Sim, eu penso que eu não existo’

<sup>283</sup> Outros autores que discutem questões semelhantes são por exemplo Sidney Shoemaker, John Perry, Judith Thomson, Bernard Williams, Thomas Nagel, John McDowell, David Wiggins, Eric Olson.

pelo facto de seja quem for que fale decididamente não ser um desses constituintes últimos. De facto, de acordo com o reducionismo constitutivo de Parfit as pessoas definitivamente existem e é importante poder distinguir pessoas de corpos e de séries de experiências. O que interessa saber é *como* é que pessoas existem.

Utilize-se uma analogia. Num exemplo como ‘A Vénus de Cellini é um pedaço de ouro’ as insuficiências quer do reducionismo identificativo quer do reducionismo eliminativo são visíveis. A estátua, embora ‘feita do ouro’, consistindo em ouro, é distinta desse ouro em que consiste, o que faz com que seja uma entidade separada. A estátua e o ouro são distintas na medida em cada um/uma pode por exemplo ‘durar’ mais do que o o outro/outra (por exemplo derretendo o pedaço de ouro destrói-se a estátua mas não o ouro, e derretendo o ouro do interior da estátua, deixando apenas a superfície da estátua e substituindo o ouro interior por outro material, destrói-se o ouro mas não a estátua). O caso não é semelhante nem à identificação do cometa Haley com um pedaço de gelo nem à ‘eliminação’ do Americano Médio.

Note-se que pelo facto de considerar que pessoas existem de uma forma que não é propícia à eliminação ou à identificação com um corpo animal humano (ou parte suficiente deste) um reducionista constitutivo como Parfit não pretende imediatamente que as experiências requerem sujeitos, e que estes sujeitos ou pessoas ficam assim imediatamente ‘provados’. Precisamente, a intenção é oferecer uma teoria do tipo de entidades que têm corpos e experiências e nomeadamente explicar a unidade da vida mental destas entidades através do tempo, em lugar de a pressupor. Como se afirmou, Parfit pensa que pessoas são distintas dos seus cérebros e corpos mas não existem separadamente, como entidades independentes. A identidade de um pessoa ao longo do tempo consiste em algum tipo de continuidade física e/ou psicológica. Isto significa que a identidade pessoal é distinta dos factos acerca destas continuidades mas não um facto independente. A partir do momento em que se descreveu todos os factos relativos a corpos e cérebros está descrito tudo o que há a descrever. Noutras palavras e assumindo e um ponto de vista epistemológico, descrever a existência de pessoas não é descrever factos a mais (*further facts*).

Como é de prever, a primeira coisa que preocupa Parfit na discussão da metafísica da identidade pessoal é a especificação do ‘consiste em’, de modo a mostrar que a sua posição acerca de pessoas não é eliminativista (como já foi acusada de ser, nomeadamente devido à possibilidade prevista em *Reasons and Persons* de uma descrição impessoal dos factos<sup>284</sup>, i.e.

<sup>284</sup> Parfit 1984.

uma descrição do mundo tal como ele é em que não aparecem pessoas, em que as experiências são descritas como ocorrendo e não como sendo ‘possuídas’ por alguma entidade). Quando Parfit afirma que factos constituem outros factos quer dizer que determinados factos fazem com que determinados outros factos obtenham, de uma maneira não causal (o ouro do exemplo da estátua não *causa* a estátua).

Com ‘consiste em’ Parfit quer, então, dizer: (1) ser necessário e suficiente para, (2) não haver mais factos (o que não significa que os factos descritos sejam os mesmos factos) e (3) uma dependência assimétrica (não relacionada com causalidade).

Várias consequências interessantes decorrem do reducionismo constitutivo de Parfit. Nomeio quatro:

- (1) Se este tipo de reducionismo é verdadeiro a identidade pessoal de cada um de nós não é determinada (embora nós tenhamos a tendência irresistível a pensar que a nossa identidade é determinada, que por exemplo num instante de tempo futuro ou seremos ou não seremos esta pessoa),
- (2) A co-pessoalidade das experiências (noutras palavras, a unidade da vida mental) não é auto-explicativa. Parfit, repita-se, parte do princípio de que a unidade das experiências de uma pessoa e a unidade da pessoa ao longo do tempo não pode ser explicada adscrevendo essas experiências a uma unidade, unidade essa que seria a pessoa, pois essa unidade não pode ser pressuposta. A unidade tem que ser explicada evocando a forma como as experiências se relacionam entre si e nomeadamente o cérebro da pessoa em causa: é isso que significa afirmar que a co-pessoalidade das experiências consiste noutros factos,
- (3) Porque consistem noutros factos, a existência de pessoas e a Identidade Pessoal ao longo do tempo não podem por si ter efeitos,
- (4) Uma (talvez) inevitável ilusoriedade da nossa auto-concepção: embora de um ponto de vista teórico muitas pessoas não estejam dispostas a afirmar que são Egos Cartesianos, elas não têm (nós não temos) outra maneira de pensar em si próprias a não ser como uma unidade e co-pessoalidade de experiências que se mantém ao longo do tempo. Assim, para Parfit o normal é termos crenças inconsistentes acerca da nossa própria natureza, fazendo constantemente suposições que o reducionismo proíbe (nomeadamente supondo que a nossa identidade deve ser determinada).

Note-se que nada nas posições de Parfit nos obriga a crer que pessoas sejam entidades conceptuais, que vêm à existência através do uso do conceito de pessoa. A existência de pessoas não é uma existência que

dependa de conceitos, nesse sentido, a discussão de Parfit não é acerca de conceitos mas acerca do tipo de realidades que são as pessoas. Todas as consequências nomeadas são importante para a ética, nomeadamente como ponto de partida de uma ética não religiosa, ou pelo menos que não pressuponha a determinação e o suporte da identidade pessoal, um tipo de ética que segundo Parfit, está ainda nos seus primórdios. Não se avançará aqui com análises das posições de Parfit acerca da metafísica da identidade pessoal. Parfit pretende defender as suas teses perante outras posições reducionistas e prolongá-las no âmbito da teoria da acção. É suficiente notar que a questão metafísica da natureza das pessoas tem uma relação directa com a questão prática das razões para agir. Aquilo que as pessoas são ou não são importa quando se trata de agir racionalmente, em função de um cuidado ou preocupação (*concern*) tanto quanto possível justificado, ou pelo menos não auto-contraditório. Se a identidade pessoal é bem menos básica e determinada do que aquilo que usualmente pensamos, então fica posta em causa a racionalidade do egoísmo racional que é, de uma forma ou outra quase consensualmente considerado como a pedra de toque da racionalidade na acção. Em *Reasons and Persons* a racionalidade do egoísmo racional como teoria da acção era já posta em causa (essa é de resto a articulação maior da questão das Pessoas com a questão das Razões para a acção em *Reasons and Persons*). De facto, a questão que se coloca é: porque é que havemos de estar especialmente preocupados com o nosso próprio futuro? Se de acordo com Parfit a identidade pessoal consiste noutros factos, a sua importância é derivada e a porta fica aberta para a defesa da pertinência do ponto de vista impessoal do utilitarismo. A metafísica da identidade pessoal é assim continuada por Parfit com uma filosofia moral, des-pessoalizada, assentando sobre a não importância (*unimportance*) da Identidade Pessoal.

Note-se que Parfit introduz uma inflexão importante na discussão da metafísica da identidade pessoal ao desviar a atenção da natureza da identidade pessoal para a importância da identidade pessoal, perguntando abertamente: será que a identidade pessoal importa? Por que razão cada um de nós há-de ter uma preocupação especial com o seu próprio futuro? O que é que faz com que sobreviver seja bom? Não será preferível pensar exactamente o que é que queremos que sobreviva? Mesmo se quase todos nós quase todo o tempo queremos, mesmo que obscuramente, continuar a ser, pode ser que isso não seja o que mais importa sobretudo se a natureza da identidade pessoal não coincide com aquilo que normalmente e ilusoriamente dela pensamos. Por exemplo relacionando a continuidade normalmente evocada acerca da identidade, é fácil verificar que temos muito mais continuidade física com o corpo a que agora chamamos nosso

do que aquela que nos importa, no sentido de a desejarmos (nomeadamente este corpo vivo é fisicamente contínuo com o nosso futuro cadáver, com os últimos fragmentos desagregados daquilo que hoje o constitui mas não parece que isso importe muito, talvez nem sequer para aqueles que esperam a ressurreição dos corpos). Por outro lado a continuidade que importa persiste com menos do que a totalidade do corpo e mesmo com menos do que a totalidade do cérebro. É possível que também a continuidade psicológica seja menos importante do que o que usualmente pensamos. Assim sendo, dada a não importância da identidade (*the unimportance of identity*, como diz Parfit), aquilo que satisfaria uma pessoa como preservação daquilo que importa, aquilo que faria com que o futuro fosse bom, pode perfeitamente não coincidir com a continuação de si, pode ser bastante mais alargado, mais neutro. Em suma, pode haver um outro tipo de prolongamento daquilo que importa que não inclui a preservação daquele indivíduo e a continuidade da sua vida mental. A questão acerca do que importa (*what matters*) seria assim bem mais básica do que a questão da identidade pessoal, nomeadamente no que respeita a decisões quanto ao futuro pessoal e social.

#### **4. Questões narrativas. Ser ou não ser um indivíduo: gastar palavras em idiossincrasias e contingências. O que é que (me) satisfaz.**

Para além da Unidade e do Centro que pensamos ser, deixando em suspenso a questão acerca daquilo que metafisicamente somos (e que podemos não vir nunca a saber que somos), nós pensamos ser (de facto pensamos ter) uma história pessoal construída em torno de tal Unidade e de tal Centro. Uma história pessoal é uma narrativa, uma agregação mais ou menos estruturada do decurso de uma vida a que chamamos nossa e na qual somos a personagem principal. Esta narrativa não se identifica com o cérebro do corpo a que chamamos nosso, embora este seja causalmente responsável por mantê-la. Trata-se antes de uma criação que acontece sobre essa base. O Eu, que do ponto de vista cognitivo merece ser considerado uma centração virtual, é o 'centro de gravidade narrativo' de uma tal história, que é criada de forma involuntária, i.e. é causada por processos que não são eles próprios nem inteligentes, nem conscientes, nem voluntários, pois são de uma certa perspectiva processos sub-pessoais no nosso cérebro (tanto quanto sabemos aquilo que causa o nosso auto-apercebimento corpóreo e narrativo é um cérebro, no entanto esse saber será sempre indirecto: nunca vimos o nosso próprio cérebro e nunca o veremos sem mediações).

Na intersecção das questões cognitivas e narrativas da identidade pessoal temos portanto a seguinte situação: uma pessoa diz 'Eu sou a Sofia'

ou ‘Eu sou o António’ e essa pessoa é como uma máquina que escreve romances<sup>285</sup>. É esta dimensão narrativa da identidade pessoal e o problema dos materiais conceptuais da narrativa que opera a ligação entre a questão cognitiva e a questão social, política, artística, etc, das ‘identidades’, pensadas como aquilo que define um indivíduo como sendo esse indivíduo específico, com determinadas características que o ligam por exemplo a comunidades e épocas históricas. Esta criação, que é de certo modo uma auto-criação, e que é, obviamente, um material importante de psicólogos, psiquiatras e outros profissionais da Identidade Pessoal, é o núcleo das questões narrativas da identidade pessoal. Não se trata obviamente de uma criação *ex-nihilo* de algo de fisicamente novo, mas da criação, sempre em curso, de uma narrativa de si, condição que é parte importante do facto de a IP não ser dada mas perseguida e conseguida ou não conseguida, num processo em que cada um se toma a si próprio como matéria susceptível de moldagem e de mais ou menos posse<sup>286</sup>. É certo que esta forma de colocar a questão da identidade pessoal tem relações complicadas com a questão metafísica da identidade pessoal, que frequentemente tem por trás uma agenda epistemológica e metafísica ‘descritivista’ (i.e. trata-se de discernir aquilo que é, a natureza de algum tipo de entidade e não de capturar o aspecto dinâmico desse tipo de entidade).

O estatuto da narrativa de si pode ser concebido de maneiras muito diferentes. Antes de considerar esquematicamente certas concepções alternativas do estatuto da narrativa de si procurarei caracterizar o teor da narrativa de si:

- (1) O sentido narrativo de si, a identidade pessoal que cada um tem que fazer para si próprio, traduz-se em grande medida num sentido idiossincrático daquilo que é importante e possível e que faz com que um Eu difira de outros eus. É em grande parte esse sentido idiossincrático daquilo que é importante e possível cuja extinção se teme quando se teme a morte (D. Parfit defende algo de semelhante, embora a sua concepção reducionista da identidade pessoal lhe permita ‘separar’ ‘aquilo que importa’ do indivíduo e da extinção deste).
- (2) Evidentemente um tal sentido daquilo que é importante e possível provem de materiais vários, contingentemente disponíveis e que podem ser tomados por muitos eus. Daí que um outro aspecto presente nesta narrativa de si, ou que a condiciona, seja,

<sup>285</sup> Cf. por exemplo Dennett 1989, Dennett & Humphrey 1989, Dennett 1991 e Dennett 1992.

<sup>286</sup> É a este sentido de produção narrativa de si que nos referimos quando dizemos que a identidade ‘faz-se, faz-se e nunca se vê nada pronto’. Este era outro dos temas organizadores da mesa redonda do Professor Vítor Oliveira Jorge.

utilizando a expressão de H. Bloom, a ansiedade da influência<sup>287</sup>. Bloom aplica, como se sabe, o conceito de ansiedade da influência ao criador artístico original. No entanto, o conceito aplica-se identicamente à narrativa psicológica de si que estou a considerar. A ansiedade da influência é o temor de ser apenas uma cópia, uma réplica, o temor de que a originalidade e a individualidade (aquilo que se perde com a extinção daquele indivíduo, daquele poeta, nos casos de Bloom) não sejam reconhecidas.

- (3) Impõe-se saber exactamente que originalidade é essa a que um indivíduo pretende, se nenhum indivíduo se cria totalmente a si próprio, a partir de zero. Este é o ponto em que as questões da auto-criação e do auto-consciência se cruzam, aparecendo a consciência dos limites da auto-criação, i.e. a consciência das razões pelas quais a auto-criação não é nunca total ou totalmente voluntária. Esta dimensão da identidade pessoal é tratada na filosofia sob muitas formas. Dou apenas alguns exemplos: os limites da auto-criação são tratados por exemplo por T. Nagel<sup>288</sup> e por B. Williams<sup>289</sup> através do conceito de 'sorte moral' (*moral luck*), por H. Frankfurt através da análise das relações entre a liberdade e o conceito de pessoa<sup>290</sup>, por Dennett através do reconhecimento dos limites cognitivos das escolhas morais<sup>291</sup> e através da ideia de condições de pessoalidade (*conditions of personhood*) 'normativamente' extraídas do funcionamento de Sistemas Intencionais<sup>292</sup>, por Nietzsche através do reconhecimento da contingência inscrita na ideia de *amor fatum*, pelos vários existencialistas através das análises da liberdade, etc.
- (4) Em contraste com as dimensões cognitiva e metafísica, é ao nível das questões narrativas da identidade pessoal que a questão da sua própria natureza é encarada pelo sujeito de forma directa (i.e. o sujeito lida com questões relativas à voluntariedade e compromisso

<sup>287</sup> Bloom 1963.

<sup>288</sup> Nagel 1979.

<sup>289</sup> Williams 1981. O conceito foi proposto por Williams e no seu núcleo está aquilo que escapa à caracterização de uma situação moral como relativa a decisões racionais, como o carácter do agente e as circunstâncias da acção. Em geral, o conceito de sorte moral pretende chamar a atenção para o facto de que aquilo que alguém faz depende de factores que estão totalmente fora do seu controlo (por exemplo uma pessoa terá ou não cometido homicídio conforme a pessoa sobre quem disparou estivesse ou não a usar um colete à prova de bala).

<sup>290</sup> Frankfurt 1971. O núcleo da questão da 'relação entre a liberdade da vontade e o conceito de pessoa' reside nas Volações de Segunda Ordem. Estas são, segundo Frankfurt, fulcrais no conceito de pessoa. A ideia de Volações de Segunda Ordem pretende nomear a possibilidade humana de querer ou não querer aquilo que se deseja. Apenas nestas circunstâncias surge a responsabilidade pelo que se é e pelo que se quer.

<sup>291</sup> Dennett 1988. A consideração de todos os factores que pesam sobre uma decisão, implícita nas éticas da maximização, como o utilitarismo e o kantismo, é computacionalmente impossível. A limitação cognitiva, a impossibilidade de considerar todos os factores, é constitutiva da forma humana de decidir.

<sup>292</sup> Cf. Dennett 1981, *Conditions of Personhood*.



com a sua própria identidade e racionalidade<sup>293</sup>). O tipo de auto-consciência ou auto-apercebimento (*self-awareness*) aqui em causa não é assim simplesmente o apercebimento de um si prévio, bruto, sub-pessoal ou a descoberta acerca de uma determinada natureza metafísica no mundo mas sim envolvimento prático da entidade que se sente e se pensa num dinamismo de auto-criação e de auto-avaliação<sup>294</sup>.

- (5) A auto-avaliação que se joga na criação de si não é apenas uma avaliação kantiana ou utilitarista daquilo que se faz mas uma avaliação daquilo que se é, uma dimensão frequentemente ocultada na filosofia moral pela exclusividade das análises deontológicas e utilitaristas. A avaliação forte, qualitativa ou reflexiva, na terminologia de C. Taylor<sup>295</sup>, característica de seres que cuidam do seu ser, que o determinam e avaliam é uma das razões pelas quais as questões da identidade pessoal não podem ser identificadas exclusivamente por exemplo com a efectuação da continuidade, da unidade e da centração que são em grande medida questões mecânicas, sub-pessoais, despossuídas, nem com a avaliação associada ao *ranking* de preferências para o agente (suposto como já constituído). Charles Taylor desde há muito faz notar que a ausência de sustentação revelada pela contigência desta auto-criação e auto-avaliação é difícil de encarar (mas afirma também apenas a capacidade de encarar essa ausência de sustentação possibilita as ‘pessoas profundas’).

Se a auto-criação não é feita a partir de zero, se os seus materiais têm origens e marcas, em grande parte saber o que é distinto e específico naquilo que se é é ter compreendido tais marcas na identidade pessoal própria, dar um determinado estatuto à narrativa de si. Ora, é precisamente nesse ponto que perspectivas filosóficas totalmente opostas se combatem. De acordo com a posição tomada será mais ou menos satisfatório para um indivíduo ser um indivíduo (i.e. apenas um, apenas este, apenas alguma coisa que não dura muito tempo). Nomeadamente, de acordo com muitas perspectivas filosóficas (de resto muito diferentes entre si<sup>296</sup>) apenas o universal satisfaz.

De acordo com uma perspectiva – a que vou por conveniência chamar nietzscheana – da identidade pessoal como auto-criação, a auto-

<sup>293</sup> Estas são questões tratadas por teóricos quinianos do mental como Dennett e Davidson quando procuram caracterizar a auto-concepção e o auto-conhecimento. Cf. Dennett 1981 e Davidson 1982.

<sup>294</sup> Separo as questões por razões analíticas. A questão não é, evidentemente, assim tão simples.

<sup>295</sup> Cf. por exemplo Taylor 1976.

<sup>296</sup> Referi atrás o utilitarismo e o kantismo, por exemplo.

criação do indivíduo tem a ver com contingências, resulta em algo que poderia não ter sido e exclui a cada passo o que não será e poderia ter sido. Pensar na auto-criação como intimamente ligada com o reconhecimento de contingências é considerar que ela não conduz ao conhecimento de uma 'lista universal do que é importante e possível para os humanos', nem a uma 'descrição única e verdadeira da condição humana'. Digamos que de acordo com esta visão, nenhum universal ou transcendência trará consolação, a única 'consolação' possível é precisamente a auto-criação. O auto-conhecimento não é assim uma descoberta da essência oculta do humano mas uma confrontação com a contingência própria. É de resto isso (a recriação do acontecer num 'eu quis assim') que cai sob o mote nietszcheano do *Amor fatum*.

Nada na forma nietszcheana de conceber o indivíduo enquanto narrativa pessoal de contingências nos impede de considerar que os humanos são produtos da natureza, sujeitos à causação natural. Tudo o que é afirmado neste âmbito aplica-se unicamente à dimensão dos humanos enquanto animais narrativos. O âmbito destas descrições são exclusivamente os assuntos humanos e não por exemplo o que se passa nos cérebros desses animais do ponto de vista neurofisiológico.

Assim, se de acordo com a perspectiva nietszcheana, a minha identidade pessoal continua a ser o meu sentido idiossincrático daquilo que é importante e possível, este deve ser concebido como opondo-se a várias ideias. De facto a perspectiva a que estou a chamar nietszcheana opõe-se a: (1) um 'desvelamento da essência universal de pessoa' em cada indivíduo, uma certa universalidade de características daquilo que é humano; (2) uma ideia de expressividade, segundo a qual na criação de si se trataria de 'raízes', de descobrir e não de moldar e de fazer; (3) um fim último para a espécie ou para a história, que se ligaria ao auto-reconhecimento da essência do humano

Note-se que se existisse um fim último como o referido em (3), as contingências de uma vida particular não seriam importantes por si e em última análise a extinção individual não importaria. Se no entanto não existe um tal fim último a individualidade do indivíduo volta a importar.

De uma outra perspectiva, a que vou chamar estabilizadora, aquela que não encontra satisfação no indivíduo e à qual apenas o universal satisfaz, algo na identidade pessoal passa por atingir a universalidade transcendendo a contingência, a particularidade de se ser apenas um indivíduo, apenas este, apenas agora (chegar a pensar, como diria F. Pessoa, «Aqui ao leme sou mais do que eu»...). De acordo com essa perspectiva estabilizadora, é possível exibir em pensamento universalidade e necessidade no individual e no contingente. Não é apenas um Hegel, por exemplo, que se deixa

tentar por esta perspectiva estabilizadora perante a história. Também o importantíssimo filão kantiano da filosofia moral contemporânea a tem no seu centro. A verdade é que, limitando a discussão ao pensamento contemporâneo, pelo menos desde Kant e a partir do seu pensamento moral e estético, um intuito moralista e um intuito romântico se guerreiam a propósito da concepção não apenas da moralidade mas do estatuto do indivíduo e da história pessoal deste. Um breve desvio pela temática artística do novo e do génio facilmente mostraria uma certa exclusão que o pensamento moral de origem kantiana opera na problemática da identidade pessoal, uma exclusão no entanto susceptível de desaparecer a partir por exemplo de interpretações românticas do pensamento estético do próprio Kant.

Destacar-se de si, do passado e fazer-se, criar-se é, como é fácil verificar, uma temática não apenas psicológica mas artística, ligada às questões do novo, da originalidade e da vanguarda e à concepção romântica segundo a qual aquilo que é mais humano no humano seria de alguma forma ‘criador’. A importância da questão nas artes é óbvia e poderia assumir a seguinte forma: tem que haver algo de novo para fazer, para ser, são o que estou *eu* aqui a fazer? Mas esse algo de novo, se não há um fim último claramente definido (social, histórico), o que será? A decisão quanto à existência ou não existência de fim último, essência, verdadeira natureza, direcção da história, determina se se considerará importante ou não ‘gastar palavras ou outros materiais com idiosincrasias e contingências’. Repare-se de resto que muita da arte do século XX, nomeadamente correntes deflacionárias, minimais, cultivadoras da irrisão, parte da decisão de que é (afinal) importante gastar materiais com idiosincrasias e contingências, abdicando do lirismo associado a um *pathos* da universalidade, da essência e da expressão de uma ‘verdadeira natureza’. O que interessa aqui notar é que na criação pessoal a questão é semelhante, que o que está em jogo a partir do momento em que se abdica do fim último é admitir que não há indivíduo humano paradigmático, que a identidade pessoal narrativa é um tecido de contingências, que qualquer coisa serve para cristalizar o sentido de identidade pessoal de uma pessoa, que não há o desumano ou o *contra natura* nas variadíssimas histórias de vida dos humanos (ou pelo menos que nenhuma forma é descontínua relativamente a formas mais ortodoxas e centradas de identidade pessoal – uma recuperação do ‘nada de humano me é estranho’).

É claro que a dimensão de auto-criação e de auto-avaliação ligada à identidade pessoal e o seu peso moral são simplesmente ocultados quando se afirma de um indivíduo ‘de um ponto de vista objectivo tu és isto: cérebro neste estado, coração neste estado, socialmente moldado desta maneira, ou quando se trabalha em teoria cognitiva, tratando os problemas lista-

dos no início deste artigo. A dimensão de auto-criação relaciona-se antes com uma questão sempre em funcionamento em cada indivíduo e que tem a forma ‘Será que eu quero mesmo ser aquilo que sou?’<sup>297</sup>. Ora, não apenas a questão está sempre em funcionamento em cada pessoa como duas respostas, sim e não, estarão provavelmente sempre a ser dadas. Ligar-se a si próprio, ‘querer-se’, comprometer-se consigo próprio, por um lado, e destacar-se de si próprio, pensar que não se é apenas o já-sido estabilizado, nem nunca apenas exactamente ‘isto’, por outro lado, são componentes inseparáveis do sentido de identidade pessoal.

Em *Contingency, Irony and Solidarity* R. Rorty<sup>298</sup> ilustra o embate das duas concepções do estatuto da narrativa de si (a que se chamou aqui nietzscheana e estabilizadora) numa mesma pessoa com um poema do poeta inglês Phillip Larkin. Mais uma vez, embora num sentido diferente do de Parfit, está em jogo a importância ou não-importância da identidade. Uma pessoa é uma, é a sua história, a sua lista de coisas idiossincraticamente importantes. Mas em que sentido é importante ser um indivíduo, este indivíduo? Quanto é que ser isso, este e apenas este indivíduo, satisfaz o próprio indivíduo? Nomeadamente, o fim da pessoa é o fim da lista contingente que compõe o seu sentido idiossincrático daquilo que é importante e possível. E que isso desapareça importa? Bom, apenas porque (ou apenas se) se é isso ou se é ser isso que é importante. Eis uma nova versão da *unimportance of identity* de que falava Parfit, agora na versão narrativa, e não já metafísica, com uma centelha de lamentação por aquilo que seria ser mais e maior do que um mero indivíduo:

*And once you have walked the length of your mind, what  
You command is as clear as a lading list  
Anything else must not, for you, be thought  
to exist,  
And what's the profit? Only that, in time  
We half-identify the blind impress  
All our behavings bear, may trace it home.  
But to confess,  
On that green evening when our death begins,  
Just what it was, is hardly satisfying,  
Since it applied only to one man once, a  
And that man dying.*

*Philip Larkin*

<sup>297</sup> You considerar que este é o ‘problema de Frankfurt’ [Frankfurt 1971].

<sup>298</sup> Rorty 1989.

Quando se trata de conceber o estatuto do indivíduo e da sua história pessoal talvez não haja escolha possível entre o intuito moralista e a ambição da universalidade, por um lado, e o intuito romântico e a importância de ser este indivíduo para o próprio indivíduo<sup>299</sup>. O objectivo do presente artigo era bem menos ambicioso do que aquilo que a exploração de uma tal questão requereria. Procurei apenas mostrar aqui algumas diferenças entre a questões (1) cognitivas, (2) metafísicas e (3) narrativas que usualmente aparecem na literatura relativa à Identidade Pessoal, enunciando as respectivas formas. Penso que qualquer investigação interessante acerca de Identidade Pessoal partirá de cruzamentos entre as três dimensões nomeadas ou mesmo da obliteração de tais distinções, uma vez que qualquer uma delas é só por si sintoma de compromissos teóricos de ordem vária.

### Referências

- Baars, Bernard, 1988, *A Cognitive Theory of Consciousness*, Cambridge, Cambridge University Press
- Bloom, Harold, 1963, *The Anxiety of Influence*, Oxford, Oxford University Press.
- Damásio, António, 1994, *O Erro de Descartes*, Lisboa, Europa América.
- Damásio, António, 1999, *The Feeling of What Happens*, New York, Harcourt Brace.
- Davidson, Donald, 1982, Rational Animals, *Dialectica*, 36.
- Dennett, Daniel, 1981, Conditions of Personhood, in Dennett, D., 1981, *Brainstorms*, Cambridge MA, MIT Press.
- Dennett, Daniel, 1983, *Elbow Room*, Cambridge MA, MIT Press.
- Dennett, Daniel, 1988, The Moral First Aid Manual, in S.McMurrin ed, *Tanner Lectures on Human Values*, III, Salt Lake City, University of Utah Press.
- Dennett, Daniel, 1989, The Origin of Selves, *Cogito*, 3.
- Dennett, Daniel, 1991, *Consciousness Explained*, Boston, Little Brown.
- Dennett, Daniel, 1992, The Self as a Center of Narrative Gravity, in F.Kessel et al., (eds), *Self and Consciousness*, Hillsdale, NJ, Erlbaum.
- Dennett, Daniel & Humphrey, Nicholas, 1989, Speaking for Ourselves, *Raritan*, 9 (1).
- Flanagan, Owen, 1992, *Consciousness Reconsidered*, Cambridge MA, MIT Press.

<sup>299</sup> Está em jogo a possibilidade de encontrar ou não encontrar qualquer 'conteúdo emancipatório' na natureza dos indivíduos, dada a ligação deste 'conteúdo emancipatório' à presunção de universalidade e a uma certa teleologia (cf. Miguens 1999).

Frankfurt, Harry, 1971, Freedom of Will and the Concept of Person, *Journal of Philosophy*, 68.

Gil, Fernando, 2000, *La conviction*, Paris, Flammarion.

Hofstadter, Douglas, 1979, *Gödel, Escher, Bach*, London, Penguin.

Hofstadter, Douglas & Dennett, Daniel, 1981, *The Mind's I*, New York, Bantam Books.

Hume, David, 1985 [1739], *A Treatise of Human Nature*, London, Penguin.

James, William, 1950 [1890], *The Principles of Psychology*, New York, Dover.

James, William, 1961 [1892], *Psychology: the briefer course*, New York, Harper and Row.

Locke, John, 1999 [1690], *Ensaio sobre o Entendimento Humano*, Lisboa, Fundação Calouste Gulbenkian.

Miguens, Sofia, 2002, Pessoas e Acções – A filosofia da mente e os fundamentos cognitivos da filosofia moral, *Análise* 23, pp. 81-133.

Miguens, Sofia, 1999, Um certo desdém de Habermas face a Derrida, *Trabalhos de Antropologia e Etnologia*, 39 (1-2). Também neste volume.

Minsky, Marvin, 1985, *Society of Mind*, New York, Simon and Schuster.

Nagel, Thomas, 1979a, Brain Bisection and the Unity of Consciousness, in *Mortal Questions*, Cambridge, Cambridge University Press.

Nagel, Thomas, 1979b, Moral Luck, in *Mortal Questions*, Cambridge, Cambridge University Press.

Nagel, Thomas, 1979c, What is it like to be a bat?, in *Mortal Questions*, Cambridge, Cambridge University Press.

Parfit, Derek, 1984, *Reasons and Persons*, Oxford, Oxford University Press.

Rorty, Amélie, 1992, Descartes on thinking with the body, in John Cottingham, (ed), *The Cambridge Companion to Descartes*, Cambridge, Cambridge University Press.

Rorty, Richard, 1989, *Contingency, Irony, Solidarity*, Cambridge, Cambridge University Press.

Shoemaker, Sidney, 1995, «Persons and Personal Identity», in Kim, J. & Sosa, E., *A Companion to Metaphysics*, Oxford, Blackwell.

Taylor, Charles, 1976, Responsibility for self, in Amélie Rorty (ed), *The Identities of Persons*, Los Angeles, University of California Press.

Williams, Bernard, 1981, *Moral Luck*, Cambridge, Cambridge University Press.



## IDENTIDADE PESSOAL E POSIÇÃO ORIGINAL RAWLSIANA

### Introdução

Não há nada de original em observar que a Posição Original (PO) rawlsiana põe em jogo uma determinada concepção da Identidade Pessoal (IP) e da forma como esta é constituída. Foi por aí que muitos críticos atacaram Rawls, declarando que a concepção rawlsiana da identidade das partes na posição original era pobre, parcial, irrealista ou mesmo discriminatória<sup>300</sup>. Já se sabe que os críticos objectam que pessoas assim (delineadas por meio da teoria da decisão racional, que é afinal uma teoria formal, ao que se junta a noção vaga de projecto racional de vida, escolhido mediante deliberação racional) não existem (frequentemente os críticos afirmam também que ainda bem que não existem pessoas assim, pois a sua vida não seria valiosa ou agradável ou proveitosa de ser vivida).

Aquilo que se pretende neste artigo é verificar qual é exactamente a concepção de Identidade Pessoal na Posição Original pressuposta por Rawls em *Uma Teoria da Justiça* e também em *Liberalismo Político* e até que ponto a neutralidade e a abstracção desta concepção são defensáveis, para posteriormente, à luz de abordagens da Identidade Pessoal e também da racionalidade na filosofia da mente e da acção avaliar a justiça de algumas das críticas endereçadas a Rawls.

#### 1. Rawls e o problema.

O problema aqui analisado não passa de todo despercebido ao próprio Rawls. Por exemplo em *Liberalismo Político* ele afirma que «a descrição das partes [na PO] pode parecer pressupor uma particular concepção

<sup>300</sup> A concepção rawlsiana foi, como se sabe, acusada de estar ligada a um particular modelo de sociedade e ao tipo de indivíduo que esta favorece (respectivamente à sociedade liberal e ao indivíduo individualista), de conceber subrepticamente como masculinas as partes condenando-se assim a ignorar por exemplo a questão da justiça no seio da família, etc. A crítica nuclear à teoria rawlsiana da justiça sublinha que esta, ao contrário do que Rawls afirma, pressupõe uma concepção determinada de bem, a concepção liberal, e também uma teoria moral determinada das relações interpessoais.



metafísica de pessoa»<sup>301</sup>, nomeadamente uma concepção essencialista de acordo com a qual as pessoas seriam independentes de atributos contingentes, incluindo os seus fins (*final ends*) e ligações (*attachments*), concepções de bem e carácter. Rawls nega que isso aconteça, considera que o véu da ignorância na Posição Original (PO) não tem implicações metafísicas específicas acerca da natureza do eu nem implica um eu «ontologicamente anterior aos factos contingentes acerca de pessoas de cujo conhecimento as partes estão excluídas»<sup>302</sup>. É certo que a teoria considera as pessoas como unidades básicas de deliberação e de responsabilidade: é relativamente a tais unidades que a PO é caracterizada. No entanto tudo o que é afirmado é que, quando se trata de definir os termos da sua associação, as partes na PO são iguais, são sujeitos livres e racionais, interessados na prossecução dos objectivos próprios e que não conhecem a sua concepção do bem (i.e. sabem que têm um plano racional de vida mas não conhecem os detalhes deste<sup>303</sup>). As partes conhecem ainda o contexto da justiça (o conflito de interesses, a escassez, etc), são caracterizadas pelo desinteresse mútuo, e estão bem providas de conhecimento: conhecem os factos gerais da sociedade humana<sup>304</sup> e toda a ‘informação geral’. É nestas circunstâncias que as partes escolhem racionalmente os princípios da atribuição de direitos e deveres básicos e da distribuição de benefícios numa sociedade.

Rawls faz uma salvaguarda relativamente à racionalidade tal como esta é entendida na teoria económica, que o guia na caracterização da decisão das partes quanto aos termos da associação: as partes aceitam uma análise do bem de acordo com a qual preferem o maior número possível de *bens primários*. Sem estes a vida seria terrível (*miserable*). A ideia de bens primários procura capturar um mínimo denominador comum de concepções substantivas de bem que podem ser totalmente diferentes e irreduzíveis: o ‘florescimento’ das diversas concepções de bem requereria algo de comum relativo a direitos, liberdades, oportunidades e, em geral,

<sup>301</sup> Rawls 1993b:27. Foi utilizada a edição americana referida na bibliografia. As traduções são da minha responsabilidade.

<sup>302</sup> Rawls 1993b: 27

<sup>303</sup> O bem de uma pessoa é definido em geral como aquilo que é para esta, dadas determinadas circunstâncias razoavelmente favoráveis, o projecto de vida racional. A motivação do indivíduo rawlsiano é sempre objecto de uma caracterização dupla, que faz apelo a (1) cálculo racional (eventualmente relacionado com a justiça) e a (2) uma concepção de bem, assim como a uma ordenação dessa caracterização, resumida no princípio da prioridade do justo sobre o bem. As concepções de bem, embora individuais, não são necessariamente egoístas. Rawls considera ainda em cada indivíduo dois poderes da personalidade moral, ambos representados na PO: a capacidade de ser racional (pela qual as partes na PO escolhem aquilo que é melhor, fazem o melhor que podem para garantir os interesses daqueles que representam) e a capacidade de ser razoável (i.e. o desejo de cooperar, de se justificar perante os outros, de honrar os termos equitativos da cooperação social), representada pelas restrições e condições da PO. Rawls considera que o razoável é público de uma forma que o racional não o é. Além disso, o razoável não pode segundo Rawls ser deduzido do racional (é algo de básico e de complementar ao racional). Cf. especialmente Rawls 1993 b, Lecture II, *The Powers of Citizens and Their Representation*.

<sup>304</sup> Entendendo por tal que compreendem os assuntos políticos, os princípios da teoria económica e as leis da psicologia humana.

às bases sociais do respeito próprio. Um dos bens primários explicitamente apontados é precisamente o respeito próprio<sup>305</sup>. Uma das razões (ou a razão) por que o respeito próprio é um bem primário é o facto de sem ele nada parecer valer a pena: sem respeito próprio não haveria motivação dos agentes. Ora, Rawls está preocupado em manter acesa a ‘motivação’ das partes na PO e dos cidadãos na sociedade. O que é aliás, a descrição da PO se não, basicamente, uma descrição da racionalidade como motivação das partes<sup>306</sup>?

Se a ideia de decisão racional, i.e. a escolha das vias mais efectivas para atingir determinados objectivos, é essencial para conceber a ‘identidade’ das partes na PO enquanto agentes isso não significa que Rawls negue que uma teoria do bem seja também necessária. No entanto, embora Rawls não negue necessitar de uma concepção do bem, ele nega que uma tal concepção de bem seja *determinada*.

De qualquer modo, e é daí que pretendo partir, a ideia de Rawls é, inegavelmente, que (1) existe uma psicologia humana que permanece quando tudo aquilo que é específico em ser-uma-pessoa-determinada<sup>307</sup> é escondido pelo véu da ignorância e que (2) essa psicologia conduz a um determinado contrato social e não a qualquer outro tipo de interacção entre as partes. Repita-se que Rawls considera que a descrição mínima da psicologia das partes de que a teoria da justiça necessita não envolve mais compromissos metafísicos do que aqueles que seriam comuns a concepções metafísicas realistas, idealistas ou materialistas, não fazendo por isso diferença alguma na questão em causa<sup>308</sup>.

<sup>305</sup> De acordo com Rawls, como pessoas livres e iguais todos os cidadãos necessitam dos bens primários. Os bens primários listados em *Uma Teoria da Justiça* são: (1) as liberdades básicas (liberdade de pensamento, de consciência, etc), (2) a liberdade de movimentos e de escolha de ocupação num fundo de oportunidades diversas, (3) poderes e prerrogativas de cargos e posições de responsabilidade, (4) rendimento e riqueza (*income and wealth*), (5) as bases sociais do respeito próprio. O respeito próprio inclui, segundo Rawls, o sentido que cada pessoa tem do seu próprio valor e a convicção de que a sua concepção de bem merece ser posta em prática, juntamente com a confiança para levar a cabo as intenções. Quanto à posição dos bens primários na teoria da justiça, cf. Rawls 1993 a, §15, *Os bens primários como base das expectativas*.

<sup>306</sup> Rawls formula frequentemente a PO afirmando que ela é um dispositivo de representação que coloca indivíduos racionalmente autónomos como representantes dos cidadãos na sociedade. Neste tipo de formulações, que fazem apelo à autonomia racional, é sublinhado o kantismo da concepção. Entende-se aqui em geral por kantismo a ideia (cognitivista) quanto à motivação de um agente de acordo com a qual é uma auto-legislação (que seria necessariamente avançada por qualquer agente racional) que move os agentes à acção. Rawls é assumidamente um kantiano mas um kantiano com salvaguardas formalistas, que faz questão de distinguir o construtivismo de Kant (a que chama construtivismo moral, substantivo) do seu próprio construtivismo kantiano (a que chama ‘político’). Apesar das salvaguardas, o kantismo de Rawls permanece notório na concepção (cognitivista) da motivação moral, na ideia de uma certa universalidade racional e na aceitação da unificação ‘natural’ do auto-controlo racional, distinta dos desejos (cf. Rawls 1993 b, Lecture III, *Political Constructivism*). Estas ideias kantianas acerca da natureza do agente racional e da deliberação marcam toda a teoria rawlsiana.

<sup>307</sup> Desde características psicológicas como o optimismo ou o pessimismo, até à riqueza, à posição social, à força física, ao grau de desenvolvimento da sociedade em que se vive.

<sup>308</sup> Aliás, Rawls afirma mesmo que a motivação das partes na PO não determina directamente a motivação das pessoas numa sociedade justa resultante do contrato, em que estas conhecem a sua posição concreta e crê que a combinação, na PO, da indiferença mútua com o véu da ignorância tem o mesmo resultado que a (aparentemente excluída) benevolência na interacção.

Até certo ponto parece razoável conceder a Rawls o que ele pretende: a teoria da justiça apenas necessitaria de uma descrição mínima de pessoa sem grandes compromissos metafísicos ou grandes reflexões quanto à natureza da IP. Utilizando um exemplo de *Liberalismo Político*, isto significará que de acordo com uma concepção política de pessoa se temos uma pessoa, Saul de Tarso, que a caminho de Damasco se converte tornando-se no Apóstolo Paulo, não devemos ter uma teoria que nos obrigue a pensar que a conversão faz dele uma outra pessoa no sentido político e institucional que interessa<sup>309</sup>. A teoria da justiça não tem necessariamente que pronunciar-se acerca de questões de IP, tais como questões metafísicas relativas ao que faz com que uma pessoa seja essa pessoa e a mesma pessoa ao longo do tempo<sup>310</sup>. A teoria da justiça necessitaria apenas de uma noção mínima de pessoa. No entanto, é o próprio Rawls que admite que «a natureza humana e a sua psicologia natural são permissivas»<sup>311</sup>, que há lugar para a variação de descrições e ideais, e que na PO está presente uma determinada ‘concepção liberal de pessoa’, que é desde o início parte integrante da teoria da justiça. A questão é portanto saber se os problemas da IP não interferem subrepticamente com aquilo que Rawls afirma acerca das partes na Posição Original.

## 2. Três tipos de problemas da Identidade Pessoal.

Propus noutra lugar<sup>312</sup> uma classificação (entendida como provisória e metodológica) dos vários problemas tratados sob o título de Identidade Pessoal na filosofia e ciência cognitiva contemporâneas. Essa classificação previa (1) questões cognitivas, (2) questões metafísicas e (3) questões narrativas<sup>313</sup>. À partida as intenções de Rawls não se cruzariam com o primeiro tipo de questões. A questões cognitivas da IP são relativas àquilo que institui centralidade e sentido de unidade da agência em sistemas cognitivos onde estes não são dados. O sentido de unidade e de centralidade é instaurado pelo funcionamento de uma representação de si – até certo ponto *virtual* – em sistemas cognitivos que são, sob uma outra perspectiva, desunificados e descentralizados. Este tipo de questões visa por exemplo o estatuto de uma tal representação por entre as representações (para além

<sup>309</sup> Cf. Rawls 1993 b: 31-32.

<sup>310</sup> Tais questões dizem respeito não apenas à persistência de fins de um indivíduo, como no exemplo, mas também à superveniência – entendida como dependência assimétrica, não causal – dos factos acerca de Identidade Pessoal a factos físicos ocorrentes nos cérebros e corpos de determinados sistemas físicos.

<sup>311</sup> Rawls 1993 b: 87.

<sup>312</sup> Cf. Miguens, *Problemas da Identidade Pessoal*, artigo incluído neste volume.

<sup>313</sup> Era aí salvaguardado o interesse do cruzamento dos tipos de questões, mas também defendida a necessidade de um esclarecimento prévio das diferenças entre eles.

de, evidentemente, a própria noção de representação e o seu aspecto de interface), bem como o teor e o conteúdo específico do auto-apercebimento (*self-awareness*), e aquilo que é mais ou menos corporal e/ou intelectual no auto-apercebimento<sup>314</sup>. As intenções de Rawls não se cruzariam à partida com este tipo de questões. Ou melhor, não se cruzam a não ser devido à racionalidade e aos mecanismos desta. De facto, os mecanismos da racionalidade não podem ser descritos sem as noções de *agente* e de *deliberação*, que pressupõem a conjugação da referida unificação (representada de si para si) do sistema com a gestão meios/ fins. A questão será retomada mais à frente.

Quanto às questões metafísicas, elas são relativas à natureza das pessoas, i.e. ao tipo de entidades no mundo que são as pessoas, isso que é reconhecido quando alguém se reconhece a si próprio como um e como um mesmo ao longo do tempo<sup>315</sup>. É duvidoso que estas questões possam ser deixadas de lado, nem que seja porque uma determinada concepção metafísica pode conduzir alguém à ideia segundo a qual a IP não importa e aquilo que importa não está focalizado em pessoas individuais (um exemplo relevante é o filósofo inglês Derek Parfit, conduzido pela metafísica da IP à defesa de um utilitarismo evidentemente estranho a Rawls). Procura-rei exemplificar brevemente este ponto.

Considere-se a teoria lockeana da IP<sup>316</sup>. Essencialmente Locke pensava que a IP não consistia nem na identidade de uma entidade imaterial (uma alma) nem na identidade de um corpo animal (humano), mas sim na *mesmidade de uma consciência ao longo do tempo*, i.e. na *continuidade psicológica*, na ligação e sucessividade do acontecer mental através da memória. Segundo Locke, o termo ‘pessoa’ nomeia um ser pensante e inteligente que possui razão e reflexão e que é capaz de se considerar a si próprio como si próprio e como uma mesma entidade pensante em diferentes tempos e

<sup>314</sup> Cf. para uma tal abordagem, por exemplo Hofstadter & Dennett 1981, Minsky 1985, Dennett 1991, Bermudez, Marcel & Eilan 1995, Damásio 1999.

<sup>315</sup> Isto significa que a questão metafísica da IP não é uma questão sub-pessoal acerca de modelos cognitivos nem directamente uma questão sobre como é, fenomenologicamente, sentir-se ser. O que está em causa é saber que tipo de entidades no mundo são as pessoas. Frequentemente a questão assume a forma de uma busca dos traços distintivos desse tipo de entidades. Têm sido propostos traços tais como a auto-consciência, a vontade livre, a existência de estados mentais acerca de outros estados mentais, a existência de volições de segunda ordem, i.e. a possibilidade de se determinar a querer ou não querer aquilo que se deseja [Cf. Frankfurt 1971 e as caracterizações – intimamente relacionadas – de ‘liberdade’ e do ‘conceito de pessoa’]. Subjacente à discussão está evidentemente o confronto entre concepções absolutistas e não absolutistas de pessoa. De acordo com uma concepção absolutista, uma entidade que é uma pessoa é sempre e em todas as circunstâncias uma pessoa, essa pessoa, e existe uma distinção clara entre pessoa e não-pessoa. De acordo com uma concepção não-absolutista, uma entidade que é por exemplo agora uma pessoa não é sempre e em todas as circunstâncias uma pessoa, essa pessoa, não existindo uma distinção clara entre pessoa e não-pessoa.

<sup>316</sup> Cf. Locke 1999, nomeadamente Livro II, xxvii, 9.

lugares<sup>317</sup>. Isso permitiu-lhe fazer uma distinção, muito influente de então para cá, entre ‘humano’, uma classificação de espécie biológica aplicada a determinados indivíduos, que não são chimpanzês ou babuínos, e ‘pessoa’, um termo que nomeia algo de mental. Este é o princípio de uma concepção não absolutista de pessoa.

Considere-se agora a concepção (reducionista) da IP de Derek Parfit<sup>318</sup>. Segundo Parfit, embora as Pessoas não sejam Egos Cartesianos, elas não são (no sentido *identificativo* de ser) corpos nem tão pouco séries de experiências (como um seguidor de Hume ou de Buda poderia defender), antes *têm* corpos, pensamentos e experiências. Simplesmente, a sua existência não está acima e para além (*above and beyond*) dos corpos, pensamentos e experiências, antes *consiste* na ocorrência de eventos mentais e físicos inter-relacionados.

Afirmar que factos *constituem* outros factos significa que (determinados) factos fazem com que determinados outros factos obtenham, de uma maneira não causal. Com ‘consiste em’ Parfit quer dizer: (1) ser necessário e suficiente para, (2) não haver mais factos (o que não significa que os factos descritos sejam os mesmos factos) e (3) uma dependência assimétrica (não relacionada com causalidade)<sup>319</sup>.

A discussão de Parfit não é acerca de conceitos mas acerca de tipos de realidades e nada nas sua posições nos obriga a crer que as pessoas são meras entidades conceptuais, que vêm à existência através do uso do conceito de pessoa. De acordo com o reducionismo constitutivo de Parfit as pessoas definitivamente existem; o que interessa saber é *como* é que elas existem. O *reducionismo constitutivo* de Parfit, opõe-se ao reducionismo identificativo e ao reducionismo eliminativo. De acordo com o reducionismo constitutivo *Embora Xs (Pessoas) sejam diferentes de Ys (corpos, eventos mentais e físicos), a existência de Xs simplesmente consiste na existência*

<sup>317</sup> A isto Locke acrescentava que ‘pessoa’, ao contrário de (ou por contraste) com ‘humano’ é um termo forense, que avalia acções e o mérito destas e que pertence ou se aplica apenas a agentes capazes de lei, felicidade e infelicidade. A preocupação especial com aquilo que são as pessoas é em grande parte devida ao envolvimento da noção em práticas jurídicas e morais (por exemplo, apenas faz sentido considerar uma pessoa responsável por um crime se ela recordar tê-lo cometido). Aliás, no caso de Locke existia também um agenda teológica, relativa a responsabilidade e ressurreição. Locke não rejeitava a existência de uma substância imaterial ‘em cada pessoa’, discernida por ser ‘o-que-pensa’ quando existe pensamento, mas apenas que ‘ser uma mesma pessoa ao longo do tempo’ requeresse a identidade de uma substância imaterial, i.e. que fossem necessárias almas para explicar a IP. Se a consciência fosse identificada com a alma e considerada a nossa essência, essa essência poderia estar ausente (por exemplo durante o sono). Ora uma tal concepção seria indefensável, pois a ausência da essência de uma entidade tornaria impossível que essa entidade continuasse a ser a mesma entidade.

<sup>318</sup> Cf. Parfit 1984. Considerarei também o curso sobre Metafísica da Identidade Pessoal orientado por Derek Parfit na New York University no Semestre de Outono de 2000.

<sup>319</sup> A primeira coisa que preocupa Parfit na discussão da metafísica da IP é a especificação do ‘consiste em’, nomeadamente de modo a mostrar que a sua posição acerca de pessoas não é eliminativista (como já foi acusada de ser, nomeadamente devido à possibilidade, prevista em *Reasons and Persons*, de uma descrição impessoal dos factos, i.e. uma descrição do mundo tal como ele é em que não aparecem pessoas e em que as experiências são descritas como ocorrendo e não como sendo ‘possuídas’ por alguma entidade).

de Ys. Pelo contrário, de acordo com o reducionismo identificativo, o ‘é’ de identidade ou instanciação entre X e Y assinala que X e Y são uma e a mesma coisa e que tudo o que é verdadeiro acerca de X é verdadeiro acerca de Y, não havendo necessidade de distingui-los (por exemplo: ‘O cometa Haley é um pedaço de gelo’). Quanto ao reducionismo eliminativo, quando este é justificado uma determinada entidade nomeada pura e simplesmente não existe. Um exemplo frequentemente utilizado por Parfit é o do Americano Médio (*average american*), por exemplo em ‘O americano médio tem 2 filhos’.

Através de um exemplo como ‘A Vénus de Cellini é um pedaço de ouro’ as insuficiências quer do reducionismo identificativo quer do reducionismo eliminativo para conceber o estatuto das pessoas (e também da Vénus) tornam-se visíveis. A estátua, embora ‘feita do ouro’, é distinta do ouro em que consiste e uma entidade de alguma forma ‘separada’<sup>320</sup>. O caso não é semelhante nem à identificação do cometa Haley com um pedaço de gelo nem à ‘eliminação’ do Americano Médio.

Não é que Parfit pretenda que as experiências requerem por definição sujeitos, ficando a existência de pessoas imediatamente demonstrada. A ideia é que a IP ao longo do tempo consiste em algum tipo de continuidade física e/ou psicológica e que embora a IP seja distinta dos factos acerca destas continuidades ela não é um facto independente. Noutras palavras, a partir do momento em que se decreveu todos os factos relativos a corpos e cérebros está descrito tudo o que há a descrever: descrever a existência de pessoas não é descrever factos a mais no mundo.

Várias consequências interessantes decorrem daqui:

- (1) Se este tipo de reducionismo é verdadeiro a IP não é absolutamente determinada<sup>321</sup>.
- (2) A co-pessoalidade das experiências não é auto-explicativa<sup>322</sup>.
- (3) Porque consistem noutros factos, a existência de pessoas e a IP ao longo do tempo não podem por si ter efeitos.
- (4) Uma inevitável ilusoriedade da nossa auto-concepção: o normal é termos crenças inconsistentes acerca da nossa própria natureza<sup>323</sup>.

<sup>320</sup> A estátua e o ouro são distintas na medida em cada um pode por exemplo ‘durar’ mais do que o outro (por exemplo derretendo o pedaço de ouro destrói-se a estátua mas não o ouro, e derretendo o ouro do interior da estátua, deixando apenas a superfície da estátua e substituindo o ouro interior por outro material, destrói-se o ouro mas não a estátua).

<sup>321</sup> Embora a tendência para pensar que a IP é determinada seja praticamente irresistível e que por exemplo num instante de tempo futuro ou seremos ou não seremos esta pessoa, isso não é verdadeiro.

<sup>322</sup> Pelo contrário, há aí uma unidade que tem que ser explicada evocando a forma como as experiências se relacionam entre si em função de factos físicos acerca do cérebro e do corpo em causa.

<sup>323</sup> De facto, embora muitas pessoas não estejam dispostas a afirmar que são Egos Cartesianos, elas não têm (nós não temos) outra maneira de pensar em si próprias a não ser como uma unidade e co-pessoalidade de experiências que se mantém ao longo do tempo.

Ora, Parfit pretende analisar as consequências das teses reducionistas acerca de IP no âmbito da teoria da acção. De facto, a questão metafísica da natureza das pessoas tem uma relação directa com a questão das razões para agir. Aquilo que os agentes (pensam que) são importa quando se trata de saber o que é agir racionalmente, em função de um cuidado ou preocupação (*concern*) tanto quanto possível justificado, ou pelo menos não auto-contraditório. Por exemplo, se a IP é bem menos básica e determinada do que aquilo que usualmente pensamos, isso é razão para pôr de alguma forma em causa a naturalidade do egoísmo racional, em geral considerado como a pedra de toque da racionalidade na acção. Um reducionista coloca as seguintes questões: porque é que, quando se trata de preocupação com o futuro e de razões para agir, havemos de estar especialmente preocupados com o *nosso próprio* futuro? Haverá algo de necessário e de definicional na ideia segundo a qual cada agente tem uma preocupação especial com o seu próprio futuro? Se a IP consiste noutros factos e a sua importância é derivada fica aberta a porta para a defesa de um ponto de vista mais impessoal sobre aquilo que importa característico por exemplo do utilitarismo<sup>324</sup>. No caso de Parfit, a metafísica da IP prolonga-se precisamente numa filosofia ‘des-pessoalizada’ das razões para agir, assentando sobre a não importância (*unimportance*) da IP.

O reducionismo leva a uma inflexão importante na discussão da metafísica da IP: o desvio da *natureza* da IP para a *importância* da IP. E a proposta é que mesmo se quase todos nós quase todo o tempo queremos, mesmo que obscuramente, de alguma forma ser e sobreviver, isso não é o que mais importa racionalmente. Seria mais racional pensar exactamente *o que é que desejamos* que seja e sobreviva<sup>325</sup>.

Discussões acerca daquilo que são ‘pessoas’ e ‘razões para agir’ são relevantes para a caracterização da racionalidade das partes na PO. De facto, ao descrever a PO Rawls está a tomar uma posição acerca da fundamentalidade das pessoas: ele está a considerar que existem, de forma básica e não problemática, unidades determinadas da deliberação racional, fundamentalmente individuadas e definidas pelo interesse próprio.

Quanto ao terceiro tipo de questões contemporaneamente tratadas sob o título de Identidade Pessoal e a que me propus a chamar questões

<sup>324</sup> De facto, não é sem algum tipo de pressupostos acerca da metafísica da IP [nomeadamente pressupostos acerca da natural separação e do carácter fundamental das pessoas] que a globalização utilitarista é tão taxativamente afastada [e mesmo considerada um adversário por excelência] na teoria rawlsiana da justiça.

<sup>325</sup> Assim como temos muito mais continuidade física com o corpo a que agora chamamos nosso do que aquela que nos importa [este corpo vivo é fisicamente contínuo com o nosso futuro cadáver, com os últimos fragmentos desagregados daquilo que hoje o constitui, mas não parece que isso importe muito], talvez a continuidade psicológica [ser uma pessoa, esta pessoa, com memórias, desejos e finalidades continuados] importe menos do que usualmente pensamos. Assim, aquilo que satisfaria como preservação ou realização daquilo que importa poderia perfeitamente não coincidir com a continuação física ou psicológica de si, sendo bastante mais alargado e mais neutro.

*narrativas*, elas são relativas ao relato autobiográfico de si ao longo do tempo, à construção sempre mais ou menos ficcionalizada que é uma parte importante da psicologia das pessoas bem como às fontes e materiais deste relato. É este o tipo de questões mais frequentemente evocado pelos críticos de Rawls: aparentemente a teoria rawlsiana da justiça ofereceria meios muito magros para a constituição de uma história de si. Exactamente de que se trata neste tipo de questões? Para além da Unidade e do Centro que pensamos ser, deixando em suspenso a questão acerca daquilo que metafisicamente somos (e que podemos não vir nunca a saber que somos), nós pensamos ser (de facto pensamos ter) uma história pessoal construída em torno de tal Unidade e de tal Centro. Uma história pessoal é uma narrativa, uma agregação mais ou menos estruturada do decurso de uma vida a que chamamos nossa e na qual somos a personagem principal. A narrativa não se identifica com o cérebro do corpo a que chamamos nosso, nem com o corpo próprio, embora estes sejam causalmente responsável por mantê-la. O Eu é o ‘centro de gravidade narrativo’<sup>326</sup> de uma história criada de forma involuntária<sup>327</sup>. Na intersecção das questões cognitivas e das questões narrativas da IP tem-se a seguinte situação: uma pessoa diz ‘Eu sou a Sofia’ e essa pessoa é como uma máquina que escreve romances<sup>328</sup>. É a dimensão narrativa da IP que faz a ligação entre a questão cognitiva e a questão social, política, artística, etc, das ‘identidades’, pensadas como aquilo que define um indivíduo como sendo esse indivíduo específico, com características que o ligam a comunidades e épocas históricas determinadas. Esta auto-criação<sup>329</sup> – um material importante de psicólogos, psiquiatras e outros profissionais da IP – é o núcleo das questões narrativas<sup>330</sup>.

Situados os três tipos de questões, retomar-se-á a partir de agora pontos específicos, começando pela questão cognitiva, da racionalidade dos agentes e portanto por aquilo que, na PO, é pensado a partir da teoria da decisão racional.

<sup>326</sup> A metáfora é de D. Dennett. Cf. Dennett 1991 e Dennett 1992.

<sup>327</sup> I.e. causada por processos que não são eles próprios nem inteligentes, nem conscientes, nem voluntários (tanto quanto sabemos aquilo que causa o nosso auto-apercebimento corpóreo e narrativo é um cérebro, no entanto esse saber será sempre indirecto: nunca vimos o nosso próprio cérebro e nunca o veremos sem mediações).

<sup>328</sup> Cf. Dennett 1989, Dennett 1991, Dennett 1992.

<sup>329</sup> Não se trata obviamente de uma criação ex-nihilo, mas da criação, sempre em curso, de uma narrativa de si. Esta condição de auto-criação é de resto a razão pela qual a IP não é dada mas perseguida e conseguida ou não conseguida, num processo em que cada um se toma a si próprio como matéria susceptível de moldagem e de mais ou menos posse.

<sup>330</sup> Esta forma de colocar a questão da IP tem relações complicadas com a questão metafísica da IP, que frequentemente tem por trás uma agenda epistemológica e metafísica ‘descritivista’, em que se trata de discernir aquilo que é, a natureza de algum tipo de entidade e não por exemplo de capturar o aspecto dinâmico desse tipo de entidade.



### 3. Problemas cognitivos: o agente e a deliberação – mecanismos e limitações da racionalidade.

O *insight* rawlsiano foi conceber a teoria da justiça como uma parte da teoria da escolha racional<sup>331</sup>. Convém então antes de mais esclarecer o que se entende por racionalidade, de um ponto de vista tanto quanto possível exclusivamente cognitivo.

Várias disciplinas utilizam e desenvolvem hoje teorias da racionalidade. É uma questão delicada dizer quais são exactamente essas disciplinas. O caso não polémico, por excelência, de teoria da racionalidade é a teoria da decisão racional, que inclui a teoria dos jogos e a teoria da escolha social (*social choice theory*) que tem aplicações na economia, na gestão, na elaboração das políticas públicas (*public policy*) e mesmo na biologia evolucionista<sup>332</sup>. A teoria das probabilidades é um exemplo mais polémico<sup>333</sup> e apontar a lógica como teoria da racionalidade é já bastante polémico<sup>334</sup>.

A forma mais simples de encarar a questão disciplinar da racionalidade é estabelecer que só é estritamente legítimo falar de teorias da racionalidade se estas prevêm *agentes*, i.e. entidades cujo comportamento adaptado a um ambiente determinado resulta da interacção entre crenças e desejos. A noção neutra e abstracta de agente, utilizada na economia, na teoria da decisão racional ou na Inteligência Artificial, é o ponto de referência incontornável da teoria da racionalidade.

Basicamente, a teoria da decisão racional analisa condições definidas sobre as preferências de agentes, assumindo que utilidades e probabilida-

<sup>331</sup> Apesar de recuos e salvaguardas posteriores, relativos nomeadamente à distinção entre 'razoável' e 'racional' [cf. por exemplo Rawls 1993 b, Lecture II, §1].

<sup>332</sup> A teoria da decisão racional foi desenvolvida no século vinte por matemáticos, estatísticos, economistas e filósofos e constitui a visão normativa actualmente dominante da natureza da racionalidade [cf. por exemplo Nozick 1993: 41]. De acordo com a teoria da decisão, um agente racional é uma entidade tal que tem preferências determinadas e ordenadas e age de acordo com elas, escolhendo por entre as opções de acção que se lhe oferecem aquela que maximiza a utilidade esperada [essa utilidade é a utilidade atribuída pelo agente ao estado do mundo resultante da acção escolhida]. Numa primeira caracterização a teoria da decisão apresenta modelos das situações em que um agente racional escolhe sozinho [cf. por exemplo Von Neumann & Morgenstern 1944 ou Ramsey 1926], a teoria dos jogos apresenta modelos das situações em que um agente racional escolhe em situações em que a sua escolha depende daquilo que outros agentes escolherem, i.e. situações de co-agência [co-agency] ou confronto [cf. por exemplo Von Neumann & Morgenstern 1944] e a teoria da escolha social apresenta modelos para situações em que um agente racional escolhe por outros agentes racionais, i.e. em função das preferências e interesses desses outros agentes, como seu 'delegado' [a obra inicial de referência da teoria da escolha social é Arrow 1963]. A teoria da escolha social é muito importante na teoria da governação [government theory] e na teoria da democracia.

<sup>333</sup> A teoria das probabilidades pode ser considerada uma teoria da racionalidade tanto quanto for tomada como uma lógica da crença parcial e do argumento inconclusivo, na terminologia de F. Ramsey. Cf. Ramsey 1926. Esta não é obviamente a única utilização da teoria das probabilidades – pense-se na estatística e na física – e nem sequer a mais importante mas apenas aquela pela qual a teoria das probabilidades pode ser considerada como uma teoria da racionalidade. Para considerar a teoria das probabilidades como uma lógica da crença parcial Ramsey põe de lado, evidentemente, a concepção frequentista comum de probabilidade.

<sup>334</sup> Menos polémico é indicar a exploração de sistemas lógicos na Inteligência Artificial, considerada como ciência experimental, como uma investigação da natureza da racionalidade. A Inteligência Artificial lida com o desenho de agentes artificiais, por contraste com a lógica, que lida com estruturas formais abstractas.

des estão implícitas nas preferências demonstradas, e enuncia determinados constrangimentos sobre a estrutura do agente racional, tais como a existência de uma listagem ordenada e hierarquizada (*ranking*) de preferências, a transitividade das preferências, a coerência pelo menos mínima das crenças em qualquer corte temporal, a cada instante, e ao longo do tempo<sup>335</sup> bem como a capacidade de fazer inferências válidas. Um agente é racional se escolhe por entre os meios de que dispõe aqueles que são adequados para atingir os fins que deseja atingir mediante um cálculo de utilidades e probabilidades<sup>336</sup>.

São estes componentes da racionalidade do agente que a teoria da decisão racional fornece como ‘esboço de uma psicologia’. Esta psicologia é em última análise um egoísmo racional, historicamente ligado à forma de conceber o comportamento de agentes no mercado e com a concepção de mercado utilizada na teoria económica clássica<sup>337</sup>, e uma concepção perfeitamente apriorista, que não está baseada em quaisquer resultados experimentais, e que tem por isso mesmo muito mais pertinência como pretensão normativa do que como descrição de factos. No entanto, sem psicologia da decisão racional não haveria teoria rawlsiana da justiça<sup>338</sup>.

Algumas das grandes críticas à teoria da decisão racional como teoria da racionalidade dizem respeito (1) à ideia de maximização e à unificação que esta envolve, (2) ao intelectualismo ou cognitivismo da formulação, que oculta o funcionamento dos desejos dos agentes como factor motivacional, e também ao facto de (3) aquilo a que nas teorias da acção se chama ‘racionalidade’ ser (apenas) uma interpretação de ocorrências físicas em

<sup>335</sup> Embora o desconto temporal (i.e. o facto de os agentes racionais humanos, por exemplo, não serem temporalmente imparciais relativamente aos resultados desejados das suas acções: em princípio tendemos a descontar um benefício futuro no presente) seja um dos problemas mais interessantes tratados neste contexto.

<sup>336</sup> De modo a manejar teoricamente as situações de escolha, a teoria da decisão atribui valores numéricos às utilidades e probabilidades e propõe um cálculo sobre esses valores. A teoria da decisão lida assim com medidas e comparações das preferências dos agentes, das utilidades dos resultados [outcomes] e da probabilidade de obtenção dos resultados dadas as acções. Diferentes axiomatizações são propostas. Cf. por exemplo Von Neumann & Morgenstern 1944, *Theory of Games and Economic Behavior*. Os autores apresentam uma teoria matemática da decisão que se aplica quando um agente está perante opções exclusivas de acção. A axiomatização da utilidade apresentada por John Von Neumann e Oskar Morgenstern, à qual os estudos da decisão racional usualmente se reportam, toma as probabilidades como dadas. Cada acção tem resultados [outcomes] aos quais são atribuídas utilidades, i.e. valores. Atribui-se probabilidades condicionais a cada resultado possível *R* em relação com uma acção, por exemplo a acção *A*. A utilidade esperada de um dado resultado *R* de uma acção *A* é dada pela seguinte fórmula:  $u(R) \times P(R|A)$ . De acordo com a teoria, a racionalidade do agente envolve a prática do acto com a maior utilidade esperada. Antes de Von Neumann e Morgenstern já F. Ramsey tinha elaborado uma teoria da crença parcial e do raciocínio inconclusivo baseada numa teoria subjectivista da probabilidade, e nas noções de preferência e expectativa matemática do agente quanto a resultados das suas acções, apresentando propostas para a manipulação numérica desta lógica [cf. Ramsey 1926, *Truth and Probability*].

<sup>337</sup> Segundo a qual o comportamento do mercado resultaria da interacção entre agentes completamente racionais cada um prosseguindo os seus interesses egoístas e decidindo-se por aquelas acções às quais atribuem a maior probabilidade de originar as melhores consequências para si de acordo com os seus objectivos e com a informação de que dispõem.

<sup>338</sup> Isto não significa que Rawls não reconheça as várias críticas a que a teoria da decisão racional está sujeita. Cf. nomeadamente Rawls 1993 a, nota 14 na p. 125.

sistemas físicos. A particular crítica que se aborda em primeiro lugar tem a ver com o irrealismo da ideia de maximização para caracterizar situações reais de deliberação<sup>339</sup>.

Aparentemente, um processo de *satisficing*<sup>340</sup> é fundamental em grande parte dos processos reais de decisão, desde as ‘escolhas’ da selecção natural que conduziram ao *design* de agentes até aos processos humanos de decisão. Nestas condições, agentes resultantes de evolução por selecção natural não podem ser considerados como ‘agentes perfeitamente racionais’: o seu *design*, inclusive o *design* para a cognição, resulta ele próprio de um processo de *satisficing*.

De modo a pôr em relevo as características de um tal processo e suas implicações na concepção de agente D. Dennett propõe o seguinte exemplo de deliberação seguida de decisão<sup>341</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. Perante o número das candidaturas e a quantidade de informação de cada uma, a primeira coisa a constatar é que (1) *é impossível considerar todos os factores relevantes no tempo disponível*. O (2) *uso cego e impiedoso de regras grosseiras de selecção e de decisão* (por exemplo um critério não absolutamente confiável mas não irrelevante, tal como as médias das classificações) é imprecindível. Além disso, (3) *o decisor terá que abdicar 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 desses processos de modo a cercar as alternativas em consideração, procura-se (4) *salvar alguma coisa do processo selvagem anterior, nomeadamente procedendo a uma análise exaustiva e cuidadosa das candidaturas que ficaram*, das quais se vem a escolher uma. Este ultimo processo é uma tentativa, resultante de auto-monitorização, de corrigir o resultado final.

Qualquer decisão tomada nestas circunstância é, evidentemente, enormemente vulnerável a um olhar retrospectivo: não existe qualquer garantia de que a alternativa que acabou por ser seleccionada seja *a melhor alternativa*<sup>342</sup>. No entanto, o decisor pode ainda assim considerar que a sua decisão foi *a melhor possível* dadas as circunstâncias. Dennett sugere que

<sup>339</sup> A crítica parte de considerações relativas ao *design* de um deliberador e aplica às teorias da escolha como maximização racional em geral e não apenas à psicologia da decisão pressuposta por Rawls.

<sup>340</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 em situações nas quais é impossível medir todas as alternativas por uma função de utilidade comum.

<sup>341</sup> Cf. Dennett 1988.

<sup>342</sup> É mesmo provável que os melhores candidatos tenham sido afastados.

o processo retrata de forma bastante fiel (e de acordo com ideias acerca do processamento cognitivo como paralelo e distribuído), as características dos processos de deliberação e escolha em deliberadores reais.

Ora, estas características da deliberação e da decisão não são usualmente encaradas de forma séria pelas principais correntes da filosofia moral (ou melhor, são afastadas com a consideração de que qualquer situação real está longe de ser aquilo que deveria ser por princípio). O utilitarismo e a deontologia identificam-se nomeadamente num ponto importante: ambos exigem que sejam (ainda que idealmente) *consideradas todas as coisas* anteriormente à decisão. Ora, de acordo com Dennett, «nenhum sistema de ética remotamente persuasivo foi alguma vez tornado computacionalmente manejável, mesmo que 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>343</sup>.

Para agentes reais, objectivamente limitados, é impossível considerar todas as coisas, e Dennett pensa que é importante que a filosofia moral leve a sério tal impossibilidade. É por isso preferível conceber a deliberação moral como um exercício de disposições e não como maximização racional (entre outras coisas entende-se por maximização racional que ‘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*’).

É mais realista do ponto de vista cognitivo considerar que as pessoas são deliberadores tais que são guiados por algo como um ‘cuidado não maximizante’ e não por maximização racional, o que significa nomeadamente que elas fazem a todo o momento escolhas não maximizantes, que não podem ser identificadas com a escolha que qualquer agente racional faria. Essa estratégia não é aliás irracional, dadas a insuficiência da informação e a pluralidade dos ‘bens’ em consideração.

Como se sabe, as éticas da virtude propõem a substituição do cálculo racional por considerações acerca do carácter dos agentes morais<sup>344</sup>. Ora, Dennett defende que somos constituídos, enquanto deliberadores,

<sup>343</sup> Dennett 1988:129.

<sup>344</sup> Corrigindo assim ao mesmo tempo a desconexão entre as decisões racionais e a vida e a pessoa dos agentes que é característica das éticas da maximização racional.

por imperativos biológicos e dada a pressão temporal, de modo a agir de uma forma que se aproxima da situação caracterizada pelas éticas da virtude. O que Dennett faz é transpôr a noção (de âmbito pessoal) de carácter para o âmbito cognitivo sub-pessoal. É possível no entanto estabelecer uma ligação entre a caracterização cognitiva da deliberação e posições em teoria moral. Parte-se do seguinte: o que a caracterização estrutural dos deliberadores reais mostra é a falta de pertinência cognitiva das éticas da maximização racional. Dois tipos de considerações nascem daí. Um primeiro tipo, tratado em seguida, diz respeito a estratégias de melhoramento das decisões. Um segundo tipo diz respeito àquilo que é possível, em tais circunstâncias, afirmar acerca da natureza da racionalidade, do ponto de vista de uma teoria filosófica da racionalidade e será tratado no ponto 4.

Relativamente à questão do melhoramento das decisões morais, Dennett desenvolve a ideia de um manual de primeiros socorros morais<sup>345</sup>. Este constituiria uma necessidade para agentes que agem em tempo real, sendo ao mesmo tempo uma tentativa de incorporação na teoria filosófica da moralidade das imperfeições constitutivas dos agentes reais. O manual diria o que fazer imediatamente e teria como finalidade criar estados disposicionais em agentes que têm que decidir sem tempo para calcular todas as implicações dos seus actos. O que poderia conter um tal manual? Uma proposta decorre do exemplo dado atrás da selecção de um candidato apesar dos constrangimentos objectivos à decisão: tendo discernido os meta-problemas do processo de decisão o deliberador pode desenvolver hábitos de pensamento que lhe permitam lidar com tais problemas. Estes estariam contidos no manual. Por exemplo, perante a proliferação das considerações na situação de deliberação (o ‘excesso’ das candidaturas), o deliberador pode desenvolver interruptores da ponderação, que terminem arbitrariamente a geração das considerações. Estes ‘interruptores’ são maneiras de lidar com a intratabilidade computacional e impedem de facto a consideração de 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 (bastante problemático) do manual de primeiros socorros morais seria a ideia segundo a qual os manuais podem ser vários e alternativos (o que supõe, evidentemente, que do ponto de

<sup>345</sup> Dennett 1988.

vista cognitivo noções que fazem apelo a Uma-ideia-de-racionalidade, tais como ‘A audiência racional’ ou ‘A transparência ético-política das sociedades’ seriam duvidosas). De qualquer forma, e essa é a sua importância, a ideia do manual ergue-se contra o erro comum das éticas da maximização racional, a ideia segundo a qual «deve haver uma única (melhor ou mais alta) perspectiva a partir da qual é possível avaliar a racionalidade ideal»<sup>346</sup>.

Embora Dennett apresente as suas sugestões sob a forma de uma investigação do *design* do deliberador e rejeite as aproximações com as éticas da virtude, as estratégias de meta-controlo da deliberação apontadas no manual parecem sustentar as propostas destas. Os problemas conceptuais com que a ideia de maximização racional se depara no domínio ético e que conduzem às propostas das éticas da virtude revelam assim ser o reflexo de características estruturais dos deliberadores, que a filosofia moral deveria considerar e aquilo que começa por ser apenas uma descrição das características sub-pessoais poderia conduzir a propostas substantivas em filosofia moral. Poder-se-ia por exemplo imaginar uma situação em que, para além da sua utilidade cognitiva e estrutural, os interruptores de considerações tivessem uma motivação moral e em que o abandono do ideal da maximização racional conduziria à defesa de uma concepção não racionalista de moralidade e motivação moral<sup>347</sup> e à aceitação do facto de o cuidado não maximizante ser uma estratégia racional perante a pluralidade de objectos de cuidado.

Em conclusão, a ideia segundo a qual a deliberação é maximização auto-controlada e unívoca, totalmente dominada pelo agente, resultando numa máxima racional que rege a acção ou num cálculo de utilidades inobjectável, é irrealista. O que ‘constitui’ um deliberador real é mais propriamente monitorização de uma multiplicidade de produções: a unidade deste é mais uma ‘unificação de segunda ordem’ do que ‘uma unidade no comando’. Do ponto de vista cognitivo, nem a unidade do agente nem o auto-controlo na deliberação que conduz à acção podem ser tomados como garantidos. As condições reais do ‘agente’ e da deliberação devem

<sup>346</sup> Dennett 1988: 139.

<sup>347</sup> De facto, o que seria um agente moral totalmente racional? Um tal ser não gozaria sequer a vida própria nem gozaria da existência de outros seres, [por exemplo em termos de amor ou de amizade] considerando-os apenas por ocasião da [não] estrita submissão das acções a regras racionais ou do cálculo do bem comum. Estas críticas aplicam-se quer à deontologia quer ao utilitarismo precisamente tanto quanto estas são éticas da maximização racional. Aliás, de acordo com uma teoria humeana radical da motivação um ‘agente’ totalmente racional nem sequer agiria: o agir só pode ser movido a desejos e esse ser estaria apenas informado. Como é sabido, nenhuma escolha particular face a uma situação determinada é, de acordo com o humeano, despoletada pela mera informação. Como diz Blackburn (Blackburn 2000), a razão apenas determina o entendimento das situações, o input, e não a forma como o input é convertido em output, i.e. em acção. Para um humeano como Blackburn, toda a motivação, mesmo a mais elevada ou intelectual, depende de desejos. Note-se que o humeano, ao contrário do kantismo, não estabelece uma cisão entre os desejos e o dever-ser, nem elimina o lugar dos sentimentos e atitudes na motivação para agir.

assim tornar-nos cautelosos face ao perfil de agentes racionais rawlsiano-kantianos.

#### 4. Um indivíduo e um projecto racional de vida.

Entender-se-á em seguida ‘pessoalidade’ como pessoa no sentido pleno e concretizado e não já apenas agente-delineado-a-partir-da-decisão-racional. Essa é a questão que Rawls aborda através da ideia de *projecto racional de vida* como bem das pessoas. Na PO as partes sabem que as pessoas terão um projecto racional de vida e uma das coisas que deve ser possível na sociedade regida pelos princípios da justiça, uma vez estabelecida esta pelo contrato das partes, é que os indivíduos se tornem pessoas desenvolvidas enquanto tal (*full blown persons*), exercendo poderes morais e prosseguindo concepções de bem, ‘preenchendo’ assim o esqueleto fornecido pela teoria da decisão. Rawls evoca portanto duplamente a racionalidade para caracterizar as suas ‘pessoas’: quando as caracteriza como *agentes racionais* e quando afirma que elas têm um *projecto racional de vida*.

O projecto racional de vida é uma noção estranha<sup>348</sup>, na qual está em causa o uso da racionalidade para se conceber aquilo que se é enquanto ‘pessoa determinada’, no sentido ‘denso’ e não apenas esquemático (o sentido que caracteriza o agente racional). Aparentemente aquilo que Rawls propõe é a utilização do esquema da decisão racional como substituto daquilo que normalmente os profissionais da IP (como psicólogos e psiquiatras) estão mais prontos a conceber como uma história que cada pessoa tem acerca de si própria, uma história circundada de afectos, relativamente aos seus objectos e objectivos.

Procura-se em seguida, antes de mais, saber o que é que, de um ponto de vista filosófico, pode ser afirmado sobre a natureza de tal racionalidade que opera na constituição de si. Ora o uso da racionalidade sobre si pode ser caracterizado, a partir da filosofia da mente<sup>349</sup> precisamente como uma característica do tipo especial de sistemas cognitivos físicos que são ‘pessoas’<sup>350</sup>. Estas constituem um ponto de aplicação específico de uma teoria geral da mentalidade no mundo físico. De acordo com a Teoria

<sup>348</sup> Rawls caracteriza a noção da seguinte forma: «O projecto de vida de alguém é racional se e só se 1) for um dos projectos que são coerentes com os princípios da escolha racional quando aplicados a todas as características relevantes da situação em que encontrem e 2) for o projecto que, de entre todos aqueles que respeitam esta condição, foi escolhido pela pessoa em causa a partir de uma deliberação inteiramente racional, isto é, tendo total consciência dos factos relevantes e depois de uma cuidadosa análise das consequências» (Rawls 1993 a: 314).

<sup>349</sup> Escolhe-se aqui como referência sobretudo a filosofia de D. Dennett e a concepção de *pessoalidade* (*personhood*) no âmbito da Teoria dos Sistemas Intencionais, a teoria geral do estatuto da intencionalidade ou mentalidade no mundo físico.

<sup>350</sup> Dennett 1981, *Conditions of Personhood*. Cf. Davidson 1982, *Rational Animals*, para uma concepção menos continuísta do que a de Dennett do estatuto da racionalidade e da *pessoalidade* relativamente a outras formas do ‘mental’.

dos Sistemas Intencionais (TSI), a pessoalidade (*personhood*) é uma forma particular e sofisticada da mentalidade. De acordo com a caracterização da Teoria dos Sistemas Intencionais, pessoas são entidades tais que são não apenas susceptíveis de serem interpretadas como Sistemas Intencionais, como são capazes de assumir a estratégia intencional sobre outros sistemas e de assumir que outros sistemas assumem a estratégia intencional sobre eles (i.e. que lhes atribuem mentalidade ‘de volta’). Abre-se assim o espaço da verdade, da mentira, da confiança e do logro que caracteriza as relações inter-mentais e comunicacionais humanas, que é uma condição essencial do tipo de mentalidade que é a pessoalidade. Mas as pessoas são capazes de mais: são capazes de avaliar racionalmente a o seu próprio uso da racionalidade em todo o processo interpretativo e nomeadamente na constituição da sua própria identidade.

Mas qual é a natureza desta racionalidade que conduz interpretativamente a sucessivos ‘aprofundamentos’ da mentalidade de parcelas do mundo físico até chegar à pessoalidade? A racionalidade é o ponto de apoio de teorias quinianas do mental, como por exemplo as de D. Davidson e de D. Dennett, de acordo com as quais nada no mundo físico<sup>351</sup> que seja insuficientemente racional é interpretável como mental. A esta posição está associado um princípio de caridade, de acordo com o qual um agente tem crenças que são na sua maioria verdadeiras e faz inferências que são na sua maioria racionais. Este é um princípio ou constrangimento apriorista que se aplica a tudo o que é mental<sup>352</sup>.

Já se viu que a abordagem cognitiva dos mecanismos da racionalidade revela algo a que poderia chamar uma imperfeição constitutiva da racionalidade dos agentes reais<sup>353</sup>. Ora, é impossível não notar o choque

<sup>351</sup> Ou melhor, no mundo pensado, do ponto de vista metafísico, como sendo basicamente físico.

<sup>352</sup> O núcleo das teorias quinianas do mental é este princípio de caridade. O lugar originário do princípio da caridade é o pensamento de Quine [cf. o parágrafo 13 do Capítulo 2 de *Word and Object*]. O princípio da caridade acompanha a filosofia da mente contemporânea a partir do momento em que Quine declara [cf. a observação acerca da tese de Brentano §45 de *Word and Object*] que o mental ou intencional tem um estatuto interpretativo. Estão assim geradas as teorias do mental que são teorias da interpretação, nomeadamente a tradução radical do próprio Quine, a TSI de Dennett e a interpretação radical de Davidson. A proposta de Quine, que rege todas as teorias da interpretação, nomeadamente as de Dennett e de Davidson, é a seguinte: para além de um certo ponto, a [suposta] irracionalidade do indivíduo sob interpretação deve fazer com que o intérprete duvide das suas próprias hipóteses de tradução. É mais provável que estas hipóteses estejam mal formuladas do que que o indivíduo a ser interpretado seja desse modo irracional. O princípio afirma a existência de uma conexão incontornável entre intencionalidade e racionalidade: a atribuição de crenças a um sistema pressupõe a racionalidade do sistema, a qual é comportamentalmente [e portanto instrumentalmente] avaliada.

<sup>353</sup> Embora não haja problemas em continuar a considerar a racionalidade como ‘a melhor adaptação possível do comportamento de agentes ao seu ambiente, consideradas as suas crenças e desejos’ bem como uma adaptação evolutiva com propósitos determinados. A aceção mínima e incontornável de racionalidade de que se trata aqui é a racionalidade instrumental, uma caracterização de racionalidade que vale para qualquer sistema cognitivo, consciente ou não. Agir de acordo com ‘razões’, é evidentemente um funcionamento que pode ter bases diversas nos sistemas cognitivos, desde a pura e simples instalação [wiring] que propicia a adaptação das funções do sistema ao seu ambiente, até ao condicionamento, até ao reconhecimento consciente de conexões entre conteúdos. Este último, embora possa parecer destacar-se relativamente às ‘bases’ anteriores pode igualmente ter sido seleccionado [cf. Nozick 1993: 108].



entre as teorias quonianas do mental, que supõem de algum modo uma impossibilidade de irracionalidade nos sistemas interpretáveis como mentais e a perspectiva cognitiva que sublinha as limitações constitutivas ou mesmo a ‘generalização’ da irracionalidade<sup>354</sup>.

Tomando de novo como exemplo a TSI, é mesmo possível considerar que ela envolve um argumento a favor da *impossibilidade* de irracionalidade<sup>355</sup>. Aliás, o argumento aparece sob um forma dupla: por um lado fazendo apelo, à maneira de Quine, à conexão intencionalidade-racionalidade, por outro fazendo apelo à submissão dos sistemas cognitivos à selecção natural. Que o argumento seja duplo revela, de resto, o próprio cerne do problema das teorias filosóficas e cognitivas da racionalidade: a ambiguidade relativa à natureza da racionalidade. Nas investigações da natureza da racionalidade nunca é totalmente claro se a teoria da racionalidade é uma teoria cognitiva naturalista, relativa a comportamentos adequados de sistemas físicos resultantes de selecção natural<sup>356</sup>, ou uma teoria normativa, com apoio na qual sistemas físicos são interpretados como mentais e que permite uma ‘exigência de correcção’.

Reportar a racionalidade à evolução por selecção natural é bastante comum nas teorias psicológicas e filosóficas. O que importa desde logo notar é que o facto de se considerar o design para a racionalidade da perspectiva da selecção natural, estende a este o estatuto geral dos produtos da selecção natural (um *design* não óptimo mas apenas satisfatório, suficientemente bom). A reportação da racionalidade à selecção natural é uma reportação ao êxito no comportamento global do agente e apenas isso é e portanto também, inevitavelmente, um passo no sentido de duvidar de qualquer qualidade intrínseca dos produtos e resultados da racionalidade.

Se o ponto de vista evolutivo, apesar de oferecer um argumento a favor da impossibilidade de irracionalidade, não coloca qualquer constrangimento de ‘qualidade’ intrínseca sobre os processos e produtos da racionalidade a abordagem feita por meio do princípio da caridade pare-

<sup>354</sup> Cf. Tversky & Kahneman 1981, Tversky & Kahneman 1993. Mesmo que se trate de sujeitos academicamente treinados em lógica ou teoria das probabilidades há certas irracionalidades recorrentes [dir-se-ia na terminologia da teoria dos jogos que as estratégias dominantes são irracionais...]. Nos aludidos estudos chama-se irracionalidade por exemplo à existência de tendências ou inclinações prévias (*bias*) que determinam o enquadramento (*framing*) das escolhas [supostamente racionais] de agentes, às más estimativas de probabilidades, à intransitividade [ou mesmo inversão] das preferências dos agentes, ao facto de os agentes incorrem sistematicamente em falácias, etc. Em suma, chama-se irracionais a [1] processos pelos quais agentes chegam a conclusões que não podem ser justificadas pelo seu próprio conhecimento, e a [2] processos que conduzem a uma conclusão ou decisão que não é a melhor que poderia ter sido alcançada à luz da evidência disponível, com os recursos temporais disponíveis. São assim possíveis situações em que agentes acreditam em contradições, não acreditam nas consequências daquilo em que acreditam, incorrem em fraqueza da vontade e em auto-engano, etc. Em todos estes exemplos trata-se de irracionalidade na gestão dos meios, e portanto de racionalidade instrumental e não de racionalidade dos fins

<sup>355</sup> Tem aliás sido acusada disso, cf. Stich 1993, Capítulo 2, *Good Reasoning and Intentional Content: How Irrational Can We Be?*.

<sup>356</sup> Por contraste com, aliás, um mecanismo de representações e computações à la Fodor, que seria uma caracterização puramente interna aos sistemas cognitivos daquilo que se entende por racionalidade.

ce obrigar à atribuição de uma racionalidade mais perfeita. Os quinianos podem, é certo, ter posições diferentes quanto ao estatuto da racionalidade suposta na interpretação.<sup>357</sup> Mas todos concordam que a suposição de racionalidade é criterial para identificar e caracterizar a intencionalidade, não constituindo uma opção mas uma condição.

O ponto de apoio do intérprete (quiniano, davidsoniano ou dennettiano) é a ‘racionalidade instrumental’, comportamentalmente avaliada. Mas um tal ponto de apoio deixa em aberto saber que racionalidade é essa (e quanta) que é assim inapelavelmente suposta, e quanta racionalidade tem que ser suposta<sup>358</sup>. Esse ponto de apoio não tem também por que envolver qualquer ponto de vista exterior em função do qual se possa falar de Racionalidade perfeita. No caso particular da TSI, uma hipótese que se coloca é mesmo que a racionalidade não seja nem uma noção muito profunda nem uma noção unificada. Antes de mais, a racionalidade é uma noção do nível do agente, i.e. do nível de sistemas cognitivos físicos que se comportam adequadamente, do ponto de vista instrumental, nos seus mundos, e não uma noção apoiada na existência de géneros naturais representacionais, uma noção por princípio insusceptível de caracterização *precisa* (e de *uma só* caracterização). É neste sentido, o sentido em que a racionalidade só pode ser avaliada instrumentalmente, que a ela não tem ‘valor intrínseco’. Este ponto é importante quando se coloca a questão da comprovação empírica da irracionalidade de sujeitos em estudos empíricos<sup>359</sup>.

Davidson tem uma proposta de resolução do problema do choque entre estudos empíricos da racionalidade e a necessidade de usar o princípio de caridade na teoria da mental: tudo depende do que se entende por comprovação empírica da racionalidade<sup>360</sup>. A suposição de que estudos empí-

<sup>357</sup> Isto é claro comparando os casos de Dennett e de Davidson: Davidson sugere que a verdadeira natureza da racionalidade suposta na interpretação é revelada pela teoria da decisão, Dennett considera a racionalidade um conceito sistematicamente pré-teórico. Cf. Davidson 1980c, Hempel on Explaining Action e Dennett 1987.

<sup>358</sup> O que as teorias quinianas da interpretação normalmente fazem é apelar a intuições acerca de irracionalidade (o exemplo de irracionalidade improvável do próprio Quine é um indivíduo que crê *p* e  $\neg p$ ). Mas onde se vai buscar os *standards* de racionalidade quando se apela intuições de irracionalidade? Que *standards* de racionalidade mantêm ainda alguém que afirma, como por exemplo Dennett faz, que a racionalidade não consiste na consistência das crenças, nem em fechamento dedutivo nem em capacidade inferencial perfeita? No caso específico da TSI, passa-se por situações em que se afirma, verificado o choque entre a racionalidade e a incorrigibilidade dos SI, que a atribuição de racionalidade é sempre global ou holista e de certo modo permissiva (no sentido em que admite alguma inconsistência), por situações em que a insuficiência computacional, a incapacidade de inferências perfeitas considerando todos os factores, é, como se viu, evocada como razão suficiente para a oposição às éticas da maximização racional, por situações em que se afirma que a solução para [aquilo a que na Inteligência Artificial se chama] o problema do enquadramento passa, nos humanos, por uma não obrigação à coerência das crenças do sistema. [Nesse contexto Dennett defende que nada na implementação dos hábitos – que são ideias práticas, razão por instinto, feixes de como-lidar – na criatura obriga a preservar a consistência do que estes instauram em termos de imagem manifestal].

<sup>359</sup> Estes teriam conclusões pessimistas, provariam que agentes reais são maioritariamente irracionais em grande parte das circunstâncias de inferência e decisão. Pelo menos é isso que eles aparentemente mostram, mas é claro que uma tal avaliação ‘supõe a racionalidade’.

<sup>360</sup> Davidson 1980.

ricos poderiam fornecer a comprovação (ou infirmação) da racionalidade dos agentes não é razoável. Experiências empíricas acerca de racionalidade não podem infirmar as caracterizações que a teoria da racionalidade (que é para Davidson a teoria da decisão, tomada como definindo a verdadeira natureza da racionalidade<sup>361</sup>) faz da estrutura do agente racional. De facto, não sabemos imaginar o que é para um agente ser racional (ou irracional) sem o quadro teórico da teoria da decisão: a caracterização do agente (feita pela teoria da decisão) é *constitutiva* daquilo que é ser racional e agir racionalmente. Nenhuma interpretação ou avaliação dos dados experimentalmente obtidos pode ser feita sem utilizar as próprias propostas da caracterização daquilo que é um agente racional. Pura e simplesmente não sabemos nem podemos pensar de outra maneira, não podemos pensar que nós próprios não pensamos (ou que qualquer outra pessoa não pensa) geralmente de forma racional (a não ser deixando de pensar que pensamos, por exemplo atribuindo erros a irrupções de distúrbios ao nível do *hardware* cognitivo).

É óbvio que há uma espécie de circularidade na posição de Davidson: a racionalidade constitutiva do agente é encontrada porque é 'lá posta'<sup>362</sup>. Mas Davidson crê que a teoria tem que ser preservada porque é tudo o que temos. A circularidade na tentativa de justificar ou fundamentar a racionalidade de agentes parece de qualquer modo ser incontornável, estando provavelmente relacionada com o tipo de seres que elaboram teorias da racionalidade, seres instrumentalmente e minimamente racionais<sup>363</sup>.

Quando se substitui a teoria da decisão proposta por Davidson como única possibilidade de chegar à verdadeira natureza da racionalidade pela ideia dennettiana de uma caracterização sistematicamente pré-teórica da racionalidade obtém-se uma análoga obrigação à idealização mas agora sem qualquer obrigação de propor uma teoria da verdadeira natureza do objecto idealizado (ou de propor que o objecto idealizado tem uma verdadeira natureza).

Um problema adicional, voltando à conexão 'conceptual' entre racionalidade e intencionalidade suposta pelos quinianos, é saber quanta racionalidade será necessário supôr quando se interpreta. Um ponto de acordo não muito promissor, aparentemente, é o facto de, se nos usamos a nós próprios como modelo, ser consensual que a nossa racionalidade

<sup>361</sup> Davidson pensa que apenas os humanos são verdadeiramente racionais; a TSI não tem que fazer restrições idênticas.

<sup>362</sup> Na terminologia de A. Zilhão [Zilhão 1998/1999] e forçando um pouco os termos do próprio Davidson (Davidson 1980: 261, Hempel on Explaining Action) isto significa que os axiomas da teoria da decisão são verdades sintéticas a priori acerca de seres racionais quaisquer. Essa é a razão pela qual Davidson pensa que os elementos da teoria da decisão [por exemplo as considerações acerca da transitividade das preferências] têm que ser preservados face a resultados como por exemplo os de D. Tversky e A. Kahneman que mostram como é comum a inversão de preferências em agentes reais, bastando que existam variações do enquadramento da decisão em causa [Tversky & Kahneman 1981].

não é muita ou muito perfeita<sup>364</sup>. De novo no caso específico da TSI, e ao contrário do que alguns críticos de Dennett afirmam<sup>365</sup>, não faz sentido considerar que a TSI, enquanto teoria quiniana do mental, supõe uma racionalidade perfeita. A melhor proposta quanto à 'quantidade' de racionalidade necessária à suposição de racionalidade da TSI é talvez a de uma racionalidade mínima. Na proposta de C. Cherniak<sup>366</sup>, 'racionalidade mínima' significa que não existe um conjunto específico de inferências que os estados mentais de um agente tenham que manifestar para este ser candidato à descrição intencional. O agente deve manifestar apenas um conjunto razoavelmente substancial das inferências requeridas a um agente racional perfeito. O que é importante na noção é o facto de ela se opôr a um apriorismo definitivo quanto a racionalidade: não é possível enumerar a priori um núcleo de crenças verdadeiras e de princípios racionais sem os quais não existem agentes racionais<sup>367</sup>.

Os núcleos da racionalidade que são a racionalidade instrumental e a racionalidade mínima, dão-se tão bem com imperfeições constitutivas como com o conceito de optimização. Nenhum requer (pelo contrário parecem permitir excluir) uma racionalidade óptima ou perfeita. Evidentemente, nenhuma qualidade intrínseca dos produtos e processos de uma racionalidade que não é óptima nem perfeita (embora seja exigida) fica garantida. Mas nada comprometia à partida a TSI com semelhante ideia. Tudo o que há na proposta da TSI como teoria da interpretação supondo a racionalidade é a afirmação de que a racionalidade é uma condição da caracterização intencional. Nada na TSI obriga a que a racionalidade *perfeita* seja uma condição *necessária* da caracterização intencional.

Constitui evidentemente um problema, dada uma tal 'natureza' da racionalidade, o facto de a racionalidade ser considerada como chave para a psicologia e para a constituição da pessoalidade. De facto, a pessoalidade 'densa' dependeria em última análise de uma interpretação, de uma 'suposição de racionalidade' acerca de si. Mas talvez a situação não seja tão estranha assim: se em termos de racionalidade se encontra aquilo que se

<sup>363</sup> Qual seria a alternativa? Como, senão através de actividade cognitiva, se avaliaria a actividade cognitiva?

<sup>364</sup> Partindo do constrangimento ou obrigação de nos supormos a nós próprios como modelo, S. Stich [Stich 1993:50] fala de um princípio de chauvinismo intencional. Ser intencionalmente descritível requer a presença uma dinâmica cognitiva semelhante à nossa e criaturas muito diferentes de nós não serão, assim, intencionalmente descritíveis. No entanto, um tal limite imposto à [ir]racionalidade tem razões relativas ao observador, que são perfeitamente desinteressantes: a in-descritibilidade intencional não assinala, nestas condições, nenhuma fronteira natural. O que Stich procura é precisamente extrair as consequências deste chauvinismo intencional.

<sup>365</sup> Cf. por exemplo Stich 1993.

<sup>366</sup> Cherniak 1986.

<sup>367</sup> Dada a ligação entre racionalidade e intencionalidade e esta vagueza da racionalidade é claro que uma caracterização intencional não poderá também ser uma questão de tudo ou nada – mas essa ideia está, precisamente, de acordo com o anti-representacionismo da TSI enquanto teoria do mental.

coloca a partir do ponto de vista do intérprete, o que isso significa é que o ponto de vista do intérprete é auto-constitutivo. Por trivial que pareça a observação, isto significa por exemplo que a racionalidade envolvida na constituição de si não é uma característica do funcionamento neurofisiológico dos neurónios ou de outro *hardware* qualquer do agente, nem do processos computacionais definidos sobre entidades intrinsecamente representacionais. O seu estatuto é o do artificial de H. Simon, o interface<sup>368</sup>.

Por outro lado, a indispensabilidade da racionalidade na teoria do mental, não parece ser necessariamente acompanhada pela convicção de que alguma teoria (por exemplo uma teoria formal como a teoria da decisão ou mesmo a lógica) eventualmente exibirá *uma verdadeira natureza* da racionalidade. A situação é mesmo compatível com a ideia segundo a qual a racionalidade não tem uma verdadeira natureza<sup>369</sup>. O ponto de referência é apenas a adaptação suficientemente boa ao ambiente do comportamento, guiado por ‘representações’, de agentes cognitivos físicos dotados de uma estrutura de ‘finalidades’. Será que nestas condições a racionalidade poderá ter pretensões a fundar alguma coisa? Parte do trabalho de uma teoria filosófica da racionalidade consistirá mais propriamente em justificar o estatuto de equilíbrio reflectido entre práticas de racionalidade, formalizadas e comuns, e respectivas intuições, por um lado e formalização de princípios e regras por outro do que na descrição de alguma realidade substantiva<sup>370</sup>. O mínimo que se pode concluir é que não é imediatamente legítimo evocar dados empíricos ou a evolução para falar de ‘Uma Racionalidade’. De forma mais extrema, dir-se-á, como S. Stich em *The Fragmentation of Reason*<sup>371</sup>, que é impossível formular constrangimentos a priori para todos os agentes racionais possíveis. Não existe A Racionalidade, o Agente Racional.

<sup>368</sup> Simon 1969.

<sup>369</sup> A racionalidade não consiste por exemplo em fechamento dedutivo. Só este, note-se, implicaria a capacidade de fazer todas as inferências logicamente válidas a partir das nossas crenças. Ora isso não nos acontece e nem é possível. Se somos racionais, a nossa racionalidade não é desse tipo. A racionalidade não consiste também nem numa capacidade inferencial perfeita, nem numa coerência total no corpo das crenças.<sup>370</sup> Assim, deve ser por exemplo encarada a possibilidade de a força das nossas intuições a favor de certos ‘princípios de racionalidade’ (por exemplo princípios lógicos) não ser prova definitiva de nada. A aparente auto-evidência de conexões de conteúdos pensados, a força do ‘segue-se de...’, não é, em tais circunstâncias, garantia alguma de que uma Conexão se sustente. A inexistência de qualquer ponto de apoio exterior ao funcionamento desses mesmos princípios justifica por exemplo o facto de a investigação de leis mais profundas da racionalidade (por exemplo leis mais profundas da lógica) não continuar do mesmo modo que continua a investigação de leis mais profundas da física (ideia explorada (ou melhor, problema caracterizado) em Nozick 1993).

<sup>371</sup> Stich 1993. Que de resto Stich pensa poder ser acompanhada por um pluralismo cognitivo normativo, o que significa que embora não exista uma relação entre o bom raciocínio e a Verdade, existem relações entre bom pensamento e verdade. Isto significa que não se abdica da avaliação de estratégias cognitivas, mas não se ignora a sua dimensão de estratégia (é isto o pragmatismo). Apenas uma posição semelhante leva a sério, de resto, o facto de o núcleo mínimo de racionalidade ser um núcleo instrumental. Uma teoria do mental como interpretação não pode deixar de considerar esse núcleo instrumental, o que significa que ela pressupõe uma avaliação consequencialista dos sistemas cognitivos. A questão seguinte é saber o que importa aos agentes conseguirem, atingirem (porque é que havia de ser a Verdade? Porque é que crenças verdadeiras seriam preferidas relativamente a processos pragmaticamente sancionados, com uma conexão óbvia com aquilo que o agente deseja, quer?). Esta situação é o princípio do aludido pragmatismo e é desenvolvida em Stich 1993.

Que conclusão final se deverá retirar das discussões acerca da natureza da racionalidade na filosofia da mente e da acção e na psicologia cognitiva? Pelo menos que subsiste uma grande ambiguidade entre um uso ‘descritivo’ da noção de racionalidade, que pareceria inocentemente cognitivo, e um uso normativo da noção, o que mostra que não está disponível nenhuma noção descritiva simples de racionalidade e portanto que é desde logo bastante problemático evocar sem mais a racionalidade de um agente. Mas, sobretudo, deve-se concluir que a passagem de uma atribuição de racionalidade a sistemas cognitivos, provavelmente inadivável na teoria da cognição e da acção, para alguma pretensão de universalidade ou de natureza profunda e fundamentadora da Racionalidade das pessoas não é legítima.

**5. Questões narrativas. Ser ou não ser um indivíduo: gastar palavras e outros materiais em idiosincrasias e contingências. O que é que (me) satisfaz.**

Voltando à descrição e à avaliação que as pessoas fazem de si próprias voltamos ao terceiro tipo de questões da IP, as questões narrativas. O estatuto desta narrativa de si pode ser concebido de maneiras muito diferentes. Antes de considerar esquematicamente certas concepções alternativas do estatuto da narrativa de si procurarei caracterizar o teor desta.

A narrativa de si traduz-se em grande medida num (1) *sentido idiosincrático daquilo que é importante e possível* e que faz com que uma pessoa difira de outras pessoas. É em grande parte esse sentido idiosincrático daquilo que é importante e possível cuja extinção se teme quando se teme a morte. Evidentemente um tal sentido daquilo que é importante e possível provem de materiais vários, contingentemente disponíveis e que podem ser tomados por muitas pessoas. Daí que um outro aspecto presente nesta narrativa de si, ou que a condiciona, seja, utilizando a expressão de H. Bloom, (2) a *ansiedade da influência*<sup>372</sup>, o temor de ser apenas uma cópia, uma réplica, o temor de que a originalidade e a individualidade não sejam reconhecidas.

Mas que originalidade é essa a que um indivíduo pretende, se nenhum indivíduo se cria totalmente a si próprio? A consciência dos (3) *limites da auto-criação*, i.e. a consciência das razões pelas quais a auto-criação não é nunca total ou totalmente voluntária, é o ponto em que as

<sup>372</sup> Bloom 1963. Bloom aplica, o conceito de *ansiedade da influência* ao criador artístico original. No entanto, o conceito aplica-se identicamente à narrativa psicológica de si que estou a considerar.

questões da auto-criação e da auto-consciência se cruzam. Esta dimensão da IP é tratada na filosofia sob muitas formas<sup>373</sup>.

Em contraste com as dimensões cognitiva e metafísica, é ao nível das questões narrativas da IP que (4) *o sujeito encara a questão da sua própria natureza da forma mais directa possível*<sup>374</sup>. O tipo de auto-consciência ou auto-apercebimento (*self-awareness*) aqui em causa não é simplesmente o apercebimento de um si prévio, bruto, sub-pessoal ou a prova da existência de uma determinada natureza metafísica no mundo mas sim um *envolvimento prático da entidade que se sente e que se pensa num dinamismo de auto-criação e de auto-avaliação*<sup>375</sup>.

A auto-avaliação não é apenas uma avaliação (por exemplo kantiana ou utilitarista) daquilo que se faz mas uma (5) *avaliação daquilo que se é*. Esta é uma dimensão frequentemente ocultada pela dominância na filosofia moral das análises deontológicas e utilitaristas. A importância da avaliação forte (avaliação qualitativa ou reflexiva, na terminologia de C. Taylor<sup>376</sup>), característica de seres que ‘cuidam do seu ser’, é uma razão pela qual as questões da IP não podem ser identificadas exclusivamente com as questões cognitivas ou metafísicas atrás descritas.

Se os materiais da auto-criação têm origens e marcas, saber o que é distinto e específico naquilo que se é será em grande parte ter interpretado essas marcas. É essa interpretação que inclina a conceder um determinado estatuto à narrativa de si e é precisamente nesse ponto que perspectivas filosóficas totalmente opostas se enfrentam. De acordo com a posição assumida será mais ou menos satisfatório para um indivíduo ser um indivíduo, (i.e. apenas um, apenas este, apenas alguma coisa que não dura muito tempo). Nomeadamente, de acordo com muitas perspectivas filosóficas (de resto muito diferentes entre si<sup>377</sup>) apenas o universal satisfaz.

De acordo com uma perspectiva – a que vou por conveniência chamar *nietzscheana* – da IP como auto-criação, a IP do indivíduo tem a ver

<sup>373</sup> Dou apenas alguns exemplos: tais limites são tratados por T. Nagel e B. Williams através do conceito de ‘sorte moral’ (*moral luck*), por H. Frankfurt através da análise das relações entre a liberdade e o conceito de pessoa, por Nietzsche através do reconhecimento da contingência inscrito na ideia de amor fatum, pelos vários existencialistas através das análises da liberdade, etc (cf. Nagel 1979, Williams 1981, Frankfurt 1971). O conceito de sorte moral foi proposto por Williams e no seu núcleo está aquilo que escapa à caracterização de uma situação moral como sendo relativa a decisões racionais. Aquilo que escapa é por exemplo o carácter do agente e as circunstâncias da acção. Em geral, o conceito de sorte moral pretende chamar a atenção para o facto de aquilo que alguém faz depender de factores que estão totalmente fora do controlo dessa pessoa (por exemplo uma pessoa terá ou não cometido homicídio conforme a pessoa sobre quem disparou estivesse ou não a usar um colete à prova de bala). Em Frankfurt 1971 o núcleo da questão da ‘relação entre a liberdade da vontade e o conceito de pessoa’ reside nas *Volições de Segunda Ordem*. Estas são, segundo Frankfurt, *fulcrais* no conceito de pessoa. A ideia de *Volições de Segunda Ordem* pretende nomear a possibilidade humana de querer ou não querer aquilo que se deseja. Apenas com esta possibilidade surge a responsabilidade pelo que se é e pelo que se quer.

<sup>374</sup> I.e. é a este nível que se lida com questões relativas à voluntariedade e compromisso com a identidade e racionalidade próprias.

<sup>375</sup> Separo as questões por razões analíticas. A situação não é, evidentemente, assim tão simples.

<sup>376</sup> Cf. Por exemplo Taylor 1976.

<sup>377</sup> Referi atrás o utilitarismo e o kantismo.

com contingências, é a cada passo algo que poderia não ter sido e exclui a cada passo o que não será e poderia ter sido. Não existe nela um núcleo que seria uma ‘lista universal do que é importante e possível para todos os humanos’, uma ‘descrição única e verdadeira da condição humana’, um ‘desvelamento da essência universal de pessoa’<sup>378</sup>. De acordo com esta visão, nenhum universal ou transcendência trará consolação (a única ‘consolação’ possível é a própria auto-criação) e o auto-conhecimento não é uma descoberta da essência oculta do humano mas uma confrontação da contingência própria<sup>379</sup>. Ela opõe-se à ideia de expressividade, segundo a qual na criação de si se trataria de descobrir e não de moldar e de fazer e à ideia de um fim último para a espécie ou para a história, que se ligaria então ao reconhecimento em si próprio da essência do humano

Se existisse um fim último semelhante, ou uma essência universal de pessoa (por exemplo identificada com a racionalidade), as contingências de uma vida particular não seriam importantes por si e em última análise a extinção individual não importaria. Se no entanto não existe um tal fim último a individualidade do indivíduo volta a importar.

De uma outra perspectiva, a que vou chamar estabilizadora, que não encontra satisfação no indivíduo e à qual apenas um certo universal satisfaz, algo na IP passa por atingir a universalidade transcendendo a contingência, a particularidade de se ser apenas um indivíduo. De acordo com essa perspectiva estabilizadora, é possível exibir universalidade e necessidade no individual e no contingente. Não é apenas um Hegel que se deixa tentar pela perspectiva estabilizadora perante a história e a moralidade. Também o importantíssimo filão kantiano da filosofia moral contemporânea a tem no seu centro e pelo menos desde Kant, um intuito moralista e um intuito romântico se guerreiam a propósito da concepção não apenas da moralidade mas do estatuto do indivíduo e da história pessoal deste. Um breve desvio pela temática artística do novo e do génio facilmente mostraria uma certa exclusão que o pensamento moral de origem kantiana opera na problemática da IP, uma exclusão no entanto susceptível de ser ‘corrigida’ a partir de interpretações românticas do pensamento estético do próprio Kant<sup>380</sup>.

<sup>378</sup> Todas as expressões referem uma certa universalidade de características daquilo que é (narrativamente) humano.

<sup>379</sup> É essa recriação do acontecer contingente num ‘eu quis assim’ que cai sob o mote nietszcheano do *Amor fatum* (nada na forma nietszcheana de conceber o indivíduo enquanto narrativa pessoal de contingências nos impede de considerar os humanos são produtos da natureza, sujeitos à causalção natural, relativamente aos quais se colocam as questões cognitivas e metafísicas até aqui caracterizadas).

<sup>380</sup> Destacar-se de si, do passado e fazer-se, criar-se, é, como é fácil verificar, uma temática não apenas psicológica mas artística, ligada às questões do novo, da originalidade e da vanguarda e à concepção romântica segundo a qual aquilo que é mais humano no humano seria de alguma forma ‘criador’. A importância da questão nas artes é óbvia e poderia assumir a seguinte forma: tem que haver algo de novo para fazer, para



A dimensão de auto-criação e auto-avaliação da IP cujas duas versões procurei caracterizar é ocultada quando se afirma que de um ponto de vista objectivo uma pessoa é por exemplo ‘cérebro neste estado’, ‘coração neste estado’, ‘socialmente moldado desta maneira’, ou ‘agente racional’. Ela relaciona-se com um aspecto da relação de si a si diferente do cálculo racional meios-fins, que está sempre de algum modo em funcionamento em cada um de nós e que tem a forma ‘Será que eu quero mesmo ser aquilo que sou?’<sup>381</sup> Não apenas a questão está sempre em funcionamento em cada pessoa como duas respostas, uma positiva e uma negativa, estão provavelmente sempre a ser dadas. Ligar-se a si próprio, ‘querer-se’, comprometer-se consigo próprio, por um lado e destacar-se de si próprio, pensar que não se é apenas o já-sido estabilizado, nem nunca apenas exactamente ‘isto’ por outro lado são provavelmente componentes inseparáveis do sentido narrativo da IP.

Em *Contingency, Irony and Solidarity*<sup>382</sup> R. Rorty analisa o choque numa mesma pessoa das duas concepções da narrativa de si a que se chamou aqui nietzscheana e estabilizadora. Mais uma vez, embora num sentido diferente do de Parfit, agora numa versão narrativa e não já metafísica, está em jogo a importância da identidade própria para o indivíduo. Uma pessoa é uma, é a sua história, a sua lista idiossincrática de coisas importantes. Mas em que sentido é importante ser um indivíduo, este indivíduo? Em que medida é que ser este e apenas este indivíduo satisfaz o próprio indivíduo? O fim da pessoa é o fim da lista contingente que compõe o seu sentido idiossincrático daquilo que é importante e possível. E que isso desapareça importa? Bom, apenas porque (ou apenas se) se é isso ou se é ser isso que é importante.

De facto, quando se trata de conceber o estatuto do indivíduo e da sua história pessoal talvez não haja escolha possível entre a concepção moralista e a ambição da universalidade por um lado, e a concepção romântica por outro. De um lado, fica a satisfação naquilo que é mais do que a

*ser, são o que estou eu aqui a fazer? Mas esse algo de novo, se não há um fim último claramente definido [social, histórico, civilizacional], o que será? Uma determinada decisão ou inclinação quanto à existência ou não existência de fim último, essência, verdadeira natureza, direcção da história, determina mesmo que implicitamente, se se considerará importante ou não ‘gastar palavras ou outros materiais com idiossincrasias e contingências’. Repare-se por exemplo que muita da arte do século XX, nomeadamente correntes deflacionárias, mínimas, cultivadoras da irrisão, parte da decisão de que é [afinal] importante gastar materiais com idiossincrasias e contingências, abdicando do lirismo associado a um pathos da universalidade, da essência e da expressão de uma ‘verdadeira natureza’. Na criação pessoal a questão é semelhante. O que está em jogo a partir do momento em que se abdica do fim último é admitir que não há indivíduo humano paradigmático, que a IP narrativa é um tecido de contingências, que qualquer coisa serve para cristalizar o sentido de IP de uma pessoa, que não há o desumano ou o contra natura nas variadíssimas narrativas de si dos humanos lou pelo menos que nenhuma forma é descontinua relativamente a formas mais ortodoxas e centradas de IP narrativa].*

<sup>381</sup> Este é o ‘problema de Frankfurt’ (Frankfurt 1971).

<sup>382</sup> Rorty 1989.

individualidade, do outro fica a importância de ser este indivíduo para o próprio indivíduo<sup>383</sup>.

Voltando a Rawls, é inegável que o seu objecto (a concepção política de pessoa, aquilo que deve-racionalmente-ser como base de motivação moral para um indivíduo) se identifica muito mais com uma concepção estabilizadora de IP narrativa, em cujo núcleo se encontra nomeadamente a racionalidade, do que com uma concepção nietzscheana. Sem isso seria por exemplo impossível separar num indivíduo o sentido da justiça da concepção de bem. O problema é que desta forma de certo modo se diminui a importância para o indivíduo de ser esse indivíduo, que excede aquilo que no interesse e cuidado do indivíduo se refere à sociedade na qual vive e aos funcionamentos desta. Mas precisamente e como se fez notar no início do presente artigo, uma das principais intenções de *Liberalismo Político* é separar o *comprehensive* do *political* na própria noção de pessoa, sendo desse modo que Rawls pretende legitimar o pluralismo das identidades narrativas nas sociedades contemporâneas, insistindo no estatuto do 'razoável', ligado à coexistência de diferenças irreduzíveis na esfera pública, por contraste com 'o racional'. Esse desejo de Rawls é ele próprio razoável, mas talvez Rawls devesse admitir outra coisa: talvez ele devesse admitir que sem a introdução de uma concepção determinada de pessoa não se disporia do puro *political*, totalmente distinto do *comprehensive*, para conceber a identidade pessoal e consequentemente toda a teoria da justiça. Sem a introdução dessa concepção não seria nomeadamente legítimo supor um dualismo simples entre fins pessoais e princípios com os quais todos os agentes racionais devem concordar.

### Conclusão

Procurei neste artigo evidenciar alguns aspectos do inevitável compromisso de Rawls com determinadas doutrinas acerca da identidade pessoal na caracterização da Posição Original e da noção de pessoa em geral. De facto, a justiça como equidade está comprometida com algo como um agente racional idealizado, um deliberador kantiano naturalmente unificado, uma motivação cognitivista e uma universalidade da Racionalidade, nuclear nos indivíduos, que seria fundadora. Assim, mesmo admitindo as vantagens da independência da concepção política de pessoa relativamente a assuntos empíricos (tais como os assuntos em causa nos problemas cognitivos e narrativos da identidade pessoal), mesmo admitindo que a

<sup>383</sup> Está aqui também em jogo a possibilidade de encontrar ou não encontrar qualquer 'conteúdo emancipatório' na natureza dos indivíduos, dada a ligação deste 'conteúdo emancipatório' à presunção de universalidade e a uma certa teleologia.

psicologia moral em causa na teoria da justiça é, nos termos de Rawls, filosófica e não psicológica, não é possível a Rawls abordar o seu assunto – indivíduos que deliberam racionalmente na Posição Original acerca do tipo de sociedade em que viverão – sem ideias definidas, mais ou menos declaradas, acerca da natureza da personalidade e da racionalidade. Apenas estas ideias permitem a Rawls utilizar conceitos tais como *agente racional*, *projecto racional de vida* e *pessoa como envolvendo primeiro um sentido de justiça e depois uma concepção de bem*. Acontece que vários pontos das investigações acerca de racionalidade e de identidade representam obstáculos no caminho de Rawls. Nestas circunstâncias, o mínimo que se pode afirmar é que a posição rawlsiana acerca da identidade pessoal das partes na Posição Original é uma decisão, uma escolha, que não se encontra pura e simplesmente fundada na Racionalidade.

Mas qual seria a consequência de reconhecer que se está a lidar com uma construção específica da noção de pessoa para sustentar a teoria da justiça, algo que de resto o próprio Rawls incipientemente faz? Antes de mais, a ideia segundo a qual alguma psicologia humana é descritível de forma neutra quando ‘os fins e as ligações estão cobertos pelo véu da ignorância’ tem que ser reconsiderada. Mas essa ideia, enquanto abstracção da identidade narrativa das partes, não é propriamente problemática, e pode até ser bastante útil. O maior problema da figuração das pessoas na Posição Original não é propriamente a abstracção mas sim a concepção da motivação.

Como afirma Blackburn, o liberalismo só ganharia «em reconhecer que certamente está a promover uma concepção de bem entre outras»<sup>384</sup> e que o modelo de sociedade descrito por Rawls é uma questão de opção. Esse reconhecimento corresponde a uma alteração na psicologia moral subjacente à teoria da justiça, a uma correcção pela qual se reconhece que não é legítimo supor um dualismo simples entre fins pessoais e princípios com os quais todos os agentes racionais devem concordar. A ‘autoridade’ da proposta rawlsiana não lhe advém de uma fundação na Racionalidade mas da vida (boa) que ela pode proporcionar aos indivíduos, que é admirável e desejável e como tal motivadora. Tais reinterpretações impedir-nos-iam além do mais de ver na justiça rawlsiana uma imagem de sociedade como rede de contratos para benefício próprio, susceptível de ser abandonada a qualquer momento por razões de interesse egoísta, já que seriam razões de interesse egoísta, racional-calculatório, a conduzir ao contrato inicial. Tudo isto significa que as pretensões de neutralidade e abstracção na concepção da identidade pessoal na Posição Original caem: existe na

<sup>384</sup> Blackburn 2000: 275.

concepção de identidade pessoal na Posição Original uma concepção determinada de bem, que não é mínima nem neutra. Nada nos impede no entanto de considerar que a suposta neutralidade era desnecessária. Como afirma Blackburn, «A autoridade do contrato justo é uma coisa: o mito de que ele se auto-sustenta por se fundar na racionalidade é uma coisa totalmente diferente, e não lhe faz serviço nenhum»<sup>385</sup>.

O que tudo isto significa também, evidentemente, é que contrariamente às suas intenções mais recentes, Rawls não chega a libertar-se dos resquícios substanciais do kantismo, nomeadamente da metafísica do eu e da deliberação. O kantismo é uma concepção determinada – precisamente uma concepção cognitivista como se diria na literatura contemporânea dedicada à questão das razões para agir – daquilo que é um agente racional, a sua unidade e a sua motivação para agir, centrada não naquilo que os indivíduos desejam mas naquilo que deve-ser. Uma profissão de fé kantiana faz-nos fechar os olhos perante a possibilidade de agentes morais humanos serem constituídos de uma forma tal que o facto de alguma coisa dever-ser os deixa totalmente indiferentes. O afastamento do compromisso da teoria da justiça com uma metafísica kantiana do eu e da deliberação não tem no entanto que deitar a perder a teoria rawlsiana, mas apenas que lhe atribuir um outro estatuto. A situação poderia ser formulada do seguinte modo: devemos ver a sociedade a que a interacção entre as partes na Posição Original conduz como aquilo que pessoas como Rawls (e talvez alguns de nós, mas certamente não todos) *desejam*, mas que não se apoia certamente numa incontestável universalidade da Racionalidade.

### Referências

Arrow, Kenneth, 1963, *Social Choice and Individual Values*, New York, John Wiley and Sons.

Bermudez, J.L., Marcel, A. & Eilan, N, 1995, *The Body and the Self*, Cambridge MA, MIT Press.

Blackburn, Simon, 2000, *Ruling Passions – A Theory of Practical Reasoning*, Oxford, Clarendon Press.

Bloom, Harold, 1963, *The Anxiety of Influence*, Oxford, Oxford University Press.

Cherniak, Christopher, 1986, *Minimal Rationality*, Cambridge MA, MIT Press.

Damásio, António, 1999, *The Feeling of What Happens*, New York, Harcourt Brace.

<sup>385</sup> Blackburn 1998: 271

Davidson, Donald, 1980, Hempel on Explaining Action, in Davidson, D., 1980, *Essays on Actions and Events*, Oxford, Oxford University Press.

Davidson, Donald, 1982, Rational Animals, *Dialectica*, 36.

Dennett, Daniel, 1981, Conditions of Personhood, in Dennett, D., 1981, *Brainstorms*, Cambridge MA, MIT Press.

Dennett, Daniel, 1987, *The Intentional Stance*, Cambridge MA, MIT Press.

Dennett, Daniel, 1988, The Moral First Aid Manual (in S. McMurrin ed., *Tanner Lectures on Human Values*, III, Salt Lake City, University of Utah Press).

Dennett, Daniel, 1991, *Consciousness Explained*, Boston, Little Brown.

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

Frankfurt, Harry, 1971, Freedom of Will and the Concept of Person, *Journal of Philosophy*, 68.

Locke, John, 1999 [1690], *Ensaio sobre o Entendimento Humano*, Lisboa, Fundação Calouste Gulbenkian.

Miguens, Sofia, Problemas da Identidade Pessoal, *Revista da Faculdade de Letras-Série de Filosofia*, nº18. Também neste volume.

Nagel, Thomas, 1979b, Moral Luck, in *Mortal Questions*, Cambridge, Cambridge University Press.

Nagel, Thomas, 1995, Rawls on Justice, in *Other Minds – Critical Essays 1969-1994*, Oxford, Oxford University Press.

Nozick, Robert, 1993, *The Nature of Rationality*, Princeton, Princeton University Press.

Parfit, Derek, 1984, *Reasons and Persons*, Oxford, Oxford University Press.

Quine, W. V., 1960, *Word and Object*, Cambridge MA, MIT Press.

Ramsey, F., 1926, Truth and probability, in H. Mellor (ed), 1990, *Philosophical Papers*, Cambridge, Cambridge University Press.

Rawls, J., 1993 a [1971] *Uma Teoria da Justiça*, Lisboa, Presença.

Rawls, John, 1993 b, *Political Liberalism*, New York, Columbia University Press.

Rorty, Richard, 1989, *Contingency, Irony, Solidarity*, Cambridge, Cambridge University Press.

Simon, Herbert, 1998 [1969] *The Sciences of the Artificial*, Cambridge MA, MIT Press.

Stich, Stephen, 1993, *The Fragmentation of Reason*, Cambridge MA, MIT Press.

Taylor, Charles, 1976, Responsibility for self, in Amélie Rorty (ed), *The Identities of Persons*, Los Angeles, University of California Press.

Tversky, A. & Kahnemann, D., 1981, The Framing of Decisions and the Psychology of Choice, *Science* 211, 453-458.

Tversky, A. & Kahnemann, D., 1993, Probabilistic Reasoning, in A. Goldman (ed), *Readings in Philosophy and Cognitive Science*, Cambridge MA, MIT Press.

Williams, Bernard, 1981, Moral Luck, in *Moral Luck*, Cambridge, Cambridge University Press.

Von Neumann J. & Morgenstern, O., 1944, *Theory of Games and Economic Behavior*, New York, Wiley.

Zilhão, António, 1998/1999, Folk-psychology, Rationality and Human Action, *Grazer Philosophische Studien*, 56.



## I. Introdução

Ouvimos de manhã uma introdução ao problema metafísico do livre arbítrio, que nos deu o estado da questão do ponto de vista dos argumentos e posições em campo (pelo André Barata<sup>387</sup>), e também uma discussão do papel das razões na explicação da acção (feita pelo Ricardo Santos<sup>388</sup>). Embora eu vá falar de filosofia da mente, o que vou fazer retoma quer as questões metafísicas quer o papel das razões na explicação da acção. O que vou fazer é falar sobre as posições de D. Dennett acerca de pessoalidade (*personhood*) e liberdade. Estas posições (com as quais muitos terão tido ocasião de contactar quando Dennett proferiu aqui em Lisboa, em 2004, a primeira das suas Petrus Hispanus Lectures, intitulada *Rational Avoidance in a Deterministic World*) constituem o núcleo da filosofia prática de Dennett e estão expostas em obras como *Elbow Room* (1983) e *Freedom Evolves* (2003), e também em artigos importantes como *Conditions of Personhood*, *Giving Libertarians What They Think They Want*, *Mechanism and Responsibility* e *How to change your mind* (reunidos em *Brainstorms*, 1978, sob o título ‘Free Will and Personhood’).

A primeira coisa a dizer é que as posições de Dennett sobre pessoalidade e liberdade (respectivamente uma posição normativa e anti-essencialista e uma posição compatibilista) pressupõem as suas teses gerais acerca da natureza do mental, nomeadamente as (i) posições acerca da natureza da intencionalidade e da consciência (desenvolvidas no âmbito da Teoria dos Sistemas Intencionais<sup>389</sup> e do Modelo

<sup>386</sup> Este texto corresponde à comunicação apresentada no 3º Encontro Nacional de Professores de Filosofia organizado pela Sociedade Portuguesa de Filosofia e enquadra-se ainda no Projecto POCTI/FIL/55555/2004 (*Rationality, Belief, Desire II – from cognitive science to philosophy*).

<sup>387</sup> Filósofo, UBI.

<sup>388</sup> Filósofo, U. Évora.

<sup>389</sup> Cf. Dennett 1987.



dos Esboços Múltiplos<sup>390</sup>) e as (ii) posições acerca da evolução do design por seleção natural, e da natureza do design<sup>391</sup>. São essas as teses que nos permitem conceber o lugar da mente no mundo físico e é esse o âmbito no qual Dennett discute a liberdade. Como o nosso propósito aqui não é discutir o mental em geral mas a liberdade, vou passar por cima dessas questões (acerca de intencionalidade, consciência, design) e avançar para os problemas e as teses da filosofia prática.

Antes de mais queria fazer uma pergunta e responder-lhe. A pergunta é a seguinte: o que pode trazer a filosofia da mente a uma discussão acerca da liberdade, uma discussão que parece ser de âmbito ético-político e metafísico, e que é frequentemente tratada de forma directa através de argumentos nesses campos? Basicamente, pode trazer uma *teoria do agente*, daquilo a que chamamos (a partir de dentro, i.e. fenomenologicamente) 'agir' e daquilo que deve ser o caso em termos de cognição (portanto ao nível sub-pessoal) para que descrições mentalistas do mundo tais como 'uma pessoa age livremente' se apliquem àquilo que acontece. Este é o caminho através do qual Dennett articula a teoria da cognição com questões metafísicas acerca de liberdade, e parece útil e necessário percorrê-lo, sobretudo porque os argumentos em ética, filosofia política e em geral nas discussões metafísicas do livre arbítrio demasiado frequentemente dão como garantida e não problemática a noção de agente e conceitos adjacentes tais como a unidade e a continuidade do agente, a natureza da deliberação, a voluntariedade e a racionalidade da acção, etc. Cabe portanto à filosofia da mente enquanto meta-teoria da ciência cognitiva impedir suposições irrealistas do ponto de vista cognitivo quanto à natureza dos agentes humanos.

Em segundo lugar, um esclarecimento. A suposição metodológica fundamental da filosofia prática de Dennett é a seguinte: para virmos a compreender de forma empiricamente informada a natureza da acção livre no tipo de agentes que somos, devemos começar por ter claro aquilo que, enquanto agentes, não somos. E Dennett pensa que *enquanto agentes não somos máquinas semânticas com centros reais* mas sim sistemas cognitivos físicos, agentes racionais reais. O que significa dizer que não somos máquinas semânticas? 'Máquina semântica' é o termo que Dennett utiliza para aquilo que na literatura filosófica costuma ser chamado o Agente Racional, um agente que seria movido a significações e razões por oposição a causas (esse agente seria por exemplo uma vontade pura, física e cognitivamente ilimitada). Ao dizer que não somos máquinas semânticas Dennett está a dizer que enquanto agentes não somos nem vontades numéricas

<sup>390</sup> Cf. Dennett 1991.

<sup>391</sup> Cf. Dennett 1995.

(«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>392</sup>), nem calculadores omniscientes utilitaristas, perfeitos maximizadores da utilidade esperada. Nada disto significa que não sejamos agentes capazes de deliberação e escolha.

O que significa dizer que os nossos eus não são centros reais? Significa que mesmo sendo o eu um centro virtual de controlo essencial em agentes como nós, na medida em que um eu é uma auto-representação no cérebro e enquanto tal efeito do funcionamento sub-pessoal do sistema cognitivo, ele não pode ser simplesmente assumido como dado a priori, como lugar de origem das acções. Em suma, não dispomos à partida de um não problemático sujeito da vontade e aliás, o facto de ser possível uma teoria sub-pessoal do Eu constitui por si evidência da existência de um fundo irreduzivelmente não-mental em qualquer entidade que venha a aperceber-se de si mesma ao nível pessoal como una e unificada, autora e iniciadora de acções próprias, e moralmente uma pessoa. Qualquer pessoa tem que ser primeiro um Eu, e essa é uma questão de organização cognitiva sub-pessoal, permitindo uma ‘centralização virtual’ apoiada numa auto-representação. Dito isto, Dennett pensa que humanos podem ser pessoas e agir livremente – é preciso, no entanto, investigar como é que isso é possível, não podemos simplesmente assumir que tais conceitos se aplicam.

Como sabemos, as abordagens do problema da liberdade em filosofia são muitas e distintas entre si: a liberdade pode ser discutida em termos de determinismo e indeterminismo, desaparecimento dos agentes, regimes políticos e instituições, etc. Assim, para termos uma noção do alcance, mas também das limitações, da particular abordagem de Dennett, devemos ter presente que Dennett aborda o problema da liberdade em termos de teoria da mente, mais concretamente como um problema relativo a (i) sistemas cognitivos e descrições mentalistas destes (i.e. descrições em termos de representações, desejos, intenções), (ii) e como um problema relativo a tipos de mentes – mentes humanas versus mentes de outros agentes, por exemplo animais. Segundo Dennett, todas as mentes são possibilitadas pela evolução por selecção natural, especificamente pela evolução do design. No entanto, a liberdade é característica de um único tipo de mentes, as mentes humanas. Para que certas entidades no mundo, sistemas cognitivos físicos, viessem a aperceber-se fenomenologicamente como pessoas e como agentes livres, foi necessário que evoluíssem *deliberadores*

<sup>392</sup> Dennett 1983: 49.

*deterministas capazes de auto-controlo meta-reflexivo*. Do ponto de vista cognitivo, este é conceito básico para falar de liberdade. No entanto, a marca da abordagem de Dennett como teórico da mente é nunca ficar apenas nem pela teoria da cognição nem pela fenomenologia, antes procurar sempre esclarecer a relação entre teoria cognitiva e fenomenologia. Por isso a sua abordagem da liberdade percorre um caminho duplo: por um lado, trata-se de descrever fenomenologicamente aquilo que queremos analisar (acção intencional, pessoalidade), por outro, trata-se de adoptar o ponto de vista subpessoal da teoria da cognição, procurando explicar em termos de ‘engenharia de agentes’ o que tem que ser o caso para que determinadas descrições mentalistas se apliquem a determinados agentes. Vou chamar a estas duas abordagens (i) abordagem *top-down* ou fenomenológica, ou por análise conceptual (i.e. aquele que procede a partir da percepção e descrição mentalista de si por agentes), e (2) abordagem *bottom-up* (a partir dos mecanismos de sistemas cognitivos físicos)<sup>393</sup>.

Como vamos ver, a questão assim colocada (em termos de descrição mentalista e design de agentes) não se cruza sequer com a questão do determinismo ou indeterminismo físico - Dennett crê que agentes livres existem num mundo causalmente determinado (i.e. defende uma posição compatibilista quanto ao livre arbítrio), o que lhe interessa é compreender *como*. Porque é que Dennett crê que agentes livres podem existir num mundo causalmente determinado? Basicamente porque pensa que as questões da liberdade e da responsabilidade de humanos dizem respeito a um particular entendimento ou interpretação de eventos físicos como acções<sup>394</sup>, e que essa é uma questão de macro-nível e não relativas ao *micro-nível* do sistema cognitivo, explicado pela teoria física. O esclarecimento do estatuto de acções de certos eventos não passa assim pela física mas pela teoria (normativa) da mente. Nos termos de Dennett: a vontade livre que vale a pena querer (free will *worth wanting*) não é nem vontade livre numérica nem indeterminismo físico e as questões importantes quanto à liberdade de acção (responsabilidade, culpa, mérito) não são nem relativas a uma vontade não física nem ao determinismo ou indeterminismo ao micro-nível físico antes envolvem apercebimento mentalista de si em agentes com características determinadas.

Apenas uma nota final para acabar de situar a abordagem de Dennett: ela coloca-se, como veremos, por um lado (i) contra a ideia de uma

<sup>393</sup> Este sentido de *top-down* e *bottom-up* afasta-se do usual, mas penso que ajudará a compreender o que está em causa.

<sup>394</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.

‘erosão total daquilo que fazemos por aquilo que acontece’,<sup>395</sup> mas por outro situa-se também (ii) contra a ideia libertária de acção livre como irrupção do agente livre no mundo determinista, iniciação não causada de uma cadeia causal<sup>396</sup>.

## **2. Top-down: a partir da fenomenologia, por análise conceptual. Descrição da acção intencional. Condições de pessoalidade.**

De forma a saber o que convém explicar quando se fala de liberdade devemos começar com a fenomenologia, i.e. com aquilo que aparece aos próprios agentes, em primeira pessoa. Para isso Dennett vai buscar a Elizabeth Anscombe a descrição da intenção na acção (Anscombe 1957, *Intention*). O que diz então Anscombe acerca de acções intencionais? Embora assumindo pressupostos wittgensteinianos, que Dennett não aceita (para Anscombe razões não podem ser causas de acções – o Ricardo Santos falou deste assunto), Anscombe oferece-nos dados importantes. Ela afirma que (i) apenas sob uma determinada descrição movimentos corporais podem ser ou não ser intencionais, (ii) essas descrições podem ser várias, (iii) que não é suficiente que exista apercebimento de comportamentos próprios sob uma descrição mentalista para que a acção seja intencional, o apercebimento tem que ser apercebimento sem observação; e (iv) que apenas satisfeitas estas condições, o agente ele próprio é capaz de oferecer a razão para a sua acção. Só assim a ordem que está lá está de forma tal ‘possuída’ pelo agente que nos permite falar de uma acção intencional tal que merece ser chamada ‘livre’. Recapitulo a definição: agir intencionalmente é aperceber, sem observação, comportamentos próprios sob uma descrição mentalista, sendo o agente capaz de dar as razões desses comportamentos. Noto que 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. Dizemos que agem intencionalmente sistemas capazes de agir por razões e cujas acções podem ser explicadas e avaliadas nos termos das razões pelas quais agiu. Um dos maiores problemas aqui reside no facto de esta pretensão se cruzar com a consciência, trazendo a Dennett um problema muito específico: conceber a passagem – que para ele envolve

<sup>395</sup> Thomas Nagel formula muito bem o risco enfrentado quando se procura conceber a acção de um ponto de vista de terceira pessoa: «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» (Nagel 1979, *Moral Luck*, *Mortal Questions*, p. 38).

<sup>396</sup> Dennett subscreve a seguinte observação de Bernard Williams quanto à ‘vocação extra-mundana 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», B. Williams 1973: 228.

continuidade – entre *free floating rationales* e *full blown reasons* existentes em agentes humanos. *Rationales* para o comportamento de agentes antecedem o uso de razões para controlar o comportamento próprio por agentes - encontrámo-los em qualquer organismo e também em máquinas. Os *rationales* 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. Razões têm que ser possuídas conscientemente pelos agentes e utilizadas para o controlo da acção.

Mas o que nos importa aqui é que já em *Content and Consciousness* Dennett defende que a eficácia das razões no controlo da acção de uma entidade é fundamental para a existência de pessoalidade. Ligar intenção na acção e pessoalidade é por isso um passo essencial, um passo que envolve a questão da consciência (pois sem esta não são possíveis *full blown reasons*). Mas o que é pessoalidade? Olhemos para a questão a partir deste ângulo: Dennett fala de liberdade falando de pessoas e acções. Ora, pessoas e acções são conceitos mentalistas. Por essa razão Dennett não aceita concepções absolutas de pessoa e de acção, antes vê as coisas da seguinte forma: 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). Uma entidade no mundo constitui-se enquanto pessoa através de uma relação com acontecimentos que têm determinadas características e que são as suas acções. Vou procurar explicar melhor este ponto em função da definição lockeana de pessoa, que presumo que conhecem, e da qual Dennet se serve para propor a noção de condições de pessoalidade (*conditions of personhood*).

A noção lockeana tem dois componentes: (i) pessoa é uma entidade auto-consciente e racional (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») e (ii) uma entidade responsável pelas acções próprias (por essa razão, pessoa é ‘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>397</sup>).

A ideia dennettiana de condições de pessoalidade<sup>398</sup> liga um ao outro estes dois componentes da noção lockeana de pessoa. Gostaria de fazer

<sup>397</sup> Usei livremente as seguintes passagens: Locke 1999: 443 [Livro II, XXVII, 11] e Locke 1999: 459 [Livro II, XXVII, 28].

<sup>398</sup> Todas as condições de pessoalidade apontadas são condições necessárias da pessoalidade. No entanto, elas não constituem, em conjunto, condições suficientes.

notar, pensando naqueles que estão interessados noutros temas de filosofia da mente, nomeadamente no tema da consciência, que o facto de o conceito de condições de pessoalidade ligar a auto-consciência às acções próprias é só por si suficiente para mostrar que para Dennett a consciência não é identificável como sensiência ou *qualia*, e que pelo menos na sua perspectiva os problemas da consciência, da acção e da racionalidade não podem ser separadamente tratados. Mas para o que aqui nos ocupa é particularmente importante o facto de a consciência especificamente humana (*full blown consciousness*, aquilo que é característico de um adulto humano, face a um cão ou um gato) depender da agência racional: é isto que torna *o esforço para entender a liberdade essencial para compreender as mentes humanas como tipo específico de mentes entre os outros tipos de mentes*.

Mas a ligação significa também que para uma entidade ser uma pessoa é necessário que determinadas condições relativas à mente sejam preenchidas: pessoa não é um conceito absoluto, aplicável a qualquer indivíduo humano, mesmo que legalmente venhamos a estipular que isso deve ser assim. Condições respeitantes a intencionalidade, consciência, e reconhecimento de outras mentes através de comunicação têm que ser satisfeitas para que a noção se aplique. Resumo-as: o conceito de pessoa aplica-se a: 1. agentes que são sistemas racionais (sentido instrumental), 2. susceptíveis de descrições intencionais, 3. capazes eles próprios de considerar intencionalmente outros sistemas, 4. capazes de comunicação verbal (com o aprofundamento mental inerente, revelado pelas análises griceanas: ter crenças acerca de crenças acerca de crenças), 5. e que são auto-conscientes. O ponto central do artigo *Conditions of Personhood* é que a auto-consciência essencial ao conceito de pessoa depende dos factores enumerados.

Apenas algumas observações acerca desta enumeração: (i) há certas características que as mentes de criaturas linguísticas possuem *por serem linguísticas*, nomeadamente serem Sistemas Intencionais de terceira ordem<sup>399</sup> (terem crenças acerca de crenças acerca de crenças); isto pressupõe que a segunda ordem (ter crenças acerca de crenças, o que ocorre em muitos animais não humanos) não é suficiente para a reciprocidade aqui em causa. O reconhecimento tácito nas crenças acerca de crenças acerca de crenças envolve algo como uma expectativa mútua de cooperação (aliás nestas condições da pessoalidade estão envolvidas, note-se, condições de possibilidade de verdade, sinceridade, justiça – valerá a pena ir por aqui).

<sup>399</sup> Para existir significação não natural, um sistema A deve pretender que um sistema B reconheça que A (in) tenciona que B produza uma dada resposta. Para começar, isto mostra que é a condição 4 (a linguagem) que depende da condição 3 (a reciprocidade) e não vice-versa.

Segunda observação: nada, nas primeiras condições, explica porque é que podemos (se é que podemos) considerar por vezes pessoas individuais como responsáveis pelas suas acções. A chave para a pessoalização da pessoa (no sentido de IP - o que é preciso para uma pessoa ser especificamente esta ou aquela pessoa) é outra e permite-nos aliás compreender por que é que para Dennett a auto-consciência envolve mais do que auto-apercebimento e auto-representação, mais do que ser considerado como racional e intencional por outras entidades que por sua vez se considera como racionais e intencionais. A auto-consciência necessária para a constituição requer auto-avaliação, e um comunicador griceano (que tem crenças acerca de crenças acerca de crenças) pode não ser capaz desta auto-avaliação. Para compreender esta última condição de pessoalidade, Dennett apela às ideias de Ch. Taylor e H. Frankfurt, respectivamente a ideia de avaliação forte (avaliação de si e dos desejos e crenças próprias e não apenas de acções) e a ideia de volições de segunda ordem. O que nos interessa é que na condição 6 está envolvida qualquer coisa como: avaliação como auto-criação e compromisso do agente com a vida mental própria.

Harry Frankfurt - que provavelmente conhecerão - liga a liberdade da acção e o conceito de pessoa. 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 supõe 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). Um comportamento voluntário de tendência para o objecto do desejo apenas constitui uma acção livre se envolver volições de segunda ordem (nos termos de Dennett: apenas com o controlo resultante da avaliação reflexiva que entra assim em jogo está satisfeita a última condição de pessoalidade). A sugestão de H. Frankfurt sobre a relação da liberdade da vontade com o conceito de pessoa traduz-se numa teoria sobre a natureza da liberdade de acordo com a qual seria *a incapacidade de deliberar racionalmente sobre o querer, e não o determinismo*, que impede a liberdade de acção.

Além de volições de segunda ordem, a sexta condição envolve avaliação forte. O conceito é de Ch. Taylor, que o reporta à responsabilidade

pela identidade própria<sup>400</sup>. 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?’ e avalia assim não apenas aquilo que fez/faz mas aquilo que é<sup>401</sup>.

De qualquer forma, e é isso que nos interessa aqui, em *Conditions of Personhood* Dennett defende que apenas cumpridas estas condições a ‘ordem que está lá’ a aparência de propósito, os *free floating rationales* observáveis no comportamento de um sistema, que pode ser de tão variadas maneiras não consciente, não está lá a não ser que esteja em episódios de pensamento consciente. E isso será pessoalidade e liberdade. Até agora Dennett fala de liberdade por análise conceptual. Mas o maior interesse da sua posição é que a esta análise conceptual corresponde uma investigação cognitiva sobre agentes que se apercebem fenomenologicamente como pessoais e livres. É do conjunto das teses que sai a defesa do compatibilismo.

### 3. *Bottom-up*: a partir da teoria da cognição.

Segundo Dennett, aquilo a que chamamos liberdade evoluiu em sistemas físicos finitos, agentes desenhados para lidar com informação incompleta num mundo determinista, e evoluiu como auto-controlo meta-reflexivo. Afirmar que a liberdade *evoluiu* é defender que ela é uma questão de design de agentes (e não de física ou de qualquer coisa como uma dotação metafísica a algumas partes do mundo). Para compreender a liberdade deste ponto de vista da cognição e do design de agentes precisamos antes de mais, segundo Dennett, de clarificar ideias erróneas acerca do determinismo, muito particularmente a ideia segundo a qual o determinismo (formulável como ‘há, a cada instante, exactamente um futuro possível’, como diria por exemplo P. Van Inwagen) envolveria inevitabilidade, ausência de alternativas de acção. Em contrapartida o indeterminismo tornaria possível a liberdade, tornando possíveis alternativas de acção (num mundo determinista não haveria opções reais). Todas estas ideias são falsas, segundo Dennett e de qualquer forma capturam mal a indeterminação requerida para se falar de acção.

De que indeterminação precisamos então para vir a falar de acção? Dennett vai buscar a Hume uma pista «Liberty or chance (...) is nothing

<sup>400</sup> Cf. Taylor 1976, *Responsibility for self*.

<sup>401</sup> Segundo C. Taylor o vocabulário da avaliação é uma parte importante do vocabulário moral pós-nietzscheano e é esquecido pelas acima chamadas éticas da maximização racional.



but the want of determination and a *certain looseness* which we feel in passing or not passing from the idea of one to that of the other» (David Hume, *Treatise of Human Nature* II, III, 2). Tomemos como referência essa *certain looseness* humeana (liberdade do ponto de vista do agente) – o que é que ela envolve?

Para essa *certain looseness* ser assim para nós o indeterminismo não tem que ser verdadeiro: evitabilidade, aleatoriedade, exploração inteligente da aleatoriedade e auto-controlo serão suficientes. Dessa forma, a *certain looseness* será possível num mundo determinista. Quando Dennett defende que para as coisas futuras parecerem poder vir a ser de várias maneiras para um agente não é necessário indeterminismo mas apenas evitabilidade, um leque de opções para um agente, o que pretende sublinhar é que a evitabilidade resulta do design de agentes e não do mundo, directamente, em termos de determinismo ou indeterminismo. É esta posição que determina a sua indiferença perante a colocação do problema do livre arbítrio em termos de determinismo / indeterminismo físico. Um exemplo favorito de Dennett é o Mundo da Vida de J. Conway. A capacidade de passar por cima de micro detalhes de situações é a chave da evitabilidade. (Para uma simulação cf: <http://www.bitstorm.org/gameoflife/> ).

Vejamos, para compreender melhor a posição de Dennett perante o debate canónico entre compatibilistas e incompatibilistas, por que razão no artigo *On Giving the Libertarians What They Say They Want* 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 esta supostamente não poderia existir liberdade da vontade. Como frequentemente acontece, Dennett coloca uma pergunta cândida: onde, em que ponto, no processo de deliberação /decisão num agente é suposto localizar esse indeterminismo? Que trabalho está ele a fazer que não possa ser feito por outra coisa? Nenhum: a indeterminação que segundo Dennett importa para a acção é apenas aleatoriedade, ligada à inexistência de padrões. Ora, aleatoriedade é perfeitamente possível num mundo determinista.

O problema com a aleatoriedade é a conotação de arbitrariedade que parece acompanhá-la. 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. Como se pode retirar a conotação de arbitrariedade à aleatoriedade? Os libertários inclinam-se para fora da natureza: por exemplo de acordo com a doutrina da causação do agente

(*agent-causation*), agir seria provocar determinados eventos sem que o agente seja causado de modo a causar esses eventos. A acção livre seria assim a iniciação não causada de uma cadeia causal (o que chamei irrupção do agente). Dada a irrupção do agente, a sua insubmissão à causalidade, o agente seria capaz de agir exclusivamente por razões.

Mas, diz Dennett, a indeterminação desejada pelos libertários não tem que supôr uma injustificada irrupção do agente puramente movido a razões, nem indeterminismo físico. Porquê? Porque 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 acções, e acontecimentos só podem ser considerados como acções através de descrições mentalistas e normativas. Só é possível fazer sentido de uma acção como constituindo uma particular acção através de uma descrição intencional. Ora, tal acção pode ser realizada de diversas formas, e quando queremos 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. Isto faz com que o comportamento humano possa ser, de forma independente, determinado ou indeterminado do ponto de vista físico e previsível ou imprevisível intencionalmente. Uma acção pode não ser arbitrária (i.e. pode ser intencionalmente previsível) mesmo que consista numa descrição de eventos resultantes de indeterminismo físico, e um agente pode ser intencionalmente imprevisível mesmo se o determinismo físico for verdadeiro. No artigo *On Giving Libertarians What They Say they Want* Dennett oferece uma ilustração deste ponto: o comportamento de uma ‘caixa de respostas’<sup>402</sup>. Trata-se de um dispositivo electrónico imaginado que regista respostas de pessoas a questões simples. Tem dois botões (sim e não) e dois pedais (sim e não) e um écran dividido ao meio, onde aparece de um lado a instrução ‘use os botões’, do outro lado a instrução ‘use os pedais’. Na situação experimental apenas metade do écran será iluminada de cada vez. O sujeito é exposto a questões simples, às quais deve dar como resposta sim ou não, seguindo as instruções do écran. A ‘escolha’ da metade do écran que está em causa de cada vez resulta, de cada vez, de um processo aleatório no dispositivo. É prometido ao sujeito um bónus pelas respostas correctas.

Esta é uma situação-teste para a seguinte questão: em que situação relativa estarão o físico e o teórico intencional relativamente à previsão do comportamento do sujeito, supondo que este é um sistema físico determinista que está a ser afectado pela aleatoriedade no processo de escolha? Trata-se também de um exemplo de previsão intencional de acontecimentos

<sup>402</sup> Dennett 1981e: 290.

baseados em processos físicos cujas características micro-físicas não se conhece e que são aleatórios num determinado sentido. Ora, segundo Dennett, 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 aqui não são os movimentos físicos do sujeito e sim as suas acções.

De resto, 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. Por exemplo: uma acção é praticável de diversas maneiras e são factores aleatórios que determinam a maneira particular de realizar a acção intencionada. 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’, mesmo se existem ‘pontos de deliberação/rejeição/escolha’ possivelmente ubíquos a nível inconsciente e que não aparecem fenomenologicamente. O que interessa é que o facto de a implementação de qualquer acção supôr vários destes pontos de escolha não torna o agente intencionalmente imprevisível.

Continuando: segundo Dennett o grande responsável para que se comece a falar de acção na evolução dos agentes é o espaço de manobra (*elbow room*) e a exploração inteligente da aleatoriedade num mundo determinista. É este espaço de manobra que vale a pena querer quando queremos vontade livre (não o indeterminismo físico nem uma vontade livre não física). Noutras palavras: nenhum agente, nem mesmo um humano, poderá alguma vez ser totalmente livre no sentido de não ser um agente para um mundo determinado e sim um agente ‘para um mundo qualquer’, tendo sempre, em qualquer mundo, possibilidade de acção significativa e de espaço de manobra. Não existe um sistema físico capaz de resposta apropriada a toda e qualquer condição ambiental<sup>403</sup>. Mas o que existe certamente são agentes com mais ou menos *elbow room*: é nesse sentido que a liberdade evolui. Ter mais *elbow room* significa apenas ter um mundo menos determinado, mais próximo do ‘mundo qualquer’ do que do mundo extra-determinado porque muito pouco discriminado das entidades muito simples que são já agentes ou proto-agentes. Será possível dar um exemplo daquilo que Dennett entende por *elbow room*?

<sup>403</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].

Vejamus um exemplo de comportamento animal<sup>404</sup>, que leva Dennett a falar de *sphexishness* do comportamento (em referência à rigidez revelada no comportamento animal, mesmo no comportamento animal aparentemente inteligente, mesmo no nosso). O exemplo é de *Elbow Room* e é do comportamento de uma vespa (a vespa escavadora *Sphex ichneumoneus*). 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 rigidez, esta repetição das rotinas definidoras do comportamento, existe em qualquer agente, em maior ou menor grau—em qualquer animal, para não falar em máquinas e programas, sendo exactamente esta uma objecção constantemente levantada em relação à simulação de inteligência em máquinas (ela torna-se mais óbvia nas mentes humanas em situações patológicas, como por exemplo situações de mania e de obsessão).

O que temos é então o seguinte: o leque de possibilidades comportamentais de agentes começa, em seres simples, por consistir em rotinas rígidas e chega até à ‘liberdade’. Nas rotinas rígidas já existem *rationales*. O problema que enfrentamos – e que é essencial quer na teoria da consciência quer da liberdade – é conceber a passagem de *free floating rationales* a *full blown reasons* (usadas por um agente para controlar o comportamento próprio). Como já disse, *rationales* para o comportamento de um agente antecedem o uso de razões para controlar o comportamento próprio por agentes. Os *rationales* 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

<sup>404</sup> 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 delimitação 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).

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. Entre ela e nós o que aconteceu não foi a suspensão do determinismo pela irrupção do agente (o aparecimento de ‘máquinas semânticas’) mas o aumento do *elbow room*. Dado o *elbow room* tornou-se possível, a determinados agentes, o controlo num mundo determinista, através da previsão (mentalista) de padrões. Controlo e evitamento, e não liberdade numérica ou indeterminismo, são assim o que realmente queremos, e mesmo em agentes outros que não os humanos constituem o âmbito de uma proto-acção. Um ponto importante para a sofisticação deste controlo em agentes como nós é centralização do controlo por um eu (essencial para ser possível vir a falar de ‘auto-controlo meta-reflexivo num mundo determinista’). Outro ponto é a análise das características da deliberação em agentes como nós (que deliberação é essa, como pode ela dar-se, se não somos vontades numéricas nem eus dados a priori). Introduzo apenas alguns detalhes: a própria ideia de deliberação seria absurda se não existisse um espaço de possibilidades para o agente; o problema é saber de que tipo de possibilidade se trata. Dennett propõe que se trata de possibilidade epistémica (e não lógica, ou nómica). É a possibilidade epistémica –subjectiva, que diz respeito ao que o agente sabe ou não sabe—que importa para a questão da acção e que subjaz ao *design* de um deliberador. 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 incontroláveis e imprevisíveis, estados controláveis e estados previsíveis. É aí que entra o papel das razões: e então como diz Dennett, «we captured reasons (*free floating rationales*) and made them our own (*full blown reasons*)» - as razões usadas por humanos no controlo de acções próprias são razões de agentes para si próprios (recorde-se os critérios de Anscombe). A teoria cognitiva da agência envolve muito mais, mas vou parar por aqui.

#### 4. Conclusão

Que conclusões tirar de tudo isto? Que objecções formular? O que fica dito sobre liberdade? Vou formular algumas conclusões e pôr em relevo alguns pontos de possível contestação.

Pensando em termos de método e temas o que vemos em Dennett é que os temas tradicionais da filosofia, neste caso, a liberdade, são tratados em ligação com a ciência cognitiva. Mas o facto é que teses positivas são avançadas quanto a saber se somos ou não somos livres, o que é a liberdade, e como se relaciona ela com a natureza física e causal do mundo. A estratégia adoptada por Dennett no tratamento do problema da liberdade é idêntica à usada quanto à consciência e é aquilo a que podemos chamar uma estratégia deflacionária - a vontade livre não é *um* problema, mas vários problemas (relativos ao evitamento, aleatoriedade, controlo, ao Eu, deliberação e decisão, às razões na acção, ao 'poder ter feito de outra maneira', às atribuições de culpa), problemas que se interpenetram aliás com os problemas da consciência e da racionalidade. Não se trata de supor a existência de um ingrediente a mais na realidade que seria a vontade livre. Será isto uma passagem do testemunho à teoria da cognição? Bom, o que é facto é que são obtidas teses, respostas a problemas que são problemas filosóficos. A particular teoria do mental defendida por Dennett faz com que a questão da vontade livre seja 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. Ilusões na própria delimitação do problema da vontade livre são assim afastadas. A ilusão central é precisamente o suposto contraste entre a vontade livre e o determinismo; defende-se que a discussão acerca de determinismo ou indeterminismo físicos não tem sequer relação com o conceito de liberdade, porque este é um conceito normativo; e é porque se argumenta que a liberdade é possível num mundo determinista que cabe a uma teoria cognitiva da agência esclarecer a natureza de deliberadores deterministas.

Poderíamos dizer que a abordagem a partir da teoria da mente retira importância à noção ética de pessoa, tratando-a como questão cognitiva (em terceira pessoa). Bom, é indesmentível que a teoria da mente tem reflexos directos na ética. É certo que a noção de condições de personalidade opõe-se desde logo a uma concepção absoluta de pessoa (segundo a qual por exemplo um ser humano é sempre e em qualquer circunstância uma pessoa). A dissociação (ou pelo menos a distinção de conceitos) entre ser humano e pessoa vem de longe (basta recordar como J. Locke acentua o facto de 'pessoa' ser uma categoria *moral*, assegurada por uma continuidade psicológica - que Locke não identificava nem com o corpo nem com a alma). O que acontece aqui é que esta posição é motivada pela teoria da natureza do mental (não se pode falar de pessoa ou liberdade sem termos uma teoria acerca da natureza do mental).

Um último comentário acerca de mente, ética e política, já que perguntei no início o que é que uma perspectiva da filosofia da mente pode

trazer a uma discussão que é usualmente ética, política e metafísica. Quanto à metafísica, já vimos o que ancora a defesa do compatibilismo. Quanto à ética, como ficamos? Se cada um de nós fosse um eu natural e essencialmente uno ou uma pessoa ‘absoluta’, se um conceito essencialista de pessoa fosse defensável, estas questões acerca de agentes e controlo, o auto-controlo meta-reflexivo, controlo de si através de razões pareceriam redundante para aquilo que somos. Mas esse não é, de acordo com Dennett, o caso. Logo, o que Dennett propõe tem reflexos não apenas teóricos mas também pessoais. (Dennett avança uma pretensão forte de resto apenas compreensível levando em conta também a sua teoria da consciência: o auto-controlo meta-reflexivo é uma condição essencial da auto-consciência genuína (*full blown consciousness*), o que traduzido em termos mais filosóficos significa: a auto-consciência está ligada à agência racional). De resto, estas posições de Dennett em filosofia prática, mostram-nos mais uma razão pela qual o eliminativismo em teoria da mente é insustentável: não podemos afastar noções semânticas e mentalistas sem nos eliminarmos a nós próprios.

Termino com uma alusão ao que vai passar-se amanhã (a discussão sobre a ‘face política da liberdade’): segundo Dennett, ao contrário do que as discussões entre compatibilistas e incompatibilistas nos levam a crer, a verdadeira ameaça à liberdade não é metafísica, colocada pelo determinismo, mas social e política – assim, da mesma forma que a liberdade teve que evoluir em agentes individuais (de forma que podemos justificadamente dizer que um humano é mais livre do que um cão ou um gato), ela continua a evoluir, para além dos agentes individuais, em tipos de sociedade e instituições, que, exactamente como os agentes individuais e exactamente pelas mesmas razões, podem ser mais ou menos livres.

### Referências

- Anscombe, G.E.M., 1957, *Intention*, Oxford, Blackwell.
- Danto, A., 1986, Review of Elbow Room, *Mind*, 95.
- Dennett, D., 1986 (1ªed. 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 Dennett1981.

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, *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 all., (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., 2003, *Freedom Evolves*, New York, Allen Lane.

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, *Journal of Philosophy*, 68, 5-20.

Locke, John, 1999 [1690], *Ensaio sobre o Entendimento Humano*, Lisboa, Fundação Calouste Gulbenkian.

Nagel, T., 1979, *Mortal Questions*, Cambridge, CUP.

Nagel, T., 1979a, Moral Luck, in Nagel 1979.

Taylor, C., 1976, Responsibility for self, in Rorty, A. (ed), 1976, *The Identities of Persons*, Berkeley, University of California Press.

Williams, B., 1973, *Problems of the Self*, Cambridge, CUP.





# LINGÜÍSTICA E FILOSOFIA



AS CIÊNCIAS COGNITIVAS E A NATURALIZAÇÃO DO SIMBÓLICO –  
- A MENTE COMPUTACIONAL E A MENTE FENOMENOLÓGICA<sup>405</sup>

*“Não há portanto diferença qualitativa entre o que se pode dizer da mente e dos seus produtos e o que se pode dizer do corpo”*  
Noam Chomsky, *Regras e Representações*

É o linguista N.Chomsky<sup>406</sup> quem pela primeira vez apresenta a Linguística como ciência cognitiva, i.e. como ciência natural dos fenómenos simbólicos enquanto ocorrências na mente-cérebro. Fá-lo ao caracterizar os fenómenos simbólicos como produções de uma gramática formal internamente representada, fundando-se no postulado explícito de uma faculdade biológica de linguagem cuja teoria deve apoiar-se na teoria das linguagens formais.

Esta caracterização não é filosoficamente indiferente, na medida em que Chomsky toma como objecto teórico a mente, que sob os nomes de ‘pensamento’, ‘espírito’ ou ‘alma’ tem sido o tradicional objecto da filosofia, e afirma explicitamente o anti-nominalismo da sua teoria da existência dos fenómenos simbólicos na parte da natureza que é a mente humana. Em suma, falando de regras e representações, Chomsky fala de processos e ocorrências na mente-cérebro sem evitar formular posições ontológicas<sup>407</sup>.

O conjunto das chamadas ciências cognitivas (nomeadamente a psicologia cognitiva, a Inteligência Artificial, as neurociências, a linguística e

<sup>405</sup> Este artigo resume a tese de mestrado apresentada em 1995 na Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa, intitulada “As Ciências Cognitivas e a Naturalização do Simbólico” e orientada por Fernando Gil (École des Hautes Études en Sciences Sociales – Paris e Universidade Nova de Lisboa).

<sup>406</sup> Vou considerar, seguindo H. Gardner, que desde 1957 (data de publicação de *Syntactic Structures*) a história da linguística moderna é a história das ideias de Chomsky e das diversas reacções da comunidade quanto a elas. Cf. Gardner 1985, *The Mind’s New Science*, New York, Basic Books, p.85.

<sup>407</sup> Cf. Noam Chomsky, *Syntactic Structures*, Mouton, The Hague, 1957; *Aspects of the Theory of Syntax*, Cambridge, Mass., The MIT Press, 1965; *Rules and Representations*, New York, Columbia University Press, 1980; *Knowledge of Language, its Nature, Origin and Use*, New York, Praeger, 1986 (tradução portuguesa de Anabela Gonçalves e Ana Teresa Alves, *O Conhecimento da Língua*, Lisboa, Caminho, 1994, p. 25).

a própria filosofia, por explícita pretensão de alguns dos seus representantes<sup>408</sup>) define-se por um ponto comum, fonte da convergência disciplinar: a constituição de uma teoria naturalista da mente tendo por horizonte a simulação dos fenómenos visados. Ao ocuparem este campo as ciências cognitivas herdaram os tradicionais problemas filosóficos das relações mente-corpo, o *mind-body problem*, na terminologia da literatura anglo-saxónica. Este problema permite, ao mesmo tempo, delimitar claramente a convergência disciplinar referida.

Mas o *mind-body problem* é redobrado pelo problema da relação entre a “mente fenomenológica” e a “mente computacional” (na terminologia do linguista Ray Jackendoff<sup>409</sup>): mesmo tendo sido estabelecido, por uma decisão anti-behaviorista, o direito de falar da mente, a mente visada pelas ciências cognitivas, e nomeadamente pela linguística, merece o nome de mente computacional pelo seu parentesco imediato com cálculos e transformações e por excluir do seu âmbito a questão da consciência e dos seus conteúdos.

Pensar que o objecto da linguística é a mente computacional é já uma tese acerca da forma do sentido enquanto forma na mente, e esta tese prescreve uma modelização algorítmica e discreta para a actividade linguística<sup>410</sup>. Ora, a disponibilidade de um modelo pode predispor ao seu uso exclusivo: a pedra de toque das teses da linguística cognitiva na sua oposição à tese cognitivista-simbólica chomskyana (a tese da mente computacional, da facultade de regras e representações simbólicas, da independência da sintaxe em relação à semântica, do carácter generativo da sintaxe e do carácter

<sup>408</sup> Por exemplo os filósofos John Searle, Daniel Dennett, Hubert Dreyfus, Paul e Patricia Churchland. Para uma quadro geral do campo e história das ciências cognitivas, cf. Dupuy, J.-P., *Aux origines des sciences cognitives*, Paris, La Découverte, 1994 ou Andler, Daniel, *Introduction aux sciences cognitives*, Paris, Gallimard, 1992. Para as relações entre a filosofia e as ciências cognitivas, cf. A. Goldman ed., *Readings in Philosophy and Cognitive Science*, Cambridge, MA, The MIT Press, 1993.

<sup>409</sup> Cf. *Semantics and Cognition* (Cambridge, Mass., The MIT Press, 1983), *Consciousness and the Computational Mind* (Cambridge, Mass., The MIT Press, 1987) e *Languages of the Mind* (Cambridge, Mass., The MIT Press, 1992).

<sup>410</sup> A origem desta tese reside na história da lógica, em particular na formulação do conceito de Máquina de Turing (uma Máquina de Turing é um autómato abstracto finito que opera discretamente, com um número finito de estados internos e uma fita ilimitada; uma Máquina de Turing Universal pode simular qualquer Máquina de Turing). Para uma explicação pormenorizada, cf. Penrose, Roger, *A Mente Nova do Rei*, Rio de Janeiro, Editora Campus, 1991, capítulo 2, *Algoritmos e Máquinas de Turing. O conceito foi criado pelo matemático inglês Alan Turing para tentar dar resposta ao problema da decisão formulado por D. Hilbert. A propósito da relação entre a lógica e as ciências cognitivas afirma J. M. Salanskis: “Não se compreenderá correctamente a articulação filosofia/ciências cognitivas e as suas possibilidades actuais de divergência e de evolução se não se pesar um acontecimento epistémico mais geral e determinante em todos os aspectos que é o triunfo disciplinar da lógica neste século (...). A lógica do século vinte é o guia e a referência tanto da filosofia analítica como das ciências cognitivas, e por isso constitui o medium da sua interacção”, (Philosophie et Sciences Cognitives, Rev. Intellectica, nº 17, Paris, 1993/2). A conjugação de resultados lógicos com a vontade de uma solução materialista para o problema da mente-cérebro leva assim à formulação da hipótese fundadora do cognitivismo simbólico: Talvez o cérebro humano seja uma Máquina de Turing Universal, uma máquina simbólica implementando algoritmos. O programa de investigação correspondente visaria o nível causal caracterizado como funcional (ou simbólico) de existência destes algoritmos. É este o lugar teórico da linguística chomskyana.*

interpretativo da semântica) é a presença constitutiva nos estados mentais linguísticos de formas geométricas figurativas, presença ocultada pela modelização algorítmica.

Se exteriormente a linguagem consiste em cadeias de representações formalmente discretas e na possibilidade de gerar ou derivar novas cadeias de símbolos, processos estes modelizáveis pela lógica, ela necessita também de um suporte material, i.e. depende de uma física interna da qual se sabe pouco. Na hipótese chomskyana, a possibilidade dos fenómenos simbólicos supõe a implementação física, neuronal, de um cálculo de representações inacessível à consciência cujos universais formais são inatamente determinados e que é necessário explicitar. A teoria formal das gramáticas que visa as gramáticas internas como um módulo da mente dos falantes tem este fundamento.

O postulado de um ‘módulo’ da mente para a linguagem é a expressão da convicção de Chomsky segundo a qual os processos algorítmicos da linguagem estão funcionalmente isolados em relação a quaisquer outros processos mentais, nomeadamente em relação às crenças. ‘Modular’ significa portanto ‘funcionalmente isolado e claramente delimitável’, nomeadamente em relação ao semantismo dos processos em estudo. Na terminologia cognitivista, um módulo da mente é um sistema cognitivo informacionalmente encapsulado. A legitimidade deste conceito levanta obviamente problemas.

Todos os linguistas cognitivos americanos a seguir referidos (Ronald Langacker, George Lakoff, Ray Jackendoff, Leonard Talmy)<sup>411</sup> se opõem à suposição da relação directa entre o ser discreto das representações simbólicas e a concepção de mente-cérebro como máquina simbólica ou computador abstracto nela implícita assim como à concepção objectivista de significação suposta quer na semântica lógica quer na linguística chomskyana na medida em que esta exclui do seu âmbito a relação entre a subjectividade e o mundo sensível, e por isso o problema do semantismo ou significatividade.

Essa exclusão, enquanto pressuposto, pode ser chamada a clausura dos agenciamentos linguísticos e essa clausura corresponde a uma concepção formalista da forma simbólica, i.e. à utilização da metáfora do sistema formal no sentido matemático técnico para conceber as línguas naturais enquanto actividade cognitiva. À hipótese da clausura, a linguística cognitiva contrapõe a ancoragem das categorizações gramaticalizadas

<sup>411</sup> Seria possível evocar também o trabalho do matemático francês J. Petitot a partir de R. Thom (cf. por exemplo, J. Petitot, *Morphogénèse du Sens*, Paris, PUF, 1985, e J. Petitot, *Physique du Sens*, Paris, Editions du CNRS, 1992) e o trabalho de J.P. Desclés.

existentes nas línguas naturais nas categorizações operadas no mundo espaço-temporal pela percepção.

Se as representações esquemáticas são possibilitadas pela percepção visual ou outra de estados processos e acontecimentos que se desenrolam no referencial espaço-temporal, a linguagem deve, citando Jean-Pierre Desclés<sup>412</sup> ser considerada um processo de *mise en forme* que articula por um lado uma física qualitativa do mundo espaço-temporal externo organizado pela percepção e por outro lado uma lógica simbólica manifestada pelas categorizações discretas e operações formais e universais como a predicação e a anaforização. A forma de linguagem estará portanto na junção de dois universos formais, o lógico e o geométrico, o que explicaria a sua complexidade e potência. A esta hipótese dar-se-á o nome de hipótese localista<sup>413</sup> e ela pretende colmatar o dualismo entre o físico e o simbólico no caso específico dos estados mentais linguísticos admitindo a possibilidade da existência de esquemas cuja forma não é lógica em acção na articulação linguística do sentido. O facto de esses esquemas provirem da percepção não implica uma reintrodução ingénuo da referência no estudo dos fenómenos simbólicos: apenas lança uma dúvida sobre a suposição chomskyana segundo a qual os agenciamentos linguísticos seriam operações formais vazias de significação e a gramática com pretensões ao património genético seria um conjunto de dispositivos algorítmicos que manipulam símbolos sem recurso ao seu significado ou a alguma capacidade cognitiva geral<sup>414</sup>.

A teoria chomskyana da linguagem como órgão mental, órgão-programa e módulo da mente foi a primeira teoria linguística cognitivista e representa, como qualquer programa de investigação, uma decisão, neste caso acerca da mente, ao pretender, como já foi acentuado, que o ser formal das representações linguísticas é algorítmico e como tal analisável e simulável: trata-se então de, ao procurar a descrição de uma gramática, fazer uma *ciência natural de um módulo da mente*. É visada em última análise uma gramática universal que permita chegar a produzir e enumerar automaticamente todas as frases gramaticais de todas as línguas humanas

<sup>412</sup> Desclés, J-P. *Le dualisme géométrique/symbolique dans l'horizon de la forme pure*, in *La Philosophie Transcendentale et le Problème de l'Objectivité*, Paris, Ed. Oisiris, 1991, p.51.

<sup>413</sup> A história desta hipótese remonta aliás aos gramáticos gregos (cf. Jean Petitot 1985) e contemporaneamente foi trabalhada na sintaxe actancial dos linguistas L. Hjemslev, L. Tesnière e Ch. Fillmore para conceber a existência de casos gramaticais. O que está em causa é a natureza formal da conexão sintáctica; a hipótese localista aponta a intuição espacial de relações entre posições como originadora. Em *A categoria dos casos*, 1935, L. Hjemslev apresenta uma interpretação transcendental da teoria dos casos.

<sup>414</sup> Na argumentação insistente do linguista George Lakoff, dissidente da escola chomskyana, as coisas aparecem da seguinte maneira: se uma gramática formal é um aparelho algorítmico cujo output são objectos abstractos e ao mesmo tempo é uma teoria da mente, a pergunta (talvez simplista) a fazer é: onde é que supostamente entra a significação na geração dos objectos abstractos que são as frases? Afirma-se usualmente que a componente semântica das gramáticas é interpretativa como num sistema formal. Mas intuitivamente a ideia de que uma língua natural seja feita de símbolos ininterpretados é muito estranha. Cf. G. Lakoff, *Women, Fire and Dangerous Things - What Categories Tell Us About The Mind*, Chicago University Press, 1987.

atestadas ou possíveis. Como Hilary Putnam faz notar<sup>415</sup>, Chomsky considera provavelmente as línguas naturais como linguagens formais muito complicadas e ainda não totalmente formalizadas. Essa concepção de linguagem natural como sistema formal, a que G.Lakoff chama (acusatoriamente, no sentido de uma imposição que é uma ocultação) uma metáfora, é a chave da linguística chomskyana e baseia-se obviamente na modularidade da gramática em relação ao resto da cognição.

A possibilidade desta ciência natural chomskyana de um módulo da mente concebido à imagem de uma linguagem formal está de acordo com os princípios gerais do cognitivismo simbólico sintetizáveis através de algumas teses do filósofo Jerry Fodor<sup>416</sup>, nomeadamente o representacionalismo, o funcionalismo (a doutrina segundo a qual a taxonomia teórica do psicólogo não tem que parecer natural do ponto de vista de uma ciência de nível mais baixo, nomeadamente da neurociência<sup>417</sup>) e o computacionalismo (a descrição sintáctica da causação simbólica/ mental)<sup>418</sup>. A célebre e contestada hipótese de Fodor de uma *language of thought* compromete-o com uma teoria da mente ou psicologia (entendida como conjunto das questões acerca de estados e processos mentais tendo no seu centro o organismo e as suas atitudes proposicionais) 'hereticamente' mentalista e representacionista além de computacional e funcionalista. O postulado da existência de representações mentais substitui as diversas versões behavioristas da descrição do mental em termos de disposições comportamentais e permite a descrição dos processos da mente-cérebro através da atribuição de fórmulas de um sistema formal a estados de organismos. Segundo Fodor esta é a melhor teoria da mente disponível e a única que evita a prática 'esquizóide' que consiste em fazer uma caracterização computacional dos fenómenos mentais sem reconhecer a necessidade de admitir um *medium* para essas computações. Jerry Fodor tem vindo a manter esta argumentação a favor do mentalismo representacionista desde *The Language of Thought* (1975) até por exemplo *Psychosemantics* (1989) e *A Theory of Content* (1990).

<sup>415</sup> É o sentido de uma sua intervenção registada em M. Piattelli Palmarini [org.], *Teorias da Linguagem, Teorias da Aprendizagem*, São Paulo, Cultrix, 1983 [o livro documenta o debate entre Chomsky e Piaget que ocorreu em Royaumont em 1975].

<sup>416</sup> Cf. Fodor 1993, *Fodor's Guide to Mental Representations*, in Goldman 1993 e também *The Language of Thought* [Harvard University Press, 1975], *The Modularity of Mind* [Cambridge, Mass., The MIT Press, 1983], *Psychosemantics* [Madrid, Tecnos, 1994].

<sup>417</sup> Do ponto de vista contrário, se não existe um nível funcional, toda a ciência comportamental deveria ser neurociência (é a posição de filósofos como os Churchland, cf. Patricia Churchland, *Neurophilosophy*, Cambridge, Mass., The MIT Press, 1986 e Paul Churchland, *A Neurocomputational Perspective*, Cambridge, Mass., The MIT Press, 1989).

<sup>418</sup> Segundo a argumentação de Fodor, devem existir símbolos porque apenas símbolos têm sintaxe e a nossa melhor teoria dos processos mentais (teoria cuja necessidade é ignorada por behaviorismos e associacionismos) necessita da ideia de mente como máquina guiada pela sintaxe. A existência de um computador corporiza a solução para o problema da mediação entre propriedades causais de símbolos e propriedades semânticas: as propriedades causais de um símbolos são conectadas com as suas propriedades semânticas por meio da sintaxe.



Poder-se-ia objectar que a decisão funcionalista chomskyana, na linha da teoria cognitivista da mente, intercepta a questão pelas condições de possibilidade da forma de linguagem no mundo: como faz notar o filósofo John Searle, a sintaxe, condição da linguagem, não existe no mundo físico. Como é então possível que exista na parte do mundo físico que é a mente(-cérebro) humana?

Para Searle, se a caracterização de um processo como computacional é atribuição de sintaxe, e se essa atribuição é relativa ao observador que trata um dado fenómeno físico como sintáctico, isso deve fazer-nos pensar que a tão proclamada realizabilidade múltipla<sup>419</sup> dos processos computacionalmente equivalentes em meios físicos diferentes não é apenas um sinal de que esses processos são abstractos mas de que eles não são de todo intrínsecos ao sistema físico<sup>420</sup>.

Por outro lado, relembrando uma objecção mais antiga de Searle à “Razão Cognitiva”, i.e., à concepção cognitivista do mental como algorítmico, a sintaxe não é suficiente para a semântica: uma caracterização sintáctica, funcional e computacional de estados e processos mentais não captura nem explica o seu semantismo, a sua intencionalidade (a experiência de pensamento do Quarto Chinês, que pretende ser uma refutação do Teste de Turing, tem como finalidade prová-lo)<sup>421</sup>.

Estes são os dois erros fundamentais, segundo Searle, da analogia cognitivista, e na sua opinião eles tornam o cognitivismo num bom trabalho repousando sobre falsas suposições. Do ponto de vista da linguística teórica tais suposições redundam na ocultação dos problema da intencionalidade e da existência de forma qualitativa linguisticamente expressa.

### **1. N. Chomsky: Pressupostos da Gramática Generativa como ciência natural de um módulo da mente**

A teoria linguística chomskyana, hoje ponto de referência inevitável do pensamento sobre a linguagem, é também ela um mentalismo computacional funcionalista, i.e. assenta na admissão de estados da mente-cérebro (sendo a mente o cérebro tomado num certo grau de abstracção, nos termos da tese funcionalista) que sustentam algo de semelhante à dedução,

<sup>419</sup> *Pedra de toque do funcionalismo, ‘realizabilidade múltipla’ significa a possibilidade de implementação de uma mesma máquina simbólica em abstractos diversos, por exemplo biológicos ou artificiais, indiferentemente. Como diz Zenon Pylyshyn, em Computation and Cognition (Cambridge, Mass., The MIT Press, 1984, p.57) “uma sequência computacional pode ser realizada por um grupo de pombos treinados para dar bicadas como uma Máquina de Turing”.*

<sup>420</sup> *Cf. Searle, The Rediscovery of Mind, Cambridge, Mass., The MIT Press, 1992.*

<sup>421</sup> *Cf. Dennett & Hofstadter The Mind’s I (New York, Basic Books, 1981) para o célebre artigo de Searle, “Minds, Brains and Programs” que foi originalmente publicado na revista Behavioral and Brain Sciences em 1980, seguido das respectivas críticas.*

i.e. transformações de *representações* abstractas segundo *regras*. Esses são os processos mentais causalmente explicativos do comportamento regido por regras que é o uso de uma língua natural. As regras e as categorias desses processos constituem conhecimento inconsciente e inacessível à consciência: para este conhecimento tácito ou implícito, subdóxico, Chomsky forja o verbo *cognize*. Podem ser objectos deste verbo os princípios e parâmetros da Gramática Universal<sup>422</sup> e as regras que constituem o estado estável da faculdade de linguagem num indivíduo adulto.

O próprio Chomsky afirma que o seu programa de investigação “foi um dos fios que conduziram ao actual desenvolvimento das ciências cognitivas”<sup>423</sup>, dada a escolha de objecto da teoria linguística. Esse objecto é e sempre foi para Chomsky a mente, e não produtos linguísticos exteriores, palavras e frases consideradas como comportamentos ou a linguagem considerada como constituída por um corpus de enunciados. Trata-se de caracterizar abstractamente, através de uma teoria funcionalista e computacional, estruturas que são propriedade de mecanismos físicos, especificamente do cérebro: trata-se portanto de um estudo da mente por princípio não distinto do estudo do mundo físico. A teoria chomskyana assume-se desde o início como um ataque frontal à plausibilidade do behaviorismo como princípio teórico para o estudo da linguagem, já que este tem como consequência a eliminação do único tipo de análise produtiva da linguagem, a análise feita em termos de sistemas formais do conhecimento acerca da estrutura que torna possível o uso de linguagem. Para Chomsky, admitir que o comportamento linguístico é causado não nos obriga a admitir que este é provocado (*elicited*) por estímulos. A atribuição de estrutura categorial e computacionalmente complexa a sinais físicos não é analisável em termos de estímulo-resposta (é este o conteúdo da polémica acerca da causação do comportamento linguístico que opôs Chomsky a B. F. Skinner no fim dos anos 50).

Esta pretensão chomskyana de uma teoria mentalista na linguística enquanto teoria psicológica, de um mentalismo que é além do mais

<sup>422</sup> Embora tenha vindo a ser diferentemente concebida desde a teoria Padrão até à teoria dos Princípios e Parâmetros, a Gramática Universal nunca deixou de ser o cerne do programa de investigação generativista na medida em que é correlativa da tese dos universais formais da faculdade de linguagem e da existência de um “órgão-mental” do qual deve ser postulado um estado inicial E-0, que é uma componente do espírito humano e que especifica uma certa espécie de gramáticas humanamente acessíveis. Deve entender-se “Gramática” no sentido formal de sistema dedutivo especificando todas e apenas as expressões bem formadas de uma linguagem. Uma Gramática atribui descrições estruturais a frases. O uso do termo gramática é conscientemente ambíguo entre a teoria do linguista e a mente do falante. A metáfora do órgão-mental, que é um órgão-programa, para o LAD (*language acquisition device*) pretende por um lado estabelecer a linguística como ciência natural da mente e por outro estabelecer uma concepção de Conhecimento da Língua como crescimento e não como generalização indutiva a partir da experiência. Também não diríamos que uma ave aprende a ter asas, e a linguagem é para Chomsky um sistema computacional tão complexo que é absurdo supô-la proveniente da interação com o ambiente à maneira de um sistema de reforços que especifica o comportamento de um pombo. Cf. Cf. Chomsky 1980, *Rules and Representations*.

<sup>423</sup> Cf. Chomsky 1994, p. 25.

matematicamente formalizado, é uma tese muito mais forte, e por isso muito mais arriscada, acerca do lugar da mente na natureza do que qualquer nominalismo linguístico ou semiótico. É uma teoria naturalista da forma do sentido, na medida em que considera que a mente releva do mundo biológico e que a psicologia faz por princípio parte das ciências naturais sendo estéril constatar a sua não coincidência no estado actual do inquérito. Nas palavras de Jean Petitot, a teoria chomskyana é a primeira tentativa de uma linguística pura para além da ciência empírica, isto porque implica uma abdução metaempírica acerca da natureza (simbólica) dos estados mentais linguísticos e do mecanismo que os gera (uma gramática), i.e. define o objecto de uma ciência determinando matematicamente uma objectividade<sup>424</sup>.

Mesmo se nas obras mais recentes a terminologia generativista muda e Chomsky fala agora de Língua-I ou interna onde antes falava de competência' e de Língua-E ou externa onde antes falava em 'performance', não muda o mentalismo nem muda o naturalismo: uma gramática generativa será sempre uma teoria que se ocupa do estado da mente-cérebro do indivíduo que conhece uma língua particular e um sistema formal que expõe explicitamente os meios finitos acessíveis à mente dos quais pode ser feito um uso ilimitado (é isso a generatividade). Isto é assim mesmo se a acessibilidade à consciência de regras e representações não é nunca critério do "conhecimento da língua" (CL)<sup>425</sup>.

Na prática da investigação, a Gramática Universal, que é a teoria daquilo que existe no estado inicial E-0, deve permitir um número indefinido de línguas humanas possíveis sem deixar de ser explicativa em relação à rigidez final do sistema de conhecimentos adquirido (para Chomsky não é verdade que as línguas humanas difiram uma das outras imprevisivelmente e sem limites a não ser em aspectos estruturalmente triviais tais como o vocabulário).

### 1.1. Esquema da teoria

Na concepção inicial da Teoria Padrão (1965), existe uma base, componente categorial, que *gera* as estruturas profundas, através de regras de reescrita categoriais, que têm o carácter recursivo necessário para gerar um número infinito de estruturas. Ao objecto formal assim gerado chama-se

<sup>424</sup> Cf Petitot 1985, *Morphogénèse du Sens*, p. 124.

<sup>425</sup> Acerca da língua natural que falamos sabemos sempre mais do que aquilo que pensamos que sabemos (este é o Problema de Platão em linguística, segundo Chomsky). Aquilo que sabemos é um saber-fazer, do qual não nos é possível fornecer um relato explícito e consciente: nada se conclui acerca da estrutura formal da língua interrogando o falante nativo.

*estrutura profunda*, e as estruturas profundas são o *input dos mecanismos transformacionais*. A aplicação das transformações determina finalmente o indicador sintagmático a que se chama *estrutura de superfície* e que corresponde ao uso. Distingue-se a componente generativa (sintáctica) das componentes interpretativa (fonológica, semântica) da Gramática.

Nas conhecida árvores da gramática generativa (que representam os processos computacionais cerebrais-mentais supostos na atribuição de significação a um sinal físico) aparecem diagramatizados estes processos bem como o facto de as frases nas línguas naturais não serem ‘entidades lineares’ mas antes estruturas hierárquicas de componentes e subcomponentes. A hierarquização e categorização (classificação de componentes estruturais) transparecem no conceito de relação de dominância.

Os fenómenos de ambiguidade estrutural (Por exemplo a compreensão da frase inglesa “Flying planes can be dangerous” como ou: 1. Aeroplanos podem ser perigosos ou 2. Pilotar aviões pode ser perigoso) são assim explicados pela possibilidade de atribuição de mais do que uma descrição estrutural, o que é feito inevitavelmente pelo falante nativo. O automatismo de compreensão alternada, por *Gestaltswitch*, evidencia a presença constitutiva da compreensão das estruturas sintácticas que a teoria chomskyana visa.

No modelo teórico mais recente<sup>426</sup>, o modelo de Princípios e Parâmetros, apesar das alterações teóricas, a gramaticalidade continua a ser considerada em termos de processos computacionais dedutivos dos subsistemas ou módulos da Gramática Universal (GU) e da exigência de legitimação de cada elemento estrutural das cadeias simbólicas da linguagem, e portanto em termos de Regras e Representações. Para o propósito da presente análise, continua a ser um modelo cognitivista simbólico

## **2. Epistemologia da linguística chomskyana: o problema de Wittgenstein (acerca de seguir uma regra)**

Chomsky admite o uso do termo vago “regra” para o seu trabalho de caracterização do funcionamento da mente computacional linguística (seguir uma regra é o que faz a identidade entre o uso de lógica e linguagem num humano e um programa numa máquina simbólica) e admite que Regras e Representações devem ser subdoxasticamente conhecidas (*cognized*) para que possamos falar uma língua natural. Ora esta afirmação encontra como limite da sua legitimidade aquela a que o próprio Chomsky

<sup>426</sup> Proposto pela primeira vez por Chomsky em *Lectures on Government and Binding*, 1981, Foris, Dordrecht. Cf. Eduardo Paiva Raposo, *Teoria da Gramática*. A Faculdade da Linguagem (Lisboa, Caminho, 1992).

chama o ‘problema de Wittgenstein’ (para além de constituir de um ponto de vista filosófico behaviorista e holista um Mito do Museu<sup>427</sup>).

Se o conceito de regra é epistemologicamente fundamental para o trabalho de Chomsky, o que está em causa não é o poder de querer, o sentimento de poder dar a si próprio a regra da acção (podemos concordar com Wittgenstein e dizer que seguimos a regra linguística *cegamente*) mas sim o paradoxo céptico que é possível levantar acerca da atribuição do seguimento de uma regra, e que é uma questão central das *Investigações Filosóficas* de Wittgenstein<sup>428</sup>: não há razão suficiente para afirmar que S está a seguir R privadamente (como podemos saber que não se tratava de R’ até aqui, sendo R e R’ até ao caso presente coincidentes nos seus produtos?)<sup>429</sup>.

Há uma nítida incompatibilidade entre a solução publicista wittgensteiniana, que parte da tese “Seguir uma regra é uma praxis”, e o foco da análise de Chomsky, que é o sujeito, no sentido psicológico de indivíduo isolado. Noutras palavras, o enfoque chomskyano seguramente envolve a convicção de que “Privadamente está a ser seguida uma regra”.

A objecção de Wittgenstein ao conceito calculatório de regra é talvez mais profunda do que a interpretação que o cognitivismo possa pensar rebater (se seguir uma regra é obedecer uma ordem, o que é uma ordem? São apenas sons ou caracteres impressos e é o uso que deles é feito que os eleva ao nível de instrução. Não é legítimo transferir para os símbolos um poder intrínseco quando eles não passam de momentos fixados, fetichizados, de uma prática, que é, ela sim, fonte de regras). No entanto Chomsky defende-se dizendo o seguinte:

1. O seu enfoque não envolve qualquer excepção à descrição wittgensteiniana da opacidade das regras ao agente do *rule-governed behavior*: o meu conhecimento de língua enquanto falante é conhecimento não fundamentado, eu não tenho qualquer ‘vivência do porque’ (é aliás para descrever essa ausência de justificação para a própria pessoa que Chomsky usa o verbo *cognize*).

<sup>427</sup> Por implicar que a significação ou o sentido supõem a existência de qualquer coisa de próprio ao indivíduo isolado, i.e. objectos mentais interiores em vez do domínio de um saber fazer específico (este é um objecto de polémica com Quine). Chomsky responde a este tipo de críticas. O adversário do (suposto) Mito do Museu vê-se comprometido com as seguintes afirmações: 1. um hipotético futuro conhecimento da física nada nos diria acerca de como o cérebro do indivíduo está estruturado de origem, 2. nada se passa no cérebro quando há uso de linguagem. Talvez o adversário não esteja disposto a sustentar tais teses.

<sup>428</sup> Cf. Wittgenstein, *Investigações Filosóficas*, Fundação Calouste Gulbenkian, 1987 e S. Kripke, *Wittgenstein on Rules and Private Language*, Cambridge MA, Harvard University Press, 1982.

<sup>429</sup> Os exemplos de Wittgenstein são: continuar uma série numérica (por exemplo 3, 6, 9...) ou ler símbolos alinhados escritos na minha língua. O problema é: poderei dizer que seguir a regra é um estado de consciência? Que há uma vivência de ser guiado? Que me ocorrem fórmulas? Serei capaz de indicar a regra que estou a seguir? Não, isto não acontece; na expressão de Wittgenstein, eu não tenho uma ‘vivência do porque’ que acompanhe a observância da regra. Não existe nenhum facto acerca de mim que [me] diga se estou a seguir R ou R’.

2. É um problema empírico difícil decidir se uma regra está ou não a ser seguida. Se se aborda esse problema cientificamente procurando a verdade acerca da faculdade de linguagem, não será a solução comunitária wittgensteiniana a auxiliar a sua resolução ou progresso: antes, ela nada acrescenta aos “factos acerca de Jones-o-falante-qualquer”. Isto é assim em contraste com a investigação em termos de regras e representações

Para Quine, outro filósofo interveniente nesta discussão, o problema com o enfoque de Chomsky é outro: Chomsky afirma demais ao supôr a codificação física das questões de significação. O linguista deveria abster-se de atribuir realidade psicológica ao objecto da descrição formal: o que existe é apenas *adequação de comportamentos* a um sistema de regras que o linguista hipoteticamente postula. Em contrapartida, para John Searle, Chomsky afirma de menos: tem que haver razão para supôr que as regras estão a funcionar causalmente e a teoria de Chomsky não nos fornece provas suficientes.

### 3. Ray Jackendoff: justificação de uma Estrutura Conceptual morfológica

Se John Searle pretende provar que a análise cognitivista da mente passa ao lado do problema da intencionalidade, que é, ele sim, o verdadeiro centro do problema do lugar da mente na natureza, e também que um computacionalismo não pode ser naturalista já que a sintaxe não é uma propriedade física, o linguista R. Jackendoff, em “A consciência e a mente computacional” (*Consciousness and the Computational Mind*, 1987) pretende mostrar que essa análise não diz nada acerca dos “ocupantes da mente fenomenológica”, quer dizer, não se pronuncia acerca da natureza da experiência consciente na medida em que esta não coincide com as regras e representações supostas na mente computacional. A experiência consciente (que é causada por ou paralela a, ou idêntica a impulsos nervosos) fica inexplicada: nada é dito acerca da *aboutness* não apenas proposicional, mas perceptiva e imaginária ou dos *qualia*.

É tão obscuro pensar que se fala da experiência consciente ao explicar agenciamentos formais de regras sobre representações como pensar que esta é explicada falando directamente de activação neuronal. Em nenhum dos casos se concebe a possibilidade de experiência consciente. É preciso saber como é que dos agenciamentos formais podem resultar experiências, *conscious awareness*. Esta não é simplesmente a capacidade de auto-referência mas o mundo externalizado das formas e dos *qualia*, e daquilo que é nomeado pela linguagem.

Para Jackendoff, falar, como faz Searle, da intencionalidade como característica dos estados mentais pela qual eles são dirigidos a ou são acerca

de objectos e estados de coisas no mundo é um posicionamento errado do problema: se a intencionalidade fosse o problema da relação entre estados mentais e mundo real não haveria lugar para a *aboutness* de estados mentais tais como alucinações ou imaginações (ou para estados mentais linguísticos, que certamente não intencionam o mundo tal como este realmente é, se por tal se entende o mundo tal como a física o descreve). A intencionalidade deve ter uma formulação mais inclusiva, deve ser a propriedade de estados mentais computacionais pela qual eles são religados ao mundo tal como este é experienciado, “real” ou não (esta formulação inclusiva da intencionalidade foi objecto de rigorosa tematização por Husserl, muito antes das teorias cognitivas que agora a reencontram<sup>430</sup>).

Na hipótese de Jackendoff a mente computacional seria uma especificação abstracta da organização funcional do sistema nervoso, mesmo se de momento não se sabe traduzi-los exactamente em termos de *hardware* neuronal, enquanto os elementos da mente fenomenológica seriam causados por (e projectados pela) informação e pelos processos da mente computacional que são (1) activos e (2) têm outras propriedades privilegiadas.

Mas porque não se pode falar de informação como se de líquido abstracto se tratasse, aquilo que é, mais especificamente, preciso explicar, é por exemplo *como podemos falar daquilo que vemos*. Essa é a raiz da proposta central de Jackendoff enquanto linguista, a proposta da existência de uma Estrutura Conceptual a que podemos chamar morfológica<sup>431</sup>. A chamada “informação” que entra num sistema cognitivo físico é recebida sob diferentes formas físicas (pense-se na diferença entre o arranjo espacial dos receptores retinianos e os detectores de frequência de som). Nenhum desses processos físicos explica por si, por exemplo, o entendimento do mundo em termos de objectos tridimensionais subsistentes e idênticos fora de nós. Teorias de módulos da mente e seus níveis devem poder ser conjugadas, de forma a ser teorizada a passagem da transdução modular para a possibilidade de processos não modulares (a não ser assim, teríamos uma desunidade da apercepção consciente). Por isso, para podermos falar daquilo que vemos, tem que existir uma *lingua franca* na mente, a mente deve poder transformar as formas de estruturas de informação diversas umas nas outras.

Jackendoff supõe que o mundo fenomenológico morfológicamente estruturado, que é o dos modelos mentais que são as línguas naturais, que

<sup>430</sup> Cf. E. Husserl, *Recherches Logiques* (Paris, PUF, 1972) e *Idées directrices pour une phénoménologie* (Ideen I), (Paris, Gallimard, 1950).

<sup>431</sup> A partir das teorias da estrutura de N. Chomsky e D. Marr. Cf. D. Marr, *Vision* (San Francisco, Freeman, 1982). Estas são teorias da estrutura para Jackendoff precisamente porque especificam, em termos de níveis de representação com primitivos e princípios de combinação, as propriedades que devem existir na informação mental para que possamos respectivamente falar e compreender linguagem e ver.

é intencionado nas nossas enunciações, mostra, ou só pode ser explicado por, uma Estrutura Conceptual para o significado, análoga às que são consensualmente consideradas pelos linguistas para a sintaxe e para a fonologia. O nível de representação mental a que Jackendoff chama Estrutura Conceptual estaria em correlação com o modelo 3D da teoria computacional da visão de David Marr. É através desta conjugação que Jackendoff pensa o enraizamento da intencionalidade das línguas naturais nas cenas perceptivas e propõe uma semântica cognitiva que pretende contrastar com uma semântica formal.

As enunciações devem poder receber descrições sistemáticas não apenas em termos fonológicos e sintáticos mas também nos termos de um nível independente, a Estrutura Conceptual (E.C), cujos primitivos e princípios de combinação Jackendoff especifica. Isto significa que as relações linguagem-realidade devem ser explicadas por teorias psicológicas de modelos computacionais que os falantes têm do mundo externo, resultantes da interação entre inputs e meios de representação disponíveis e não pela aplicação de cadeias formais em construtos conjuntistas ou no (suposto simplesmente dado) mundo real. Esta seria a única abordagem compatível com a linguística generativa (o fundo teórico do trabalho de Jackendoff) cuja pretensão fundamental é que a linguagem está mentalmente representada e não existe “fora”, no mundo. Seria interessante a comparação desta tese com a de Searle, da inexistência de sintaxe no mundo físico, já que o cérebro é certamente parte do mundo físico.

Chamando pressuposições ontológicas de uma Língua L aos tipos de entidades que, no modelo mental que é a língua, habitam o mundo, ter-se-á a prova de que para um falante de L uma entidade existe no mundo, faz parte das pressuposições ontológicas de L, quando o falante se lhe refere por anáforas pragmáticas<sup>432</sup>, como “Aquilo (apontando) é um gato”, ou através de expressões de identidade e individuação como “o mesmo”, “o outro”, etc.

As anáforas pragmáticas não são apenas possíveis com OBJECTOS como no caso acima, mas, segundo Jackendoff, com entidades usualmente não participantes em teorias semânticas tais como TRAJECTOS, EVENTOS, LUGARES, etc. Jackendoff introduz então categorias (que não estão necessariamente em correspondência unívoca com a estrutura sintática) no nível de representação a que chama E.C. São traços de categoria ontológica: (OBJECTO), (LUGAR), (TRAJECTO), (ACÇÃO), (EVENTO), (ESTADO), (QUANTIDADE, PORÇÃO), (NÚMERO), para além da distinção *type/token* que possibilita a categorização.

<sup>432</sup> Pronome demonstrativo a que se acrescenta o gesto, a referência provindo de um contexto não linguístico.



Vejamos como serve de prova da organização dos primitivos conceptuais em E.C. o comportamento gramatical das três classes de verbos ingleses para posição espacial: os verbos GO, verbos BE, verbos STAY.

Nas frases:

1. The dog run from the door to the table
2. A meteor hurtled toward the earth
3. The hawk flew over the prairie

há sempre um objecto que se move, a que Jackendoff chama TEMA (X) ao longo de um TRAJECTO (T), que pode ou não ser ainda especificado por uma FONTE ou um ALVO. Para Jackendoff a similaridade semântica destas três frases resulta de serem instanciações de (GO(X,T)), que representa o movimento de um objecto X, o Tema, ao longo de um trajecto T. (GO(X,T)) pertence à categoria ontológica EVENTO em E.C, pois acontece no tempo, enquanto as variáveis X e T pertencem às categorias (OBJECTO) e (TRAJECTO).

Mas as frases citadas são compósitas, na medida em que em cada uma existe mais do que uma Entidade de Referência além da Função-Trajecto, nomeadamente porque as preposições *from*, *to*, *toward* e *over* designam respectivamente:

1. *From*(de): um Trajecto que se inicia na Entidade de Referência
  2. *To* (para): um Trajecto que termina na Entidade de Referência
  3. *Toward* (em direcção a): um Trajecto que se fosse continuado terminaria na Entidade de Referência
  - 4 *Over* (sobre): um Trajecto que atravessa um ponto ou região situada verticalmente sobre a Entidade de Referência
- A forma geral dos Conceitos de Trajecto é então:

(trajecto FUNÇÃO-TRAJECTO(objecto Y))

Combinando as duas formas gerais apontadas, (evento) e (trajecto), obtem-se a forma geral em E.C das 3 frases que serviram de exemplo:

Evento GO ((objecto x), trajecto (FUNÇÃO-TRAJECTO (objecto Y)))

Os verbos BE<sup>433</sup> não descrevem movimento mas locação e têm a forma:

Estado BE ((objecto x) Lugar (FUNÇÃO-LUGAR (objecto y)))

<sup>433</sup> Por exemplo, "Max was in Africa" ou "The statue stands in the woods".

Os verbos STAY<sup>434</sup> têm a forma  
(Evento STAY ((Objecto x) Lugar (FUNÇÃO-LUGAR (Objecto y)))

Feita esta distinção entre as três classes de verbos para posição espacial, cujo comportamento perante juízos de gramaticalidade é aliás, de acordo com Jackendoff, determinada por esta forma em E.C., comprova-se a existência destes traços de categoria ontológica e sua combinação pela possibilidade de generalização através de campos semânticos. Verifica-se que os verbos de posse e os verbos de predicação se deixam classificar exactamente nas mesmas três classes, sendo portanto considerados como instâncias dos conceitos em E.C. descritos. Por exemplo:

Mudanças de posse são instanciações de GO:  
Harry gave the book to Betty  
Charlie bought the lamp from Max

Predicação e adscrição de propriedades são instanciações de GO, BE, STAY:

- a. The servant changed from a young man into a pumpkin
- b. The food was tasty
- c. The man remained a pumpkin

Em conclusão, se os três conceitos GO, BE STAY, na verificação minuciosa e empírica que Jackendoff faz, se aplicam a campos semânticos sem qualquer relação a priori entre si (i.e. nas várias línguas naturais verbos particulares funcionam em mais do que um campo semântico conservando a sua classificação como instâncias de GO, BE ou STAY) e se não existe nada de intrínseco no mundo que requeira que posse e a adscrição de propriedades estejam mentalmente representados num sistema paralelo à representação de Eventos e Estados espaciais, isto só pode significar que existem constrangimentos à conceptualização, i.e. à forma e aos primitivos de E.C., o nível de representação do significado.

#### **4. Ronald Langacker: Fundações de uma Gramática Cognitiva Icónica**

No seio da linguística as Gramáticas Cognitivas aparecem essencialmente como alternativa aos pressupostos de raiz chomskyana, que implicam uma concepção da natureza de estados e processos mentais linguísticos

<sup>434</sup> Por exemplo "Stanley remained in Africa".

a que temos vindo a chamar cognitivismo simbólico, postulando a sua natureza discreta e computacional, a generatividade sintáctica e o semantismo como atribuição (*assignment*).

A posição comum aos linguistas cognitivos é a rejeição de certas suposições quanto à natureza da significação e da gramaticalidade de tal modo aceites na linguística transformacional que se tornaram imperceptíveis: “A gramática cognitiva está basicamente em conflito com as tendências dominantes na teoria linguística actual. Fala de dimensão imagética convencional (*imagery*) num tempo em que a significação é generalizadamente tratada com um aparato derivado da lógica formal”<sup>435</sup>. A noção de *imagery* pretende substituir, portanto, a noção de cálculo ininterpretado. Para Langacker “não é impossível que uma descrição formal seja contraproducente ao dar uma visão falsa ou drasticamente distorcida do seu assunto. (...) Uma particular aparelhagem matemática, por exemplo a semântica verofuncional da lógica formal, pode ser utilizada simplesmente porque está disponível, não porque os factos de linguagem peçam esta abordagem enquanto oposta a outras”<sup>436</sup>.

Com o termo *imagery* (dimensão imagética convencional), Langacker não se refere imediatamente à linguagem figurativa ou metafórica nem a imagens sensoriais no sentido da ocorrência de perceptos na ausência de um input actual, mas à possibilidade, (linguística) de estruturar alternativamente a cena pensada ou enunciada. Toda a expressão linguística diz respeito a situações concebidas ou cenas; isto implica que numa Gramática Cognitiva não seja admitido o pressuposto da separabilidade da sintaxe em relação à semântica, i.e. o assemantismo da sintaxe, e assim uma diferença de natureza entre sintaxe e léxico. Toda a gramaticalidade é de natureza simbólica, consistindo na “simbolização convencional da estrutura semântica”<sup>437</sup>.

Neste fundo teórico, se a figuração, a iconicidade, o esquematismo são omnipresentes (*pervasive*) na linguagem, mesmo na estrutura sintáctica, os fenómenos figurativos e metafóricos esporádicos não podem ser considerados periféricos mas antes centrais e previsíveis.

O que os linguistas cognitivos entendem por semântica não é articulado em termos de teoria dos conjuntos, de verdade e de referência, mas

<sup>435</sup> R. Langacker, *Foundations of Cognitive Grammar* [Stanford, Stanford University Press, 1987], p. 1.

<sup>436</sup> Langacker 1987, p. 45. A teoria chomskyana é, segundo Langacker, naturalmente afim das teorias semânticas verofuncionais.

<sup>437</sup> Langacker 1987, p. 2. A estrutura semântica é definida como estrutura conceptual que funciona como pólo semântico de uma expressão linguística, sendo a estrutura conceptual a experiência mental em geral: conceitos linguisticamente acedidos, percepções visuais, imagens, etc...

sim considerado como *síntese subjectiva*, conceptualização, operação cognitiva moldada às especificações da convenção linguística, e, portanto, o que o linguista faz é caracterizar os eventos cognitivos correspondentes a uma ocorrência linguística dada enquanto experiência mental, i.e., procura fazer uma fenomenologia dos estados mentais linguísticos dentro do seu propósito de ciência empírica.

Por isso a Gramática Espacial ou Cognitiva é psicologia especulativa, e os diagramas utilizados são tentativas de restituição da conceptualização, i.e. da síntese subjectiva. Mostram graficamente, por exemplo, relações temporais ou atemporais entre entidades em domínios, considerando ser esse o objecto conveniente da descrição da experiência mental linguística. São diagramas informais, mas por serem esquemas icónicos são considerados mais adequados à experiência mental linguística do que cadeias de fórmulas lógicas ou árvores categoriais.

A descrição é feita em termos de *units*, unidades simbólicas descritas no diagrama, estruturas completamente dominadas, que o falante pode activar como um todo pré-formado, sem necessidade de consideração da composição interna nem atenção consciente ao uso<sup>438</sup>. Repare-se que a mudança no vocabulário teórico pretende ser um afastamento da noção lógica de cálculo e uma aproximação da noção neurofisiológica de activação. Qualquer pensamento ou percepção é a ocorrência de um evento neurológico, em última análise electroquímico, e dizer que alguém formou um conceito é dizer que um padrão de actividade neurológica está estabelecido e pode ser repetido com facilidade<sup>439</sup>. Eventos são para Langacker ocorrências cognitivas de qualquer grau de complexidade, desde o disparar de um neurónio até aos acontecimentos de larga escala da gramaticalidade.

Segundo Langacker existe uma habilidade cognitiva fundamental inerente ao fluxo de eventos que constitui a experiência mental, a que chama *scanning*. O *scanning* é uma operação unidireccional de comparação, conexão e detecção de contrastes e descontinuidades qualitativas entre eventos, omnipresente na experiência cognitiva em qualquer domínio<sup>440</sup>.

As unidades simbólicas (*units*) são caracterizáveis por recorte em domínios, que são definidos como o espaço representacional necessário à

<sup>438</sup> Embora unidade signifique estatuto de unidade: por exemplo os sons básicos de uma língua são *units* para o falante nativo mas não para o não-nativo.

<sup>439</sup> Cf. Langacker 1987, p.100.

<sup>440</sup> Curiosamente, Langacker exemplifica a noção de *scanning* com o exemplo de Husserl nas "Lições sobre a Fenomenologia da Consciência Interna do Tempo" (1905): a audição. A noção de *scanning* representa, afinal, a introdução do tempo na conceptualização ou síntese subjectiva.

emergência da conceptualização<sup>441</sup>, e os domínios apoiam-se sobre domínios de base (como espaço, tempo, qualidades sensíveis). Postular domínios e unidades simbólicas supõe a inexistência de primitivos semânticos que possibilitariam decomposições analíticas, e a existência de hierarquias de domínios e domínios básicos.

#### 4.1. O esquematismo das Unidades Simbólicas básicas

As unidades simbólicas básicas (os nomes ou predicções nominais) e os verbos e preposições (predicções relacionais, temporais (processos) ou atemporais) são para Langacker definíveis na medida em que envolvem a conceptualização de tempo, espaço e movimento. Esta definibilidade ergue-se contra o dogma linguístico de acordo com o qual nada haveria a dizer do ser-nome ou ser-verbo, que seriam primitivos não especificáveis.

Langacker define o nome ou nominalizável como envolvendo a conceptualização de uma COISA, definida como entidade (= região delimitada) nalgum domínio. Esta definição não visa objectos físicos mas eventos cognitivos, ocorrência de rotinas caracterizáveis independentemente de estimulação periférica. A delimitação de uma região pode dar-se em qualquer domínio<sup>442</sup>.

É possível objectar que Langacker evoca e não explica as propriedades gestaltistas da visão para definir as unidades simbólicas básicas: são primitivos da teoria noções como Perfil e Base, definições supostamente não formais, relações de Identidade, Separação, Vizinhança e Inclusão, e a assimetria, característica das predicções relacionais, entre Trajector e Entidade de Referência (*landmark*), que supõe a elevação de um participante numa organização figura/fundo ao estatuto de figura (o Trajector).

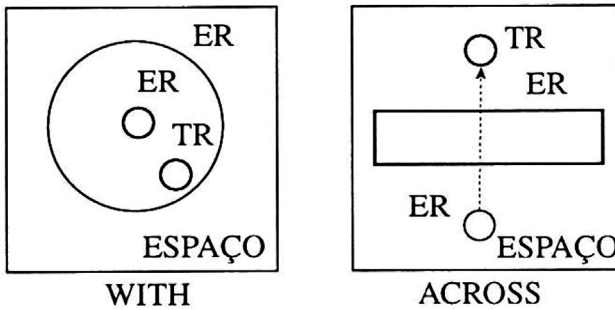
Um perfil relacional, por exemplo uma preposição ou um verbo, envolve portanto uma assimetria entre Trajector e Entidade de Referência, abaixo diagramatizada para WITH e ACROSS.

Os perfis relacionais podem ser distinguidos em termos de *scanning* pela sua temporalidade ou atemporalidade<sup>443</sup>, e a não coincidência do tempo da conceptualização com o tempo concebido é crucial nas análise de Langacker.

<sup>441</sup> Por exemplo o domínio espaço bidimensional é necessário para a conceptualização [CÍRCULO], ou o domínio braço para a conceptualização [MÃO].

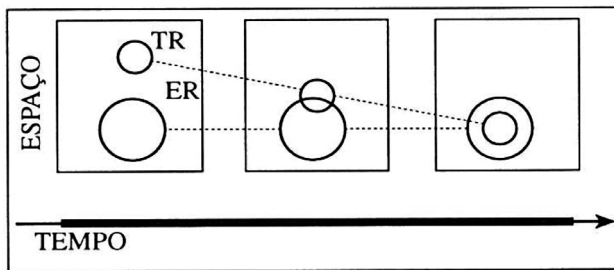
<sup>442</sup> Por exemplo, um termo como RED usado como nome perfila uma região delimitada no espaço da cor, nomes contáveis como SPOT ou STREAK ou STRIPE perfilam regiões de expansão delimitada no campo visual, BEEP perfila uma região delimitada no domínio som e no domínio tempo, BLIP e FLASH ocupam ambos os domínios visual e temporal, mas enquanto BLIP é delimitado no domínio campo visual, FLASH pode abarcar o campo visual, devendo ser estritamente delimitado no domínio tempo.

<sup>443</sup> A atemporalidade de um perfil relacional não supõe que o tempo não seja o seu domínio primário: por exemplo BEFORE não é uma predicção processual e no entanto o seu domínio primário é o tempo.



Um verbo perfila não apenas uma mas uma série de configurações relacionais envolvendo um perfil temporal positivo, e portanto “*scanning* sequencial”.

Por exemplo:



O problema para Langacker neste ponto é visar o que faz a diferença entre uma relação atemporal complexa (por exemplo, a preposição ACROSS) e uma predicação processual (por exemplo, o verbo to CROSS); a noção de *scanning* sequencial serve para explicitar essa diferença. O *scanning* sequencial caracteriza a ocorrência de uma predicação processual por oposição ao *scanning* sumário das predicações relacionais complexas atemporais.

Um processo é seguido ao longo do tempo concebido, enquanto a predicação relacional complexa não é seguida ao longo do tempo, embora obviamente a sua ocorrência requeira a passagem do tempo de conceptualização. O que faz a distinção é então para Langacker a diferença entre o modo de processamento empregue quando vemos um filme ou observamos uma bola que voa pelo ar (os estados sucessivos do evento concebido são activados serialmente, e a activação de um estado começa a declinar quando a do seu sucessor é iniciada: isto é *scanning* sequencial, seguimos o desenrolar do evento) ou quando reconstituímos mentalmente a trajectória

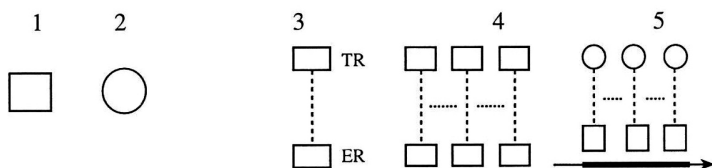
da bola no ar (os estados componentes são activados cumulativamente, mesmo como um todo: isto é *scanning* sumário).

A diferença entre as categorias de predicções referidas não diz respeito a nenhum conteúdo intrínseco mas ao modo de *scanning* empregue na sua activação, i.e. não diz respeito ao tempo concebido, mas ao tempo de processamento.

Mesmo uma predicção como SER, que já para Heidegger contem o elemento de significatividade sobre o qual toda a significação repousa<sup>444</sup>, é caracterizada nestes termos por Langacker: corresponderia ao encadeamento-consigo-no-tempo de uma relação pondo em jogo um Trajector/ Coisa e uma Entidade de Referência/ (coisa ou relação), conceptualizado segundo *scanning* sequencial...

Neste quadro, Langacker faz o seguinte inventário das categorias básicas da gramaticalidade disponíveis:

1. Entidade: coisa ou relação
2. Coisa: região em algum domínio
3. Relação atemporal simples (perfila interconexões entre duas ou mais entidades concebidas, que podem ser coisas ou relações)
4. Relação atemporal complexa (sequência de relações estáticas seguida por *scanning* sumário)
5. Processo: perfila configurações relacionais, mas
  - a. Os estados componentes são concebidos como distribuídos no tempo
  - b. São submetidos a *scanning* sequencial
  - c. O Trajector é sempre uma coisa e nunca uma relação



As considerações de Langacker acerca do carácter imagético e temporalizado das unidades simbólicas disponíveis para o falante levantam a hipótese de os juízos de gramaticalidade no sentido chomskyano clássico,

<sup>444</sup> Cf. M. Heidegger, *Qu'appelle t'on penser?* (Paris, PUF, 1959). Cf. ainda J. M. Salanskis, *Différence Ontologique et Cognition*, in *Intellectica*, n°17, Paris, 1993/2, para uma comparação da teoria da significação de Langacker com os factos de linguagem, em particular a nominalização, em torno dos quais se fazem as reflexões heideggerianas acerca do sentido do ser. Langacker explica na sua teoria as características do sistema do verbo que são os participios e nominalizações e o aspecto em termos de certas percepções do espaço e do tempo gramaticalizadas como dimensão imagética.

que são uma pedra de toque da teoria chomskyana da gramática, e que são normalmente pensados como sendo análogos à aceitação ou não aceitação de uma cadeia simbólica como expressão bem formada, poderem proceder antes da estruturação imagética e temporal da cena concebida e dos desvios a esta.

### 5. Cognição e metáforas

A centralidade da metáfora na nossa compreensão da experiência é a tese central de *Metaphors We Live By* (1980), um estudo do linguista George Lakoff e do filósofo Mark Johnson dentro do espírito da linguística cognitiva<sup>445</sup>. A capacidade de metaforizar é, dizem os autores, como um sentido, como ver ou tocar.

Os autores não falam da metáfora como fenómeno especificamente estético, nem procuram fazer uma ontologia do poético como por exemplo Paul Ricoeur na *Metáfora Viva* (1975). O objecto visado é antes a lexicalização do metafórico nas línguas naturais, os mapeamentos metafóricos que têm como fonte o corpo próprio, a sua experiência sensorial e posição no espaço e que são omnipresentes (*pervasive*) em todas as línguas.

O estudo foi feito para o inglês, língua-objecto da demonstração da tese dos autores. Os seus exemplos serão aqui retomados, sem que se considere necessária uma tradução passo a passo. Deve manter-se em mente a definição de conceptualização ou síntese subjectiva dada por Langacker para compreender o propósito do estudo. Evidentemente do ponto de vista da tradição linguística estruturalista europeia o que vai ser dito implica uma indistinção entre conceito e significado, mas essa é uma decisão da teoria.

A teoria chomskyana da linguagem tem ela própria dificuldades em lidar com a metáfora, o que é previsível dada a sua (metafórica) inspiração formalista, e por isso é frequentemente acompanhada por estratégias que podem ser consideradas *ad hoc*, como as explicações da metáfora pela Abstracção ou pela Homonímia. Se algures se produz uma metáfora, para a primeira das estratégias haveria um conceito geral neutro entre uso metafórico e não metafórico e para a segunda haveria dois conceitos independentes. Estas estratégias ocultam, segundo Lakoff e Johnson, a sistematicidade e generatividade de esquemas de imagem e a sua função cognitiva.

Como não estamos conscientes da Estrutura Conceptual da língua natural que falamos há que procurar as metáforas lexicalizadas em expressões correntes. Por exemplo dir-se-á que existe no inglês a metáfora

<sup>445</sup> G. Lakoff & M. Johnson 1980, *Metaphors We Live By*, Chicago, University of Chicago Press, 1980.



“Argument is war” documentando o uso de expressões como “Your claims are INDEFENSIBLE”, “He ATTACKED my argument”, “He SHOT DOWN my argument”. Nestas expressões não se trata, segundo, os autores apenas de como falamos mas do que pensamos que fazemos: é esse o sentido de “Metaphors we live by”, metáforas pelas quais vivemos. Nós pensamos (e em português dá-se uma metaforização semelhante) que ganhamos e perdemos discussões, atacamos e defendemos posições, ganhamos e perdemos terreno na discussão – não pensamos que estamos a metaforizar quando o fazemos.

A metáfora é definida neste estudo como compreender ou explicar um género de coisa nos termos de uma outra, *highlighting and hiding*, iluminar e ocultar – uma definição que não difere muito das inventariadas na história secular do pensamento da metáfora.<sup>446</sup> Provar a sistematicidade implica que a expressão não seja pontual ou excepcional, por exemplo um fenómeno poético num contexto de arte, mas convencional e como tal não apercebida como metafórica pelo falante. O plano de análise é estritamente sincrónico, não se procura a profundidade oculta da significação original na história de uma palavra ou expressão. Por outro lado, os exemplos não são daquilo a que os retóricos chamam catacreses, como “a perna da mesa”.

O foco da investigação é a Estrutura Conceptual, a reunião de provas tem como propósito defender que a mente não é um processador algorítmico de símbolos ininterpretados, mas que a significatividade é incorporada.

A fonte da maioria dos mapeamentos metafóricos reunidos pelos autores é a experiência do espaço (ou pelo menos são aqueles que sustentam um caso mais forte a favor das suas teses), são metáforas a que chamam orientacionais ou ontológicas a partir de pares como dentro/fora, cima/baixo, frente/trás, central/periférico, perto/longe

Um exemplo de metaforização ontológica será o recolhido em expressões correntes acerca da linguagem. Segundo os autores 70% das expressões utilizadas para falar em inglês sobre a linguagem são instanciações desta metáfora. Expressões como “to PUT INTO words”, “to PACK INTO

<sup>446</sup> Desde o lugar clássico, a *Poética* de Aristóteles, em que se encontram nomeados o primeiro problema sempre repetido do pensamento sobre a metáfora [o facto de a palavra meta-phora ser uma metáfora, um pedido de empréstimo à ordem do movimento] e o enigma da sua produção: é preciso to homoion theorein, ver as semelhanças, produzir o género daquilo que é dissemelhante, e essa é uma enigmática capacidade para Aristóteles, um sinal de dom natural. Cf. Aristóteles, *Poética*, trad. port. de Eudoro de Sousa, Lisboa INCM, 1992. Para a definição “A metáfora é o transporte para uma coisa de um nome que designa uma outra, transporte do género para a espécie, ou de uma espécie para o género, ou da espécie para a espécie ou segundo a relação de analogia”, cf. 1457b. Das quatro espécies de metáfora as que mais se fazem apreciar são as metáforas por analogia, sendo a analogia definida em 1457b: “Digo que há analogia quando um segundo termo está para o primeiro na igual relação em que está o quarto para o terceiro” e exemplificada por exemplo na *Retórica*, 1411a: “Foi assim que Péricles disse: Os jovens que morreram na guerra deixaram um vazio tão sensível na cidade como se do ano se arrancasse a Primavera”. Para uma discussão do problema filosófico da metáfora, cf. J. Derrida, J., *A Mitologia Branca*, in *Margens da Filosofia*, Porto, Rés, s/d.

words”, “To STUFF INTO words”, “words that CARRY”, “HOLLOW words”, “to get an idea ACROSS to him”, “to GIVE an idea” mostram que a linguagem é conceptualizada metaforicamente como envolvendo:

1. Ideias ou significados como OBJECTOS;
2. Expressões linguísticas como CONTINENTES de significados;
3. Comunicar como MANDAR, ENVIAR.

A existência de metáforas orientacionais na conceptualização implica a organização de vários conceitos que podem ser considerados como sendo semanticamente independentes a partir de relações espaciais básicas. Uma metáfora orientacional dá a um conceito uma orientação espacial, devido, segundo os autores, a termos os corpos que temos, uma vez que só deles pode provir a orientação do espaço, que é a orientação do corpo no espaço. Esta metaforização não é arbitrária, na medida em que tem a sua condição de possibilidade na experiência física de um corpo que se apercebe. As oposições acima referidas parecem ser aliás uma base universal de metaforização nas línguas naturais, embora as metáforas assim baseadas (i.e., que conceitos estão espacialmente/metaforicamente orientados de que modo) variem de língua para língua. A espacialização é frequentemente tão essencial à significação da palavra ou expressão que é difícil ou impossível imaginar uma metáfora alternativa.

Tomando o exemplo da oposição UP/DOWN, cima/baixo, temos as expressões:

1. Happy is UP/Sad is DOWN (feliz é em cima, triste é em baixo)  
(To FEEL UP, to BOOST my spirits, my spirits ROSE, to be in HIGH spirits, to FEEL DOWN, to be real LOW)(Em português, com o mesmo sentido, temos por exemplo, “estar em baixo”, ou “tudo em cima”)
2. Conscious is UP/Unconscious is DOWN (consciente é em cima, inconsciente é em baixo)  
(to GET UP, I’m UP, he RISES, to FALL asleep, to DROP off to sleep, he SANK) (falamos das “profundezas da mente, do espírito”)
3. Health and life are UP/Sickness and Death are DOWN (saúde e vida são em cima, doença e morte são em baixo)  
(the PEAK of health, Lazarus ROSE from the dead, to be in TOP shape, to FALL ill, to COME DOWN with the flu, to DROP dead)
4. Control/Force is UP/ Subject to control/Force is DOWN (força/controlo é em cima, ser sujeito a força/controlo é em baixo)

(to have control OVER someone, to be on TOP of the situation, the HEIGHT of power, his power ROSE, to be UNDER control, to FALL from power)

5. More is UP/Less is DOWN (mais é em cima, menos é em baixo) (the number is GOING UP, the number of errors is LOW, his income FELL).(Em português, os números, os preços, a inflação, são “altos e baixos”, “sobem e descem”)

6. Foreseeable future is UP and AHEAD (o futuro previsível é em cima e à frente) (All UPcoming events, What’sUP?, I’m afraid of what’s UP AHEAD of us)

7. Good is UP/Bad is DOWN (bom é em cima, mau é em baixo) (things are looking UP, things are at an all time LOW, he does HIGH quality work)(Em português falamos de “alta e baixa qualidade”)

8.Virtue is UP/Depravity is DOWN (virtude é em cima, depravação é em baixo) (to be HIGH-minded, to have HIGH standards, to be an UPright, UPstanding citizen, LOW trick, to STOOP to that, something is BENEATH me, a LOW-DOWN thing to do) (Em português temos “descer muito baixo”, com o mesmo sentido)

9.Rational is UP/Emotional is DOWN (racional é em cima, emocional é em baixo) (The discussion FELL into the emotional level, but I RAISED it back, he wouldn’t RISE ABOVE his emotions)(Em português, o “nível” de uma discussão também pode “descer”)

Para o par FRENTE/TRÁS , a metaforização do tempo<sup>447</sup> em inglês será a seguinte:

1. O futuro está à frente e o passado atrás:  
(the weeks AHEAD of us, that’s all BEHIND us now)
2. O futuro está atrás e o passado à frente:  
(In the FOLLOWING weeks (futuro), in the PRECEDING weeks (passado))

<sup>447</sup> Como notam os autores, são domínios por excelência a existirem lexicalmente metaforizados nas línguas naturais o tempo, a mente/espírito, as emoções, os pensamentos.

Estas metáforas “contraditórias” podem inclusivamente misturar-se sem qualquer contradição apercebida na mesma expressão corrente (por exemplo: We are looking AHEAD to the FOLLOWING weeks).

A segunda metaforização para tempo (como objecto móvel, que vem do futuro e por isso passa por ele primeiro, metáfora segundo a qual o tempo se move em direcção a nós - e objectos móveis, sem frentes inerentes, recebem frequentemente uma orientação frente/trás, sendo a ‘frente’ o sentido do movimento) é facilmente constatável também no português: ‘a semana que VEM’, ‘o ano que VEM’, ‘o ano que PASSOU’, ‘CHEGOU o momento...’

Nos exemplos dos autores para o inglês:

‘The time WILL COME when...’, ‘the time HAS GONE’, ‘the time HAS ARRIVED’.

Mas o objecto móvel que passa por nós não é a única metaforização lexicalizada para tempo: se se diz correntemente “As we GO THROUGH the years, as we GO FURTHER in the 90’s, we are APPROACHING the end of the year, - em português também podemos estar “a chegar ao fim do ano” - é o tempo que é conceptualizado como estacionário e somos nós que nos movemos através dele.

Para o par IN/OUT, dentro/fora (correspondente à metáfora do continente -CONTAINER - que é segundo os autores uma das mais difundidas em todas as línguas naturais) temos por exemplo ‘to be IN love’, ‘to come OUT of coma’, ‘to be IN the race’, ‘to be OUT of sight’, ‘to come INTO view’.

É porque o corpo está na mente - ou a constitui metaforicamente - que a experiência do espaço possibilita a significatividade directa e esquemática que os autores querem mostrar como fonte de metáfora. Parece no entanto que algumas metáforas existem que podem ser mais facilmente explicadas como um fenómeno de enciclopédia e pragmática<sup>448</sup> (não há por exemplo qualquer espacialização na ‘madeleine que está pelo tempo perdido’ de Proust). Nem toda a metaforicidade das línguas naturais pode ser identificada com a preservação de propriedades topológicas mediante mapeamento em domínios abstractos, inúmeras metáforas, poéticas ou lexicais, são questões pragmáticamente analisáveis em termos de descomposição semântica e dicionário/enciclopédia. Mas a originalidade desta abordagem é a explicação pela gramaticalização da espacialidade, a afirmação de que certos domínios-fonte das metáforas devem ser compreendidos independentemente da linguagem.

<sup>448</sup> Cf. Umberto Eco, *Artigo Metáfora, Enciclopédia Einaudi, Lisboa, INCM, e também U. Eco, Os Limites da Interpretação, Lisboa, Difel, 1992.*

Numa posterior elaboração teórica, Lakoff dá à sua teorização o nome de Hipótese da Espacialização da Forma que corresponderia a uma visão da língua como pensamento imagem-esquemático, evidenciando ainda mais a ligação deste estudo da metáfora com o estudo mais geral da sintaxe por Langacker. Esta hipótese é aliás prolongável numa teoria das bases neurológicas da competência gramatical <sup>449</sup>.

### **6. L.Talmy, a Dinâmica da Força: representações cognitivas de natureza topológico-dinâmica como estruturas semânticas profundas**

Na linha do programa de investigação da Linguística Cognitiva, o linguista Leonard Talmy pretende ter isolado uma categoria semântica na qual está gramaticalizada informação topológica e dinâmica proveniente da percepção, a que chamou categoria da dinâmica da força<sup>450</sup>. Para Talmy, está em causa na delimitação desta categoria a existência de uma estrutura conceptual comum através de domínios cognitivos, i.e. no sentido de Jackendoff, de um nível cognitivo profundo em que percepção, acção e linguagem se tornam compatíveis e que faz a mediação entre o mundo físico e a ontologia qualitativa das línguas naturais correspondente à nossa consciência linguisticamente estruturada.

Talmy parte dos seguintes pressupostos:

1. Cada enunciação numa língua natural é possibilitada por esquemas de imagem que organizam “cenas” no sentido de Fillmore<sup>451</sup> (i.e. informação perceptiva que organiza posterior material conceptual), esses esquemas constituem a Estrutura Conceptual da língua

2. Léxico e gramática são diferenciáveis pela oposição entre classes abertas (as lexicais - nomes, verbos, adjectivos - que são acrescentáveis) e classes fechadas (as gramaticais - tipos de construções sintácticas, conjunções, etc - de cardinal pequeno, fixo).

A ideia de Talmy é que as classes fechadas são uma gramaticalização de conteúdos abstractos, formais, por oposição aos lexicais: “As formas de classe fechada de uma língua tomadas conjuntamente representam um esqueleto de microcosmos conceptual que age como organizador de posterior

<sup>449</sup> Cf. A. Damásio e H. Damásio 1993, *Brain and Language*, in Goldman 1993.

<sup>450</sup> L.Talmy, *Force Dynamics in Language and Cognition*, *Cognitive Science* 12, 1988.

<sup>451</sup> Para o linguista Ch. Fillmore, o estudo da semântica é o estudo das cenas criadas ou activadas por enunciações (esta concepção cénica de sintaxe, nomeadamente dos casos nas línguas, é uma resposta à não definibilidade conceptual dos casos. Cf. Fillmore, *The case for case reopened*, *Syntax and Semantics* 8, Academic Press, 1977).

material conceptual. Por exemplo a classe fechadas das *preposições* especificaria informação morfológica<sup>452</sup> e a classes fechada dos *verbos modais* (o exemplo a seguir analisado) a categoria da dinâmica da força”.

Aquilo que é gramaticalmente especificado mostra, segundo Talmy, grande universalidade: o número, a quantificação, a estruturação do espaço, a estruturação do tempo, a tipicidade, etc. A esta lista Talmy pretende, no trabalho citado, acrescentar a dinâmica da força, categoria semântica gramaticalizada respeitante às interações de entidades em termos de força, como o exercício da força, a resistência à força, a ultrapassagem dessa resistência, o bloqueio da expressão da força, etc. Ao contrário dos sistemas de imagem gramaticalizados e estruturadores das “cenas” activadas nas enunciações em línguas naturais, mais usualmente estudados e que mostram estar relacionados com a percepção visual, a dinâmica da força relaciona-se com a sinestesia e mostrará ter representação gramatical não apenas em conjunções e preposições como na categoria dos verbos modais nos seus usos normal e epistémico/deôntico, em itens lexicais, extensões metafóricas e padrões de argumentação. Os exemplos a seguir referidos são os exemplos de Talmy para o inglês e os diagramas são adaptados do artigo citado<sup>453</sup>.

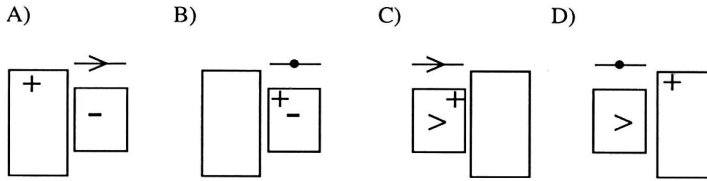
As distinções operatórias básicas para a análise de enunciações que activam um padrão de dinâmica da força em estado estável, são: a oposição de duas forças, sendo gramaticalizada a conceptualização da diferença de papel das duas entidades exercendo as forças - uma é focalizada e está em causa na interacção o facto de ela manifestar ou não a sua tendência de força (o Agonista, nos diagramas, um círculo), a outra é considerada pelo seu efeito na primeira entidade (o Antagonista, em diagrama, uma figura côncava).

O exercício de uma força é linguisticamente conceptualizado como existindo em virtude de uma tendência intrínseca para a manifestação, podendo a tendência ser para a acção (diagrama: cabeça de seta) ou para a inacção (diagrama: um ponto). Outro conceito associado à interacção é o de força relativa (diagrama: um menos e um mais). Da oposição das forças relativas das entidades em interacção resulta uma ocorrência, a acção ou inacção do Agonista (marcada por uma linha sob o Agonista, com uma seta ou ponto, para acção ou inacção).

<sup>452</sup> Note-se que grande parte das metáforas de Lakoff e Johnson revelam esquemas espaciais ou morfológicos correspondentes às preposições nas línguas naturais.

<sup>453</sup> Com permissão do autor.

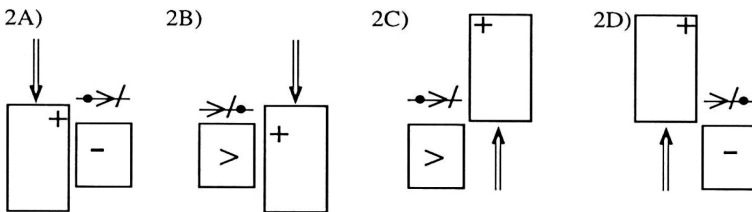
A partir destas distinções são caracterizáveis os quatro padrões mais básicos da dinâmica da força, envolvendo uma oposição em estado estável:



Exemplos.

- A) The ball kept rolling because of the wind blowing on it
- B) The shed kept standing despite the gale wind blowing against it
- C) The ball kept rolling despite the stiff grass
- D) The log kept lying on the incline because of the ridge there

Acrescentando o factor mudança no tempo, Talmy apresenta os padrões básicos da dinâmica da força com mudança de estado nos quais o Antagonista em vez de exercer estavelmente força sobre o Agonista inicia ou termina o estado de imposição de força (em termos gráficos, com uma seta dupla indica-se o início ou fim da imposição da força e com um traço na linha do estado resultante, o antes e o depois):



- A) The ball's hitting it made the lamp topple from the table
- B) The water's dripping on it made the fire die down
- C) The plug's coming loose let the water flow from the tank
- D) The stirring rod's breaking let the particles settle

Os múltiplos exemplos que Talmy enumera e analisa demonstram na sua opinião uma família de circunstâncias que a linguagem concebe (*construes*) conjuntamente e que pode ser chamada causativa, e em termos de dinâmica da força implica que um objecto tem uma tendência natural de força e manifestá-la-á a não ser que seja ultrapassado em força por um objecto exterior mais forte.

Se a dinâmica da força corresponde ao que Talmy chamará uma *folk physics*, uma física de senso comum, inicialmente respeitante a objectos, ela é metafóricamente transposta para a concepção linguística de psico e socio-interacções, i.e., é generativa de metaforizações intrapsíquicas e sociais.

Talmy investiga as configurações conceptuais gramaticalizadas e lexicalizadas do “eu dividido”, a semântica de *push* e *pressure* (Ex: He’s under a lot of pressure to keep silent/Our government exerted pressure on Nicaragua/Getting job security relieved the pressure on her to perform), de *urge*, *be reluctant*, *relent*, *give in*, *refuse to*, *will not*, etc.

A categoria da dinâmica da força é considerada então como generativa e a tese central é que esta categoria é o núcleo do significado dos verbos modais no inglês (*can*, *may*, *must*, *will*, *need*, *dare*, etc) no uso básico dos quais há uma referência nuclear à dinâmica da força, i.e. estão linguisticamente inscritas tendências e exercício e bloqueio de tendências, Agonistas e Antagonistas, cuja especificidade é fornecida pelo material lexical. O sujeito do modal é usualmente o Agonista enquanto o Antagonista está (em geral) implícito na situação concebida, ao que se acrescenta a curiosa característica dos modais no seu uso básico no inglês, que é o facto terem Agonistas sensientes e construir linguísticamente interacções psicossociais e não físicas (só *can* e *will not* têm frequente referência física).

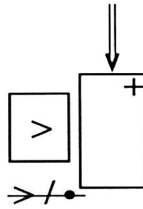
O uso dos modais que permite sujeitos gramaticais não sensientes é o uso epistémico como em “The pear *could/may/must/should/needn’t* be ripe by now”. Este uso consiste na aplicação da semântica modal ao raciocínio sobre proposições e não ao conteúdo destas. Historicamente, no inglês, os modais só adquiriram uso epistémico depois do uso deontico básico: mais uma extensão metafórica linguisticamente possibilitada pela generatividade da categoria semântica.

Segundo Talmy, a dinâmica da força está ainda inscrita na possibilidade linguística de argumentação, i.e. na retórica da persuasão como esforço para exortar, convencer, demonstrar logicamente, que envolve a oposição de posições (forças) por e contra e o desenrolar dessa oposição.

Os *logic-gaters* da dinâmica da força são segundo Talmy uma classe gramatical fechada que molda o fluxo ilocucionário e configura o espaço da argumentação como espaço de oposição: exemplos são “*yes, but*”, *besides*, *nevertheless*, *moreover*, *granted that*, *instead*, *all the more so*, *even so*, *well*, etc.

Tomando como exemplo o significado argumentativo de “yes, but”, Talmy diagramatiza-o assim:





Sendo aquilo que o diagrama cobre parafraseável como ‘o seu último argumento, inclinado para uma particular conclusão, é bom até certo ponto, mas há uma consideração mais importante em causa que conduz à conclusão oposta, por isso aquilo que eu vou dizer tem mais força do que aquilo que disse’.

Para Talmy, os modelos conceptuais gramaticalizados têm uma similaridade espantosa com concepções ingénuas do mundo físico (i.e. correspondem a uma *folk physics*, uma física de senso comum, uma concepção incipientemente teórica, inclusive da causalidade) e com pensamento científico historicamente existente: é por exemplo notório que os conceitos básicos da dinâmica da força (tais como bloquear, permitir, resistir, ultrapassar) não têm qualquer contrapartida na física (especificamente porque implicam a adscrição de identidade e tendências intrínsecas a porções conceptualmente delimitadas do *continuum* espaço-temporal). É também óbvio que estes sistemas de imagem, tendo embora inscrita uma concepção extremamente simples da causalidade, seleccionam, focalizam ou recortam a actualidade da causação: mesmo porque “eventos”, i.e. porções conceptualmente delimitadas a partir do *continuum* daquilo que ocorre, implicam uma representação fixada, extraída da causalidade espaço-temporal. As construções linguísticas apresentam eventos autónomos, sem precursores causais, sem consequentes, “esta partição extrínseca, que isola e descausaliza, que a linguagem pode impor conceptualmente ao fluxo da ocorrência está em contraste directo com a perspectiva da física, em que tudo é um *continuum causal* ininterrompido”<sup>454</sup>. O que não implica a inexistência da razão de ser ou da condição de possibilidade da ontologia qualitativa gramaticalizada, e é essa condição de possibilidade que Talmy procura relevar ao analisar a gramaticalização de esquemas de imagem que constituem a Estrutura Conceptual das línguas naturais.

## 7. Conclusão: Gramaticalidade e fenomenalização

As teses dos linguistas cognitivos até aqui analisadas mostram a necessidade de fenomenologia na descrição dos fenómenos simbólicos como

<sup>454</sup> Talmy 1988, p.93.

processos e ocorrências na mente-cérebro. Essas teses procuram, fundamentalmente, afirmar a dimensão imagética da gramaticalidade, a relação entre a espacialidade e a possibilidade de metáfora, a gramaticalização da dinâmica da força, a existência de uma estrutura conceptual. As análises de R. Langacker, feitas sob o pressuposto de que “A mente é aquilo que o cérebro faz” conduzem à revelação do envolvimento da conceptualização do tempo e do espaço na descrição da ocorrência das unidades simbólicas básicas<sup>455</sup>, icônicas e esquematicamente caracterizáveis. As análises de Lakoff, Talmy e Jackendoff, análises da metaforicidade, da dinâmica da força e da Estrutura Conceptual, supõem primitivos como posições espaço-temporais, regiões do espaço, Objectos, Trajectos, e Eventos. Estas noções, que são incompatíveis com a noção de cálculo ininterpretado fundamental na descrição lógica da forma simbólica, tornaram-se moeda corrente na teoria linguística. Como Jackendoff afirma explicitamente a propósito daquilo a que chama a conexão entre a Estrutura Conceptual (nível computacional para o significado) e o modelo 3D na teoria computacional da visão de D.Marr, “A linguagem está apenas a capitalizar sobre o que já está presente no entendimento espacial” (e temporal). Por analogia com a terminologia de Chomsky aquilo que é, para os linguistas cognitivos, subdoxástico e presente na experiência mental linguística (experiência que não é, citando Langacker, periféricamente iniciada) é uma intimidade prática como o espaço-tempo. Apesar da contestação, a directriz chomskyana que pela primeira vez afirma a linguística como ciência cognitiva<sup>456</sup> é reafirmada pelos linguistas cognitivos — apenas os pressupostos da abordagem formalista são abalados pelo novo tipo de descrições.

A necessidade de fenomenologia pode ser considerada uma consequência do imperativo de Langacker da naturalidade da descrição linguística<sup>457</sup> e contradiz a metáfora do sistema formal para as línguas naturais que é a chave, segundo G. Lakoff, da abordagem chomskyana, para a qual os estados e processos mentais linguísticos enquanto ocorrências na mente-cérebro só podem ser causalmente explicados em função de regras e representações simbólicas subdoxasticamente conhecidas, o que dado o compromisso anti-nominalista de Chomsky faz da mente computacional linguística uma máquina simbólica física.

A “viragem fenomenológica” na linguística deve ser entendida, se considerada filosoficamente, como uma consequência teórica e descritiva

<sup>455</sup> Predicações relacionais e nominais.

<sup>456</sup> Essa directriz é, sumariamente, a naturalização dos fenómenos simbólicos, do conhecimento e uso de línguas naturais.

<sup>457</sup> Não é por estarem disponíveis formalismos lógicos, algorítmicos e discretos, para a modelização de fenómenos simbólicos, nomeadamente nas línguas naturais, que eles necessariamente se adequam ao ser formal do objecto descrito.

da ancoragem dos fenómenos simbólicos na percepção. Esta ancoragem perceptiva desmente o fechamento ao semantismo dos agenciamentos formais de uma máquina-guiada-pela-sintaxe, parafraseando a formulação (com defesas diferentemente colocadas) da noção cognitivista-simbólica de mente por Jerry Fodor e John Searle. As descrições, a que podemos chamar fenomenológicas, dos linguistas cognitivos substituem as descrições formalistas que têm como paradigma a obra de Chomsky e a inspiração desta na teoria dos sistemas formais.

Entende-se por necessidade de fenomenologia o facto de se ser conduzido à consideração da subjectividade e do problema do aparecer no ser quando se põe o problema teórico e científico da descrição da forma do sentido. A descrição da possibilidade de gramaticalidade envolve para autores como Jackendoff, Langacker, Talmy, Lakoff e Petitot problemas como a intencionalidade e a conceptualização<sup>458</sup> do espaço e do tempo, que não existem em Chomsky. Por isso, mesmo se o problema teórico de fundo continua a ser nestes autores a naturalização do simbólico, i.e. o tratamento do sentido articulado como um fenómeno natural, a hipótese fundamental não é já a da existência de um cálculo ininterpretado fisicamente implementado, em primeiro lugar porque a possibilidade da forma do sentido não pode excluir a relação entre subjectividade e mundo sensível, uma vez mostrado que as categorias formais dos estados mentais linguísticos supõem a espaço-temporalidade e portanto a auto-afecção do sujeito enquanto fenómeno.

Os exemplos aqui analisados provêm da linguística, no entanto vários autores no campo das ciências cognitivas se têm aproximado, por necessidade interna à suas análises, da filosofia fenomenológica continental. Segundo Jean Petitot, “Se há uma concepção filosófica que merece ser retomada pelo programa de pesquisa das ciências cognitivas (que é um programa de naturalização da mente e do sentido) é a fenomenologia husserliana (que) permaneceu um dos mais altos lugares de reflexão ao mesmo tempo científica e filosófica sobre o espírito e a consciência”<sup>459</sup>. Mas se as ciências cognitivas retomam o problema fenomenológico fazem-no no sentido da tentativa de naturalização do conteúdo intencional que permanece na mente uma vez feita a *epochê* do mundo e portanto contradizendo o anti-naturalismo husserliano

Como pressuposto do confronto entre a inspiração lógica e a inspiração fenomenológica pela modelização da forma na mente está o facto

<sup>458</sup> ‘*Sintese subjectiva*’.

<sup>459</sup> Cf. J. Petitot, *Phénoménologie naturalisée et morphodynamique: la fonction cognitive du synthétique a priori*, *Intellectica* n°17, 1993/2, p.80.

de, para o nominalismo de qualquer semântica formal inspirada na lógica, a forma qualitativa do mundo linguisticamente descrito ser forma de linguagem irremediavelmente desligada do mundo físico<sup>460</sup>. Dado este pressuposto é necessário aceitar o comentário de Jerry Fodor, segundo o qual, num psicologia científica, que só pode ser funcionalista e computacional, é impossível introduzir a relação cognitiva que o organismo tem com o ambiente. Por isso a psicologia é dificilmente incrustável no corpo geral das ciências da natureza, devendo ser metodicamente solipsista.

Para a teoria dualista e ao mesmo tempo naturalista dos sistemas cognitivos que é o cognitivismo simbólico, cujo paradigma aqui apresentado foi a linguística chomskyana, inevitavelmente existe uma cisão entre a lógica, teoria dos sistemas formais, e a física, matematização da espaço-temporalidade, o que justifica o solipsismo metódico das descrições, uma vez que é a lógica que inspira a modelização da cognição que existe fisicamente implementada. A “informação” externa, descritível pela física, não pode ter a priori qualquer significação para o sistema cognitivo: submetida a uma transdução modular periférica cognitivamente opaca, só depois (de alguma maneira) se converterá na informação interna computacionalmente significativa que é o objecto teórico de cognitivistas como J.Fodor e N.Chomsky, para os quais a explicação causal dos estados mentais proposicionais e linguísticos é função de regras e representações (já) simbólicas. Só assim a teoria da mente pode ser a teoria de uma máquina guiada pela sintaxe.

Evidentemente, o que fica então em causa nas ciências cognitivas é a compatibilidade desta afirmação dualista<sup>461</sup> com teorias do simbólico que são teorias da naturalização, nas quais é visada a existência física e não apenas a plausibilidade descritiva de um modelo de sistemas cognitivos e perceptivos. Recordando as palavras de Chomsky, se a linguística é uma ciência cognitiva e toda a ciência é uma ciência da natureza, não há razão para estudar os produtos da mente humana de um modo diferente do que se utiliza para qualquer outra parte da natureza. Mas então “a orientação

<sup>460</sup> Como diz J. Petitot, “Du point de vue de Sirius, on ne peut qu’être frappé de la scission qui clive les disciplines positives modernes. D’un côté, il y a les sciences expérimentales comme la physique, la chimie ou la biologie, sciences objectives souvent formalisées et dont l’édifice théorique culmine dans de profondes théories mathématiques, comme la mécanique rationnelle, l’optique ondulatoire, la théorie quantique, la relativité générale, la théorie des transitions de phases, etc. D’un autre côté, il y a les analyses logico-formelles des langages. Pour les premières, l’objectivité s’enracine en dernière analyse dans la géométrie de l’espace-temps. Pour les secondes, les langages réfèrent en dernière instance au “monde” comme apparaît, c’est-à-dire, à la manifestation phénoménale qualitative. Entre physique et logique, le divorce est totale” [Petitot 1992, *Physique du Sens*, p. 20].

<sup>461</sup> Claramente defendida por exemplo em Pylyshyn 1984, p.166-167: ao discutir o problema a que chama o problema de uma *bridge from physical to symbolic*, i.e. de uma ponte do físico para o simbólico, Z. Pylyshyn faz notar que “os aspectos relevantes do ambiente não são geralmente descritíveis em termos físicos” e que “as regularidades psicológicas são atribuíveis a propriedades percebidas não fisicamente descritas”. Daí “o insucesso geral da psicologia perceptiva quando tenta descrever adequadamente estímulos em termos físicos”.

dogmaticamente proposicionalista (ou dogmaticamente simbólica) do cognitivismo simbólico (...) (torna-se) incompatível com qualquer tese naturalista, pois não existem formas simbólicas na natureza interna ou externa. Não podem existir senão formas geométricas e dinâmicas (daí que toda a naturalização da mente, da linguagem e do sentido pressuponha uma revolução da concepção do formal herdada do formalismo lógico. Ela pressupõe categoricamente que as formas da mente, da linguagem e do sentido sejam formas geométricas e dinâmicas, evidentemente simbolicamente tradutíveis e manipuláveis a níveis superiores de representação<sup>462</sup>”.

Não há objectividade originariamente simbólica: é este o limite das teses dos cognitivistas simbólicos. Nomeadamente, um computacionalismo não pode ser naturalista: é o problema que J.Searle localiza quando afirma que a sintaxe não é uma propriedade física e que por isso há uma ilusão na caracterização computacional da mente-cérebro. Se “A finalidade da ciência natural é descobrir e caracterizar traços intrínsecos ao mundo natural”, “pelas suas próprias definições de computação e cognição, não há nenhuma maneira de uma ciência cognitiva computacional alguma vez ser uma ciência natural, porque a computação não é um traço intrínseco do mundo. É atribuída relativamente a observadores”<sup>463</sup>.

Ao dualismo inerente a uma abordagem da gramaticalidade apoiada na lógica pode ser contraposta a aplicação ao problema da articulação do sentido de teses de Husserl acerca do estatuto do aparecer, correlativo da *cogitatio* ou vivência da consciência e portanto da subjectividade, dada a hipótese central da linguística cognitiva, que é a hipótese da ancoragem perceptiva das categorizações linguísticas. É essa a fundamentação filosófica evocada em vários trabalhos de Jean Petitot, nomeadamente em “*Morphogénèse du Sens*” e “*Physique du Sens*”.<sup>464</sup>

Em suma, no tratamento das questões da naturalização da mente e do mental, é preciso saber de que são feitos os conceitos (por exemplo de Chomsky) de natureza e de psicologia com que se trabalha e que lhe permitem as afirmações acerca do ser da mente como parte da natureza. É nomeadamente contestável que sejam convicções possíveis quando se trata do “estudo paralelo da mente e da natureza”, nas palavras que Francisco Varela<sup>465</sup> usa para caracterizar a ciência cognitiva, a existência de uma psicologia sem mente fenomenológica, no sentido de vivência da consciência

<sup>462</sup> J. Petitot, *Le Physique, le Morphologique et le Symbolique – remarques sur la vision*, *Revue de Synthèse*, n<sup>o</sup>s 1-2, 1990.

<sup>463</sup> J. Searle 1992, p. 212.

<sup>464</sup> Conjuntamente com um transcendentalismo de tipo kantiano estendido, nas palavras do autor, à “parte máldita” fenomenológica.

<sup>465</sup> Varela, F., *Connaitre*, Paris, Seuil, 1989.

ou o pressuposto da ausência de uma “ponte” entre o físico e o simbólico, i.e. a ausência de explicação da passagem daquilo que merece descrições físicas àquilo que merece descrições semânticas e experienciais.

É fácil imaginar que as crispações dos filósofos (especialmente dos filósofos deflacionários, eliminacionistas ou pragmatistas) perante uma teoria do tipo da de Chomsky dizem respeito à leveza de alma com que os psicólogos cognitivos e os linguistas fazem proliferar as representações mentais, o museu interior, impedindo o caminho do inquérito por sucumbirem à dificuldade de compreender cientificamente o corpo e especialmente o cérebro, desde sempre a principal razão para a necessidade de pensarmos que possuímos uma mente e para a necessidade de analogias para a capturar.

A pedra de toque do cognitivismo é de facto o afastamento de preconceitos contra entidades e processos mentais, em suma, contra o vocabulário mentalista. Mas, como Richard Rorty nota com justiça em “*A Filosofia e o Espelho da Natureza*” (1989) não se pode objectar acerca das “representações mentais” dos cognitivistas que elas *representam* bem ou mal, clara ou obscuramente alguma realidade<sup>466</sup>. A mente enquanto objecto teórico das ciências cognitivas já não é, paradoxalmente, uma mental, a mente das metáforas ópticas, a mente espelho da natureza: ela é sempre claramente considerada como uma especificação abstracta do cérebro e por isso é uma mente natural, incorporada.

Mas, se não é essa a verdadeira objecção às representações mentais dos cognitivistas, até porque não é difícil abdicar da qualificação “mental” a favor da qualificação “interno”, continuando a falar de regras e representações, o que é de facto relevante é o dualismo físico-simbólico e que a proliferação de entidades e processos mentais possa culminar numa psicologia sem realidade psicológica. Mais do que a mente computacional, a inteligência algorítmicamente caracterizável, o enigma da psicologia, mesmo de uma psicologia fazendo explicitamente parte das ciências naturais é o facto de ter por tema a subjectividade<sup>467</sup>, que não pode ser um princípio sem princípio numa teoria científica. Daqui surgem todas as dificuldades geradas pela restrição à mente computacional da noção de mente reintroduzida “após” o behaviorismo psicológico e filosófico.

De qualquer modo, e foi isso o que se tentou provar neste artigo, o problema da forma de linguagem fez com que as ciências cognitivas através da linguística cognitiva recuperassem empiricamente a *questão*

<sup>466</sup> Cf. Rorty, R., *A Filosofia e o Espelho da Natureza*, Lisboa, Dom Quixote, 1988, capítulo *Epistemologia e Psicologia Empírica*, especialmente p. 196.

<sup>467</sup> É a questão de Thomas Nagel, cf. *The View from Nowhere*, Oxford University Press, 1986.

*fenomenológica*, a questão da cisão entre física e aparecer, e essa recuperação foi um dos factores do abalo da centralidade da lógica, *medium* do seu parentesco com a filosofia, na modelização dos estados mentais linguísticos.

## LINGUISTAS E FILÓSOFOS: MANEIRAS DE FAZER TEORIA DA MENTE<sup>468</sup>

“A lógica e as teorias do pensamento e da significação devem ser claramente demarcadas da psicologia”, Michael Dummett,

*Origins of Analytical Philosophy*, p.23

“(…) a filosofia americana está, em grande parte, no mau caminho (…) a tendência geral é para ir na direcção do cientismo, o que me parece uma opção estéril”, Michael Dummett, entrevista a Fabrice Pataut, *disputatio*, 3

### 1. A revolução cognitiva na teoria da mente e os seus pressupostos filosóficos

Que pontos de comunicação podem ser encontrados hoje nas investigações de linguistas e filósofos? Penso que linguistas e filósofos se encontram em torno da teoria da mente. É preciso recolocar temporalmente a questão: esta situação de que vou falar—o encontro de linguistas e filósofos em torno da teoria da mente—começou com o trabalho de Chomsky. O trabalho de Chomsky é usualmente apontado como um dos factores responsáveis pela revolução cognitiva em psicologia. Esta revolução teria eliminado o behaviorismo e o seu tabu do mental e do interior no estudo da cognição e do comportamento, e este teria sido o passo fundamental para a constituição do campo das ciências cognitivas, no qual hoje se encontram a linguística, a filosofia, mas também as neurociências, a psicologia cognitiva e a IA.

É importante defender aqui um uso alargado do termo “psicologia” para cobrir todo o estudo do mental e da cognição, ou todo o estudo de sistemas que se comportam<sup>469</sup>. A oscilação dos termos “mente” e “cognição”

<sup>468</sup> Este texto foi objecto de uma apresentação pública na Universidade do Minho, a 9 de Abril de 1999, inserido num ciclo de encontros de linguística e filosofia. Agradeço à Prof<sup>a</sup> Doutora Pilar Barbosa, do Instituto de Letras e Ciências Humanas da UM, o convite que me fez e a oportunidade de discussão pública que ele re-presentou. Agradeço ainda ao Dr. João Alberto Pinto, da FLUP, por todas as sugestões e críticas que me foi fazendo, assim como pelo contacto com o seu próprio trabalho em filosofia da mente e em lógica.

<sup>469</sup> Talvez mesmo da cognição como natural kind, por oposição por exemplo à consciência. Para esta consideração da cognição como tipo natural por oposição à consciência, cf. Pylyshyn, Zenon, *Computation and Cognition*, 1984, *What is cognitive science the science of?*, p. 257



entre aquilo a que podemos chamar o mental-natural e o mental-artificial é comum na literatura, e a hipótese da Inteligência Artificial forte<sup>470</sup> (i.e., a ideia de que poderá vir a existir uma “simulação” de inteligência consciente que será não apenas um modelo mas uma replicação do fenómeno, dada a hipótese de a existência física da inteligência consciente ser um fenómeno de organização), é um limite teórico sempre presente. Como diz David Chalmers, um filósofo australiano, professor na Universidade da Califórnia (Santa Cruz)\*, a hipótese da IA forte supõe (a existência de) “uma classe não vazia de computações tal que a implementação física de qualquer computação dessa classe seria suficiente para a existência de um mente e em particular de uma mente consciente”<sup>471</sup>. A ideia de implementação conduz segundo Chalmers ao princípio da invariância organizacional<sup>472</sup> que sustenta a IA forte. As ciências cognitivas correspondem afinal, elas próprias, à abertura de um campo teórico transversal em que “o cognitivo” e a sua caracterização se tornam de certo modo independentes do facto de essa cognição ser natural ou artificial. Considero que esta indistinção entre natural e artificial no estudo de sistemas físicos capazes de comportamento racional, inteligente, apropriado<sup>473</sup>, que é obviamente problemática, é constitutiva do campo, e que a hipótese segundo a qual a cognição depende da existência de Sistemas Simbólicos Físicos (na expressão de A. Newell e H. Simon), hipótese que permite o transporte dos resultados obtidos no estudo de algoritmos para o estudos de sistemas físicos, é uma hipótese matricial<sup>474</sup>.

\* À data em que o artigo foi publicado. Hoje (2009) David Chalmers é professor na Australian National University (ANU).

<sup>470</sup> Cf. a terminologia de John Searle para a distinção entre Inteligência Artificial forte e fraca (Searle 1998).

<sup>471</sup> Chalmers 1996, p.314.

<sup>472</sup> O problema é saber, dado o surgimento da consciência a partir do físico, em virtude de que tipo de propriedades ela surge. Se se supuser que o abstracto químico ou quântico do cérebro são irrelevantes para a produção de consciência, a sugestão acerca dessas propriedades apontará para uma organização funcional como padrão abstracto de interacção causal entre partes do sistemas e entre estas e inputs e outputs externos. Segundo o princípio da invariância organizacional, dado qualquer sistema que tenha experiência consciente, qualquer outro sistema que tenha a mesma organização funcional terá experiências qualitativamente idênticas, podendo ser deixada inespecificada a natureza dos componentes e estados dessa organização funcional. De acordo com o princípio, a consciência é um invariante organizacional. Cf Chalmers 1996, p. 247.

<sup>473</sup> Cf. Descombes 1995, pp.194-203. Descombes nota, analisando as conferências de Herbert Simon sobre as “ciências do artificial”, que o que está aí em causa não é uma saída das ciências da natureza [“o artificial não se opõe ao natural como um reino a um outro mas como um ponto de vista a um outro. Tudo o que é artificial é igualmente natural e deve receber em última análise uma explicação física”]; o que está em causa é a consideração de sistemas segundo um interface dentro/fora e a racionalidade ou adaptatividade dessa interface, i.e., aquilo a que se poderia chamar teleologia. Um sistema natural pode então, como é óbvio, ser também analisado de acordo com estes parâmetros, o que interessa é que no centro da descrição estejam finalidades que “liguem” o sistema interno ao externo. Cf. também Herbert Simon, *The Sciences of the Artificial*, 1969. Herbert Simon inclui a psicologia, definida como ciência dos “behaving systems”, nas ciências do artificial. Essa inclusão deve ser vista em contraste por exemplo com neurofisiologia do [mesmo] sistema, que se reporta exclusivamente ao meio interno e não à interface. Vincent Descombes interpreta a descerebralização da mente, da inteligência e da racionalidade que sai da abordagem de Simon considerando [citando Aristóteles, *As partes dos animais*] que se trata aqui de casos nos quais a natureza formal tem mais importância do que a natureza material, cf. Descombes 1995, p.214.

<sup>474</sup> Cf. Miguens, Sofia, 1995, p.387. A ideia de Sistemas Simbólicos Físicos foi considerada por A. Newell e H. Simon e os seus colegas em Carnegie-Mellon a ideia nuclear da área das ciências dos computadores. A existência de um sistema simbólico físico seria necessária e suficiente para levar a cabo acções inteligentes.

A origem da ideia de mente computacional, por cuja introdução na linguística Chomsky foi responsável, pertence à história da lógica e teve como apoio uma noção do lógico e matemático inglês Alan Turing, a noção de Máquina de Turing Universal<sup>475</sup>. Pode-se dizer portanto que foi a conjugação de resultados lógicos com a vontade de uma solução materialista para o *mind-body problem*, que preparou o terreno para a formulação da hipótese fundadora dos Sistemas Simbólicos Físicos.

A hipótese dos Sistemas Simbólicos Físicos corresponde ao que os filósofos (e não só) chamam a Teoria Representacional Computacional da Mente (TCRM), a qual supõe que uma explicação cognitiva é o tipo de explicação científico-natural que é computacional e representacional.

A ser assim, deverá existir um nível independente, autónomo, dos sistemas físicos (por exemplo dos cérebros) descritível através de noções a que podemos chamar latamente noções de linguagem, noções simbólicas, como representações e computações. Esse nível é o âmbito do cognitivo e dada a autonomia da caracterização pode mesmo pôr-se a hipótese de essas descrições serem implementáveis noutras realidades materiais (esta é a ideia da realizabilidade múltipla)<sup>476</sup>.

A teoria da linguagem de Chomsky é frequentemente considerada como tendo sido em parte responsável pela própria génese deste campo teórico transversal das ciências cognitivas. É aliás por vezes apontado como um marco inicial na constituição deste o encontro de 10-12 de Setembro de 1956 do Institute of Radio Engineers do MIT, um encontro sobre teoria da informação, no qual participaram Allen Newell e Herbert Simon, George Miller e Noam Chomsky<sup>477</sup>, simbolizando uma relação de disciplinas nucleares no campo, respectivamente a IA, a psicologia cognitiva, e a linguística e apresentando respectivamente os seguintes papers: *The Logic Theory Machine*, *The Magical Number Seven, Plus or Minus Two*, e *Three Models for the Description of Language*.

Acerca de duas destas disciplinas nucleares das ciências cognitivas, diz Herbert Simon que “A linguística e a IA tiveram relações cordiais desde então e ainda bem pois assentam conceptualmente numa mesma visão da mente humana”<sup>478</sup>. A afirmação é obviamente contestável, mas o ponto aqui é a partilha de uma visão teórica determinada daquilo que constitui o mental, visão que tentarei esclarecer. Como disse, o encontro de 1956

<sup>475</sup> Uma Máquina de Turing é um autómato abstracto, sem especificações quanto à realização física de estados internos e transições de estados. Mas estes têm obviamente que existir. Uma Máquina de Turing universal pode simular qualquer máquina de Turing. As Máquinas de Turing são capazes de executar qualquer função computável

<sup>476</sup> Para a hipótese da realizabilidade múltipla que acompanha o funcionalismo e a sua contestação à Teoria da Identidade exponenciada por exemplo por J. J. C. Smart e D. Armstrong, cf. Kim 1996, cap. 3, *Mind as the Brain*, e cap. 4, *Mind as a Computer*

<sup>477</sup> Para descrições da importância desse encontro, cf. Gardner 1985, p. 28 e Dennett 1995, p. 384.

<sup>478</sup> Simon, 1969, p.47. Cf. comentário in Dennett 1995, p. 385.

do *Institute of Radio Engineers* é um marco do campo das ciências cognitivas, campo no qual hoje muitos filósofos da mente situam o seu trabalho<sup>479</sup>. Dado o seu papel central na revolução cognitiva em teoria da mente, Chomsky provocou um abalo na filosofia e foi e continua a ser muito importante para muitos filósofos (como Quine, Putnam, Fodor, Dennett, Searle, etc.)<sup>480</sup>. Estou a falar sobretudo da filosofia analítica contemporânea, da filosofia americana, e mais especialmente da *philosophy of mind*, que frequentemente se considera a si própria como tendo como problema central a questão dos fundamentos da ciência cognitiva - sendo portanto uma filosofia da psicologia, se pudermos entender esta como natural ou artificial.

As hipóteses de Chomsky acerca da linguagem e da mente, eram, quando apareceram, muito ousadas, e iam contra grande parte da ortodoxia dominante em filosofia na altura, pois prescreviam um estranho “objectualismo” mental naturalista que não parecia compatível por exemplo com teses pragmatistas muito difundidas, que consideravam injustificada qualquer referência ao “interior” no tratamento da significação e compreensão linguísticas. Pense-se por exemplo na famosa tese wittgensteiniana das *Investigações Filosóficas* segundo a qual “O significado é o uso” e compreender, por exemplo compreender uma palavra, é um domínio (*mastery*) no seio de um jogo de linguagem no qual ela entra em cena. Esta tese indica, obviamente, como método para a teoria da mente e da significação um olhar voltado para as práticas e não “para dentro da caixa” como Chomsky recomenda, e portanto uma exteriorização e não uma interiorização da teoria da linguagem e do mental.

Mas o que é este estranho objectualismo interiorista de Chomsky? De que fala Chomsky? Desde *Syntactic Structures* (1957)<sup>481</sup> a *The Minimalist Program* (1995) Chomsky fala de regras e representações realistamente consideradas através de uma teoria formalizada e afirma que isso é a nossa mente (uso o palavra realismo no sentido filosófico de existência independente do pensamento e da descrição).

Esta investigação de Chomsky é uma investigação empírica naturalista— considera-se que a linguagem é uma parte do mundo natural<sup>482</sup>—por

<sup>479</sup> O trabalho de alguns filósofos tornou-se por isso um trabalho “impuro”, na expressão de Dennett (cf. Dennett 1998, p.366) isto é, um trabalho que não se limita ao apriorismo do *armchair knowledge* mas conjuga análise conceptual e ciência empírica, seja esta psicologia cognitiva, neurociência ou IA. Pense-se por exemplo em filósofos como Jerry Fodor, Paul e Patricia Churchland ou Andy Clark. Para Dennett, esse trabalho impuro do filósofo é um reconhecimento do facto de que os problemas filosóficos mais desafiadores e pertinentes quanto à mente surgem dessa proximidade com as ciências.

<sup>480</sup> Cf. Dennett 1995, p. 385. Dennett relata ter perguntado a Quine em 1960 a quais dos seus críticos deveria dedicar-se. Quine sugeriu-lhe imediatamente o trabalho de Chomsky.

<sup>481</sup> Referido em Putnam 1960 como exemplo.

<sup>482</sup> Cf. Chomsky 1995, pp 167-168.

acaso com propriedades não usuais nos sistemas biológicos, nomeadamente a infinitude discreta, mas ainda assim parte do mundo natural—e uma investigação feita à maneira de Galileu—i.e, assim como a física moderna concede mais realidade aos modelos matemáticos do que ao sensorialmente percebido, a investigação linguística chomskyana concede mais realidade aos modelos formais de um fenómeno do que ao “sensorialmente” percebido, digamos neste caso aos acontecimentos de língua-E. Aliás, a própria divisão chomskyana entre Língua-I<sup>483</sup> e Língua-E indicia já o foco da atenção que é o interior, a generatividade como propriedade de uma entidade formal, uma gramática, e não tanto os eventos exteriores de enunciação, os acontecimentos.

Por isso Chomsky é, como teórico, rigorosamente revolucionário. Aparece como um humanista, um professor de linguística, que faz ciência natural e estuda um sistema físico, o cérebro humano, com todas as ambiguidades que advêm deste desvio das humanidades para a ciência natural, desvio que torna possíveis contestações que ainda hoje os linguistas “formais” devem sentir a partir de colegas voltados para o lado mais artístico do que científico do trabalho sobre a linguagem<sup>484</sup>.

Os filósofos, enquanto epistemólogos e metafísicos da mente, imediatamente pretenderam esclarecer os pressupostos desta investigação empírica naturalista, feita por Chomsky e outros cognitivistas (como Allen Newell, Herbert Simon, Marvin Minsky, ou mais recentemente Philip Johnson Laird, David Marr, etc, para dar alguns exemplos de autores que trabalham em modelos computacionais para a cognição, em áreas diversas desde a linguagem, à visão, à decisão racional, etc...).

Chomsky, mais familiar, penso, aqui, é um exemplo central daquilo a que a partir da filosofia interessa chamar o cognitivismo simbólico. A revolução cognitiva iniciada por Chomsky foi acompanhada por hipóteses filosóficas materialistas – hipóteses de filosofia das ciências naturais – neste caso hipóteses acerca do lugar da mente na natureza, que procuraram conceber como é que tem que ser o mundo para haver mente tal como Chomsky a descreve. Essas hipóteses são fundamentalmente o funcionalismo, o representacionismo e o computacionalismo.

Em termos de filosofia, o responsável por esse trabalho pioneiro foi o filósofo Hilary Putnam, em textos fundadores como *Minds and Machines*, (1960) onde se expõe o chamado funcionalismo das Máquinas de Turing,

<sup>483</sup> Estados da mente-cérebro por contraposição a eventos externos.

<sup>484</sup> O linguista e cientista cognitivo Steven Pinker, em *The Language Instinct*, descreve a teoria de Chomsky como uma importante e necessária demonstração das inconsistências daquilo a que chama “the standard social science model”, como um dos fundamentos da vida intelectual do século 20, o qual consideraria que tudo o que é mental e humano é resultado da cultura (cf. Pinker 1994, p. 23). Este modelo encontraria supostamente apoio por exemplo na “relatividade linguística”, tal com esta foi defendida sob o nome de tese de Sapir-Whorf.

ligado à contestação à teoria da Identidade entre Estados Mentais e Estados Cerebrais (esta seria uma posição materialista-paroquial, demasiado humana, quanto ao *mind-body problem*) ou *Psychological Predicates* (1967), depois intitulado *The Natural of Mental States*.

Do ponto de vista funcionalista, um teórico da identidade confunde estados do hardware com estados lógicos, que são o que está em causa no mental (por exemplo quando está em causa o estatuto teórico da afirmação empírica “A dor é estimulação das fibras C”). Além de defender a caracterização dos estados mentais como estados lógicos e não estados físicos, Putnam argumenta contra o fisicalismo dos tipos (que a Teoria da Identidade supõe<sup>485</sup>), e defende a ideia de realizabilidade múltipla: as propriedades mentais (por exemplo a dor) realizadas nos humanos (por exemplo por activação das fibras-C) podem ser noutros seres físicos realizadas em mecanismos diferentes, com propriedades físicas diferentes. A caracterização do mental não tem por que incluir constrangimentos quanto ao sistema físico que deve realizá-lo.

Para compreender o teor do funcionalismo pense-se: se Chomsky diz continuamente que estuda o cérebro, porque é que não fala de neurónios, neurotransmissores, neuroanatomia? E porque é que não lhe interessa referir o falante como ser consciente? Chomsky não fala nem do cérebro nem da consciência porque fala daquilo que é feito informacionalmente e computacionalmente pelo cérebro e não da máquina que faz nem da presença (ou ausência) na consciência desse fazer: a separação hardware/software é central aqui (como diz a palavra de ordem cognitivista, ‘a mente está para o cérebro como o software para o hardware’). É uma questão de níveis: grosseiramente poder-se-ia afirmar que o nível simbólico-computacional no qual Chomsky está interessado está “entre” a neurofisiologia – é realizado por ela – e a consciência—porque constrange a forma dos nossos estados mentais linguísticos pelos quais intencionamos conscientemente o mundo (uso aqui a palavra “intencionar” no sentido filosófico de *aboutness* e não como voluntaridade).

Segundo a concepção funcionalista do mental, o facto de um sistema ter mente está ligado à realização por esse sistema de uma determina-

<sup>485</sup> A identidade dor-estimulação das fibras C seria uma identidade teórica, descoberta empiricamente, tal como “A água é H<sub>2</sub>O” ou “A SIDA é infecção pelo HIV”. Identidades deste género, cientificamente relevantes, dizem o que uma coisa realmente é. Estas identidades podem ser formuladas em termos de eventos. Decidir como considerar teoricamente eventos é então um problema ontológico básico e prévio. Os eventos podem ser considerados como particulares básicos concretos do mundo, inalisáveis, ou particulares estruturados em propriedades, objectos, tempos. No último caso um evento seria uma instanciação de uma propriedade por um objecto num tempo t. O fisicalismo dos tipos, que é a maneira natural de formular a teoria da identidade, afirma “Tipos de eventos mentais são tipos de eventos físicos”, considerando que um evento é uma instanciação de uma propriedade por um objecto num tempo t; um evento mental é a instanciação de uma propriedade mental por um objecto num tempo t; um evento físico a instanciação de uma propriedade física por um objecto num tempo t. A identidade deve ser pensada aqui.

da organização funcional, organização essa que corresponde à descrição informacional-computacional do sistema físico, que é uma descrição do que o sistema faz (mas não temos que afirmar que sabe conscientemente que faz, ou que se faz fazer voluntariamente). Este é o lugar filosófico da linguística generativa: ela visa um módulo<sup>486</sup> da mente, um módulo inato e autónomo em relação ao resto da cognição, chamado *Language Acquisition Device*. Há portanto uma faculdade de linguagem cujo estado inicial seria teorizado pela Gramática Universal e cujo estado final mais ou menos rígido, fixado, num humano adulto seria teorizado pelas gramáticas das línguas que modelizariam uma Língua-I, tacitamente conhecidas (*cognized*) pelo falante. Há portanto um subsistema do cérebro para a linguagem, um órgão-mental que é um órgão-programa, em grande parte inatamente determinado, sendo a aquisição de linguagem um processo de determinação de algo já presente, por exemplo dos valores dos parâmetros (princípios abertos que só são ‘ligados’ com o desenvolvimento da Faculdade de Linguagem), porque, como Chomsky diz, “*Language [embora seja inata, wired-in] is a network, not completely wired-up at birth*”<sup>487</sup>.

Esta é a mente computacional estudada pela linguística. Vou esquematizar agora o problema mais geral da Teoria Representacional-Computacional da Mente (TRCM), cuja necessidade em linguística Chomsky mostrou, tal como ele aparece aos filósofos. Vou tomar aqui como guia um filósofo e psicólogo americano que trabalhou com Chomsky no MIT nos inícios dos anos 60, Jerry Fodor.

O ponto de partida da teoria é o facto de parecerem existir estados cuja inserção é problemática numa psicologia materialista, i.e. na teoria da mente que parte do estudo do cérebro como sistema físico. Esses estados são os estados mentais conscientes, intencionais e racionais. No entanto a inserção não pode deixar de ser feita, uma vez que se está a fazer ciência natural, i.e. não se pode simplesmente pressupor que os estados mentais são conscientes, intencionais e racionais sem se os inserir num mundo cuja descrição mais básica cabe à física. Considerar que não há propriedades mentais sem propriedades físicas, i.e. que não há objectos puramente mentais, é o primeiro mandamento do fisicalismo na teoria da mente.

O facto é que nós nos descrevemos e explicamos uns aos outros (e a nós próprios) em termos mentalistas (quando dizemos por exemplo: ele fez A porque desejava que B e pensava que se A então B, por exemplo: “o João levantou a mão porque queria agarrar o copo e pensou que se levantasse a mão então poderia agarrar o copo”). A isto chama-se *folk psychology*,

<sup>486</sup> Um módulo é um sistema informacionalmente isolado.

<sup>487</sup> Chomsky 1994, in Dancy 1994.

psicologia do senso comum e na maior parte das vezes não temos sequer que a formular explicitamente, seria até absurdo que o fizéssemos (como é absurdo no caso do exemplo que dei). É claro que nós compreendemos imediatamente os humanos e as suas acções voluntárias, racionais e com sentido, compreendemos movimentos físicos de corpos humanos mediante estas descrições/explicações mentalistas (nós somos animais mentalistas, aliás estaremos mal se o não formos: os psicólogos falam por exemplo do autismo como *mindblindness*; a não-prática do mentalismo é um défice cognitivo, uma patologia).

A *folk psychology* funciona, nós somos bons a praticá-la, mas isso não garante de modo nenhum o estatuto científico destas atribuições, porque o que acontece é que o comportamento das pessoas, que são sistemas físicos, biológicos, está a ser explicado, previsto e compreendido por referência aos conteúdos das suas crenças e desejos, mas “conteúdo” não é uma noção da física ou da biologia, parece antes uma noção semântica. Daniel Dennett diria que uma máquina semântica é uma impossibilidade física - muito embora nós nos tratemos *como se* fossemos máquinas semânticas, i.e. “movidos a pensamentos, a conteúdos” nada no mundo se move por causa do conteúdo, nem sequer as partes do nosso cérebro como parte do mundo (em última instância, as células do nosso cérebro “agirão” sempre como células, por exemplo degenerando com doença de Alzheimer, sem quererem “saber” do que pensamos com elas).

Entre os poderes causais desses estados com conteúdo parece estar o poder de causar outros estados com conteúdo, no que chamamos inferências, movimentos de pensamento. Estas inferências são (mais ou menos) racionais e responsáveis pelos nossos comportamentos. Então, coisas semanticamente avaliáveis, coisas que podem ser verdadeiras e falsas, fazem coisas acontecer, comportamentos, e inclusive movimentos de pensamento. Mas como é que uma explicação assim dos sistemas físicos que nós somos—uma explicação que supõe estados semanticamente avaliáveis com poderes causais—pode funcionar? Como nota Fodor<sup>488</sup>, as proposições são semanticamente avaliáveis, mas não supomos que tenham poderes causais, elas são entidades abstractas, por outro lado um gato ou uma mesa têm poderes causais mas não são semanticamente avaliáveis...O mental é esta parte do mundo para cuja teoria precisamos de ambas as coisas - ser semanticamente avaliável e ter poderes causais (esclareça-se que o problema só ganha esta forma numa teoria materialista da mente—também poderíamos dizer que pensamento e vontade não são físicos - mas o facto de nos termos envolvido num inquérito naturalista quanto à mente traz certas obrigações).

<sup>488</sup> Fodor 1994, “Fodor”.

Como diz Fodor, “Aí estava o problema até Alan Turing ter tido o que eu suponho que foi o melhor pensamento acerca de como a mente funciona que alguém teve até agora”<sup>489</sup>. Turing pensou numa máquina manipuladora de símbolos cujas mudanças de estado fossem provocadas pelas propriedades materiais dos símbolos sobre os quais operam – historicamente essas Máquinas de Turing, que eram objectos abstractos, e especificamente as Máquinas de Turing Universais, uma vez fisicamente realizadas, vieram a ser os computadores. A ideia de Turing subjaz à intuição cognitivista de que tenho vindo a falar. Ela dá conteúdo teórico ao nível autónomo da explicação psicológica (por oposição a neurofisiológica ou outra) numa teoria naturalista da cognição. Com ela, poderemos ter uma teoria da implementação para leis intencionais, poderemos ter uma psicologia materialista. Daí o facto de Alan Turing ser um dos ‘heróis’ das ciências cognitivas e o facto de o funcionalismo ter sido originalmente formulado por Putnam em termos de Máquinas de Turing<sup>490</sup>.

As hipóteses do funcionalismo, representacionismo e computacionalismo esclarecem o alcance epistemológico da teoria naturalista de Chomsky e da Teoria Representacional Computacional da Mente em geral. A investigação de Chomsky em gramática generativa é então filosoficamente um funcionalismo representacionista e computacionalista que tem como objectivo fazer ciência natural de um módulo da mente humana, um módulo informacionalmente isolado às crenças e aos seus conteúdos conscientes. Esse problema da linguagem é um sub-problema específico do problema filosófico do estatuto de qualquer psicologia materialista, que é o problema de Jerry Fodor. Aliás, a ideia de modularidade, que insere segundo Chomsky a faculdade de linguagem na mente-cérebro (a faculdade de linguagem é um módulo da mente) é outra das ideias filosoficamente importantes do cognitivismo, e faz-nos considerar o cérebro como naturalmente especializado, como um *team* de especialistas, trabalhando isoladamente e compartimentadamente nas suas tarefas, daí que outro problema filosófico que aqui surge seja o da unidade da vida mental sustentada fisicamente por um cérebro num eu-consciente. Tudo nos diz que é estranho que tenhamos uma vida mental consciente unificada: sendo o cérebro como é, uma associação de agências independentes, poderíamos estar sempre à beira da disrupção. Que sejamos unificados, que sejamos *uma* pessoa consciente, é outro dos problemas da filosofia da mente.

Em suma, Chomsky trouxe a teoria do mental “para dentro”. Podemos chamar a isso mentalismo por oposição ao behaviorismo dominante

<sup>489</sup> Fodor 1994, p. 296.

<sup>490</sup> Hoje será mais usual caracterizar o funcionalismo em termos de papeis causais-funcionais.



na psicologia e filosofia americanas até aos anos 50/60. Mas o mentalismo (funcionalista e por isso materialista) característico das ciências cognitivas não envolve qualquer apelo a estruturas estranhas ao mundo físico. I.e. embora se fale do interior não há aqui nenhuma interioridade imaterial, nenhuma *res cogitans* cartesiana, identificável com uma alma. Isto embora a expressão ‘*Cartesian Linguistics*’, utilizada por Chomsky para indicar o interesse por esse interior, seja ambígua. Mas o que pessoas como Chomsky e Fodor vêem de bom no cartesianismo é o reconhecimento de estados mentais e a livre atribuição a estes de poderes causais. Isto é, o que eles partilham com Descartes é uma repugnância pela ideia empirista de tábua rasa para falar do mental.

A ideia de tábua rasa é hoje, aliás, no mínimo cientificamente ignorante dada a sofisticada especialização que se encontra num cérebro humano. A originalidade de Chomsky como um dos pais da revolução cognitiva foi ter dado uma teoria para a especificidade desse inatismo biologicamente incontornável, admitindo estados da mente-cérebro que sustentam algo de semelhante à dedução, i.e., computações de representações abstractas segundo regras. Esses estados e processos da mente-cérebro seriam o objecto da linguística. A este objecto pode-se chamar a mente computacional, logo aquilo que se procura caracterizar é uma máquina simbólica física: um humano, ou pelo menos a sua faculdade de linguagem, é, neste quadro, uma máquina simbólica física.

## **2. Alguns problemas da chegada da estratégia da objectividade ao lugar físico da subjectividade, o cérebro**

A proposta de Chomsky, a que chamei cognitivismo simbólico, é a proposta de uma naturalização –estudo naturalista—da significação via a dimensão formal, algorítmica, dos fenómenos de linguagem. A ideia é ver as linguagens naturais à imagem das linguagens formais e considerar que se está a falar do cérebro, ao nível representacional-computacional.

Para a filosofia da mente e a filosofia da ciência este é o lugar do problema: estamos perante a chegada da ‘estratégia da objectividade’ (na terminologia do filósofo Thomas Nagel<sup>491</sup>), que começou com a física, no século 17, à mente-cérebro como parte da natureza. É óbvio que aquilo que as ciências cognitivas hoje fazem é física, biologia, matemática, da mente-cérebro: teorias como a teoria matemática dos sistemas dinâmicos, a física quântica, a neurobiologia são hoje todas convocadas para o estudo da cognição.

<sup>491</sup> A objectividade é, para Nagel, um método do entendimento; apenas secundariamente podemos chamar objectivas às verdades assim atingidas (cf. Nagel 1986, p. 4).

Em suma, o círculo fechou-se. O pensador estuda as condições físicas, naturais, do seu próprio pensamento com a mesma estratégia bem sucedida que tem sido usada em relação ao resto da natureza. Só que até agora ficava sempre de fora qualquer coisa, a que se chamava o sujeito, o observador, a alma, e que seria o autor do pensamento. Ainda assim, como T. Nagel nota sempre, nós não podemos pensar o pensamento totalmente a partir de fora, como se não existissemos e não se tratasse de nós. Nem na física, nem numa teoria física de um sistema cognitivo e consciente: mesmo a física – o melhor exemplo de *View From Nowhere* para Nagel – é pensamento humano e não uma identificação com a realidade em-si. Sendo teoria, é feita em primeira pessoa, mesmo que se trate de uma primeira pessoa des-psicologizada.

Será então sequer pensável uma naturalização total do pensamento, da mente enquanto sujeito pessoal, consciente, sujeito de acção voluntária, sujeito de ciência? Será que é possível uma teoria da primeira pessoa em terceira pessoa (como todas as teorias científicas), uma ‘teoria externa do mental’ como Thomas Nagel considera ser por exemplo o funcionalismo de que atrás se falava?

Vou concretizar o tipo de restos e impedimentos que podem aparecer aqui, *malgré nous*, i.e. mesmo dada a aceitação sem problemas de um fiscalismo mínimo, mesmo aceitando que não há propriedades mentais sem propriedades físicas.

Voltando a Chomsky, o facto de o ponto de referência da gramática generativa ser o indivíduo isolado coloca o problema a que o próprio Chomsky chama “o problema de Wittgenstein”<sup>492</sup> e que considera ser a mais interessante crítica ao enquadramento conceptual da gramática generativa. Segundo Chomsky, é possível caracterizar como regras os processos da mente-cérebro de um indivíduo isolado. Os Princípios e os Parâmetros podem ser considerados como regras neste sentido. É suposto então que o indivíduo está a seguir regras “privadamente”. Ora, Wittgenstein é conhecido por ter contestado esta possibilidade. A ideia de seguir-regras é uma peça fundamental da teoria wittgensteiniana da compreensão e da linguagem, e por isso é importante compreender as razões da posição de Wittgenstein quando se faz teoria da mente e da linguagem.

Ele diz no parágrafo 202 das *Investigações Filosóficas* que “Pensar que se está a obedecer a uma regra não é obedecer a uma regra”. E conviria ler esta afirmação como: Não é por se pensar que se está a obedecer a uma regra que se está a obedecer a uma regra. I.e., o ser-regra da regra, o funcionar-come-regra não é uma questão ao nível do indivíduo, da sua

<sup>492</sup> Cf. Chomsky 1986 [ed. port., p. 217, *Questões sobre regras*].

voluntariedade ou consciência: “Portanto não é possível obedecer a uma regra privadamente. Caso contrário, pensar que se estava a obedecer a uma regra seria o mesmo que estar a obedecer a essa regra”. É essencial ao cognitivismo simbólico considerar uma analogia entre os mecanismos cerebrais cognitivos num humano e o funcionamento de um programa numa máquina simbólica que pode ser expressa pela noção “seguir-regras”. Mas Wittgenstein não poderia aceitar essa analogia.

Os filósofos desenvolveram uma verdadeira indústria em torno deste problema de seguir-regras (e do problema que o acompanha, o problema da linguagem privada), como entradas no problema do que é significar e compreender. Constituem marcos centrais desta discussão as próprias *Investigações Filosóficas* e o livro do lógico e filósofo Saul Kripke, *Wittgenstein on Rules and Private Language*, de 1980<sup>493</sup>. Esta “indústria” deve-se ao facto de este não ser apenas um problema de teoria da linguagem, mas um problema de teoria do conhecimento e da ciência, da possibilidade do pensamento em geral.

Para começar, para Wittgenstein significar e compreender não podem ser considerados como estados ou processos, quando pressupomos que o são já decidimos tudo e decidimos mal. Mas, como vimos, os cognitivistas necessitam exactamente de estados e processos mentais.

Afinal o que é seguir uma regra? Ou melhor, a verdadeira questão é: será que “regra” é o tipo de conceito que pode ser usado numa explicação naturalista, em terceira pessoa?<sup>494</sup> Se a minha observância de regras é cega como Wittgenstein pensa ser o caso (cf. §219 das *Investigações Filosóficas*) e Chomsky concede, se por exemplo no uso de conceitos ou no uso de uma palavra para nomear vermelho, seguir uma regra não é um saber que eu possuo e sou capaz de explicitar, mas antes uma prática impenetrável ao (suposto) agente, se a compreensão de “vermelho” não é um estado mental consciente prévio, com uma dada duração, do qual brotam as aplicações linguísticas, se não há o processo abstracto “regra” além da prática observada, que razões restam para afirmar que é uma regra aquilo que está a acontecer fisicamente no cérebro do indivíduo?

Para Wittgenstein ao contrário da pretensão implícita na atribuição da observância privada de regras, as atribuições de observância de regras a indivíduos (para descrever a significação e a compreensão) são legítimas apenas no contexto de uma comunalidade de respostas e práticas numa comunidade. O ponto é claro: “regra” é um conceito para uma descrição

<sup>493</sup> Por acaso Kripke pensa que no citado § 202 o problema da linguagem privada está já resolvido. Cf. Zilhão 1993.

<sup>494</sup> Cf. Zilhão 1993, p.172.

exteriorizada e social do mental, para uma descrição do pensar e do falar como um fazer, e não para uma descrição individualista e mentalista do que se passa num cérebro. Mas, como se sabe, Chomsky diz assumidamente que o seu conceito de linguagem é internalista, individualista, diz respeito a estados do cérebro de um indivíduo, a sua teoria da linguagem é um mentalismo em terceira pessoa<sup>495</sup>.

Pode-se mesmo dizer que fazer asserções factuais acerca de regras que estão a ser seguidas sem serem conscientes é, dado o compromisso naturalista, a própria finalidade da gramática generativa e sabemos que Chomsky pensa que a resposta comunitarista wittgensteiniana não acrescenta nunca nada ao conhecimento acerca dos “factos acerca da mente-cérebro de Jones” que interessam o linguista.

Possivelmente os aspectos da linguagem que estão a ser teorizados por Chomsky e Wittgenstein são diferentes. Chomsky fala da fisicalidade da sintaxe, Wittgenstein persegue o problema “como é que alguma coisa pode significar alguma coisa” e é conduzido à externalização do problema semântico. Mas será que mesmo a sintaxe cuja implementação Chomsky procura descrever de modo funcionalista é uma noção física? Não será o facto de falar de “regras que estão a ser seguidas privadamente” mais um dos aspectos da metáfora do homúnculo que John Searle considera endémica no cognitivismo<sup>496</sup>?

O que é que há aí de metafórico? Metafórico é o facto de se descrever como se fosse em terceira pessoa aquilo que só é “visível”, só existe, mediante a presença de um observador (e um observador é coisa que obviamente não está lá, no cérebro, logo não poderia haver regras ou símbolos como há neurónios e iões). Para Searle, assim como a sintaxe não é suficiente para a semântica (e o ponto a célebre experiência mental do Quarto Chinês é desmentir este pressuposto cognitivista), a física não é suficiente para a sintaxe (pois é preciso um observador para considerar um processo físico como sintáctico).

Em suma, a legitimidade ou não do conceito de regra para processos físicos na mente-cérebro do indivíduo funcionalistamente caracterizada e inacessíveis à consciência é um conflito nuclear na fundamentação conceptual do programa generativista (o que que pensaríamos do neurocientista que dissesse que as bombas iónicas na membrana dos neurónios “seguem regras” deixando passar selectivamente iões de sódio e de cálcio?). Searle diria que em termos de cognição há o cérebro, há a consciência, mas não há nenhum nível algorítmico intermédio a não ser que não queiramos

<sup>495</sup> Cf. Chomsky 1994, in Dancy 1994.

<sup>496</sup> Cf. Searle 1998, *Crítica da Razão Cognitiva*, p. 229.

abdicar do antropomorfismo. A Teoria Representacional Computacional da Mente seria então um último reduto do antropomorfismo que foi sendo expulso de todo o conhecimento do mundo físico e biológico pela estratégia da objectividade. Quando esta chegou ao lugar suposto do *antropos*, o antropomorfismo ter-se-ia lá finalmente entrincheirado. Todos os termos do cognitivismo como pretensa ciência natural (termos como “símbolo”, “algoritmo”, “representacional”, “computacional”) seriam assim falaciosos. Searle pensa que a ideia de uma sintaxe realizada fisicamente no cérebro, noção sem a qual a ideia cognitivista colapsa, é insustentável pois sintaxe não é uma noção física como massa ou gravidade, que podemos considerar existirem aí, no mundo, independentemente de nós. Alguma coisa só é sintáctica aos olhos de um observador (e portanto os nossos olhos vão ilegitimamente connosco quando “descemos ao cérebro” – onde não há obviamente ninguém—e dizemos: isto é sintáctico). Este é um problema acerca de causas versus normas ou regras na descrição do mental e, como se viu, o problema é o envolvimento de qualquer coisa como consciência ou “sujeito” na noção de regra, por mais que a intenção de Wittgenstein tenha sido dar uma descrição externalista do que é pensar e falar como fazer alguma coisa<sup>497</sup>. Mesmo que esse “sujeito” não seja a interioridade cartesiana – não o é certamente para Wittgenstein nas *Investigações*— uma descrição baseada em regras precisamente não é causal e em terceira pessoa. A noção de regra é então uma noção normativa por oposição a causal (no sentido em que se podemos dizer que na ética e no direito há normas não vamos dizer o mesmo da física ou da neurofisiologia; não o poderíamos fazer legitimamente). E se a noção de regra é necessária para pensar sobre linguagem e significação, podemos argumentar que isso se passa porque é errado pensar na linguagem e na significação em termos totalmente naturalistas.

Repare-se que a origem deste problema do carácter homuncular das descrições cognitivistas é fácil de ser seguida: se um lógico ou um teórico da computação podem usar à vontade noções como sintaxe e semântica, a partir do momento em que a inspiração lógica é transportada para o estudo de um sistema físico, espaço-temporal, cujos parâmetros variam continuamente, o uso dessas noções já tem que ser justificado.

Este é o primeiro problema dos pressupostos filosóficos do programa chomskyano. Vou referir outros dois problemas para a Teoria Representacional Computacional da Mente, que surgiram de imediato perante decisões cognitivistas. Em geral podemos dizer que os *qualia* (as qualidades subjectivas do experienciar conscientemente, ligadas ao *what is it like to be*

<sup>497</sup> Cf. Nagel 1997, *Language*, p.37.

an *x?*, na expressão celebrizada por T.Nagel) e a semântica são os grandes excluídos de uma visão funcionalista / computacionalista / representacionalista da mente. Isto tem vindo a ser diagnosticado a partir da filosofia nas últimas décadas.

Celebres críticas ao funcionalismo são por exemplo as de Thomas Nagel (em Nagel 1974, *What is it like to be a bat?*<sup>498</sup>), onde Nagel defende a irredutibilidade do subjectivo ao funcionalismo fiscalista e a impossibilidade de uma “teoria externa do mental”. A pedra de toque para distinguir BATS de não-BATS (i.e. “beable things” de “non beable things”) é, segundo Nagel, a possibilidade de resposta à pergunta “*what is it like to be x?*”, pergunta que captura a existência de uma espaço em primeira pessoa, mais ou menos sofisticado. E para Nagel essa subjectividade, muito embora as manifestações do mental sejam para nós locais, é um tipo, uma aspecto da realidade cuja base geral não é local, embora só seja acessível em primeira pessoa (*takes one to know one...*). I.e. Nagel pensa que a subjectividade é objectiva, que a consciência é tão básica no mundo como a espaço-temporalidade. Há, de acordo com Nagel, uma conexão necessária entre o mental e o físico, embora as leis desta conexão não possam ser descobertas a priori. Nestas condições, uma teoria exterior do mental seria simplesmente bizarra.

Grande parte da discussão filosófica em torno dos problemas de uma psicologia materialista não atinge directamente a linguística (note-se que por exemplo Chomsky sempre teve muito cuidado a distinguir os *mistérios* – aquilo que é cognitivamente inacessível ao organismo, como no nosso caso seriam a consciência, a vontade livre e mesmo o uso criativo de linguagem (a marca cartesiana do não-autómato) e que são afinal os apanágios tradicionais da mente pensada como alma imaterial — e os *problemas*, empiricamente tratáveis. Obviamente aquilo que a linguística, como qualquer inquérito científico-natural, trata são os problemas e não os mistérios.

Chomsky separa-os, é um facto, mas já está lá instalado, no próprio terreno polémico onde os mistérios podem vir perturbar o nosso tratamento dos problemas. Os filósofos têm a obrigação profissional de considerar esses mistérios, ou mesmo de mostrar que pode ser um desvio de atenção considerá-los como mistérios e não como problemas. Mas também os linguistas de certo modo herdaram os “mistérios” nos seus problemas, via a semântica, a necessidade de explicação do *meaning of meaning* (além de, como já vimos, os herdarem através dos problemas das regras, sintaxe e símbolos e da justificação da sua natureza simbólica e discreta de processos em sistemas físicos).

<sup>498</sup> Cf. Nagel 1979.

A origem nas linguagens formais e no estudo de algoritmos do modelo generativista de mente, ao importar a separabilidade da sintaxe relativamente à semântica, traz desde logo problemas. Para seguir a argumentação de George Lakoff, que como sabem começou na escola chomskyana, se se diz que uma gramática formal é um dispositivo algorítmico cujo output são objectos abstractos e que ela é ao mesmo tempo um modelo da mente, a pergunta que se deve fazer imediatamente é: e onde entra a significação nos objectos abstractos que são as frases ou estados/processos da mente-cérebro?<sup>499</sup>. Como pode a significação advir a objectos formais? Uma solução é dizer que a componente semântica de uma gramática é interpretativa, como num sistema formal. Mas a psicologia não é interpretação de cadeias formais através da teoria dos conjuntos.

Lakoff afirma que a descoberta da ubiquidade da metáfora conceptual nas línguas naturais lhe mostrou que a semântica destas não pode ser baseada em condições de verdade<sup>500</sup>. As condições de verdade são a ideia lógico-filosófica essencial para tratar significação—compreender o significado de uma frase, i.e. o pensamento expresso pela frase é saber as suas condições de verdade e se essas condições são satisfeitas ou não, daí o valor de verdade (V,F) da frase. “Condição de verdade” é portanto a condição que o mundo tem que satisfazer para uma frase ser verdadeira. Lakoff pensa que a compreensão do *meaning of meaning* não poderia ter a ver com uma relação assim directa, abrupta, suposta, entre frases e mundo, na exclusão da consideração do entendimento (incorporado) desse mundo. A ideia de *embodied mind* poderá ter aqui importância para o estudo linguístico do *meaning of meaning*. A mente cognitivista é precisamente des-incorporada, não no sentido imaterialista cartesiano mas no sentido de ser des-biologizada para poder ser considerada como uma realidade de natureza formal. Dada a inspiração formalista ela é uma mente descerebralizada. E, de facto, era esse o compromisso do funcionalismo, para não ser paroquial, recordemos. Mas será razoável que para salvaguardar a possibilidade de o mental ter outras realizações físicas se descorporealize assim o entendimento?

Se bem a compreendo, através de alguns autores que estudei mais de perto, como Ronald Langacker,<sup>501</sup> Leonard Talmy<sup>502</sup> e George Lakoff<sup>503</sup>, e Ray Jackendoff<sup>504</sup>, a linguística cognitiva dá corpo, observada do lado da filosofia, à hipótese segundo a qual o semantismo que procuramos terá que

<sup>499</sup> Cf. Lakoff 1987.

<sup>500</sup> Cf. Baumgartner & Payr 1995, *Speaking Minds – Interviews with 20 Eminent Cognitive Scientists*, p.120.

<sup>501</sup> Langacker 1986.

<sup>502</sup> Talmy 1988.

<sup>503</sup> Lakoff & Johnson 1980 e Lakoff 1987.

<sup>504</sup> Jackendoff 1987 e Jackendoff 1992.

vir da relação entre subjectividade e mundo, relação que é originariamente perceptiva. A intuição da linguística cognitiva é que não existe objectividade imediatamente simbólica, por exemplo não há aí no mundo independentemente de seres humanos e do seu particular entendimento, como diz Lakoff, coisas que sejam “verbos” ou “nomes” ou “preposições”. No ser-verbo do verbo, no ser-nome do nome, no ser-preposição da preposição, no ser-metáfora da metáfora, na generatividade de esquemas de imagem para metaforizar (como *in/out*, *up/down*), são supostas conceptualizações, sínteses subjectivas como diz Langacker. Quer dizer, o “sujeito” está aí, nas categorias de linguagem que encontramos. Pense-se nos casos óbvios dos esquemas de imagem que originam metáforas e preposições (preposições como ‘em’, ‘entre’, ‘de’, ‘desde’, etc). Isto conduz-nos, segundo Lakoff<sup>505</sup>, a constatar que o sujeito está aí perceptivamente na génese da possibilidade de linguagem, pois essas categorias só poderiam existir numa mente incorporada, nunca numa mente des-incorporada.

Ou, para dar outro exemplo, o que está Ray Jackendoff a fazer quando defende a necessidade de uma Estrutura Conceptual na teoria da linguagem, um nível de representação que codifica a significação para além dos níveis para a fonologia e a sintaxe? Ele está a supor que é o mundo fenomenológico morfologicamente estruturado dos modelos mentais que é intencionado nas nossas enunciações e não um “mundo real” abstractamente definido. Noutras palavras, o mundo em que vivemos linguisticamente é uma interface fenomenológica definida e não “o mundo da física”.

Jackendoff parte aliás do princípio de que a primeira obrigação que nos traz o facto de subscrevermos – bem – uma teoria modular do mental é discernir os princípios de correspondência que nos permitem por exemplo falarmos daquilo que vemos. Esta é uma possibilidade estranha, dada a especialização e separação das tarefas da visão e da linguagem no cérebro, i.e. dada a modularidade da mente. Uma pista-protótipo para Jackendoff seria por exemplo esta. Não existe nada de intrínseco no mundo que requeira que, linguisticamente, a posse e a adscrição de propriedades estejam mentalmente representadas num sistema isomórfico da representação de estados e eventos espaciais, mas é isso que se passa e que Jackendoff documenta com a análise dos verbos categorizáveis como GO, BE, STAY, ao longo de vários campos semânticos – por exemplo verbos como: *give*, *buy from*, *inherit/ belong*, *own/ keep*, *retain* (verbos de posse), cairiam respectivamente sob essas categorias<sup>506</sup>. As características da Estrutura Conceptual reveladas por este tipo de análise são uma prova da *embodied mind*.

<sup>505</sup> Cf. Lakoff e Johnson 1980 e Lakoff 1987.

<sup>506</sup> Cf. Jackendoff, 1987, p.152, *Generalization of Spatial Concepts to Abstract Domains*, para a análise de verbos GO, BE, STAY.



Podemos, respondendo à pergunta inicial, falar daquilo que vemos porque a mente é *embodied*.

Quando Ronald Langacker pretende substituir a noção de cálculo ininterpretado pela noção de *imagery* para falar de gramática e mesmo das intuições de gramaticalidade<sup>507</sup> nega explicitamente que se possa tomar simplesmente como garantidos o assemantismo da sintaxe das línguas naturais e a indefinibilidade de unidades simbólicas básicas, como predicções nominais (nomes) e predicções relacionais (como verbos e preposições).

Quando George Lakoff localiza os esquemas de imagem “generativamente” disponíveis para a produção da metaforicidade, pervasiva nas línguas naturais (esquemas como dentro/fora, cima/baixo, etc), ele considera que aponta para uma concepção de linguagem como pensamento imagem-esquemático. Quando Leonard Talmy evidencia os fenómenos da categoria a que chama “dinâmica da força” num grande corpus linguístico, ele pretende ter isolado uma categoria semântica que gramaticaliza informação topológica e dinâmica, i.e. geométrica e física.

O problema da intencionalidade ou *aboutness* joga-se aqui. Pense-se no facto seguinte: aquilo que é linguisticamente intencionado com uma enunciação como por exemplo “*Está uma bola em cima da mesa*” não é o mundo segundo a física, nem a mecânica newtoniana nem a física quântica nem a física relativista, mas antes um mundo de modelos mentais ao nível de uma interface fenomenológica específica. A intencionalidade psicológica num humano não é a interpretação de cadeias formais via teoria dos conjuntos e a semântica linguística tem que considerar este facto. Os linguistas cognitivos fazem-no via o que pode ser chamado uma teoria da génese perceptiva das categorias de linguagem, procurando a origem da significação linguística no entendimento sub-linguístico do mundo.

Se a linguística é como Chomsky quis, uma ciência cognitiva, a primeira coisa a considerar é que toda a ciência é uma ciência da natureza. Logo, não há razão para estudar os produtos da mente humana de um modo diferente daquele que se utiliza para o resto da natureza. Esta é aliás uma lição de Chomsky. Mas então, como diz Jean Petitot, um matemático francês, a orientação simbólica do cognitivismo revela-se incompatível com este naturalismo pois não existem formas simbólicas na natureza interna (i.e. no cérebro) ou externa. Não podem senão existir formas geométricas e dinâmicas, daí que toda a naturalização da mente, da linguagem e do sentido deva segundo Petitot conduzir a uma revolução na concepção do formal e dos formalismo envolvidos na existência de cognição, e

<sup>507</sup> Cf. Langacker 1986.

consequentemente a um afastamento da sua inspiração inicial na lógica e do apoio nos modelos discretos<sup>508</sup>.

### 3. Semântica: sugestões do trabalho de filósofos

Sob o título de teleosemântica os filósofos lidam com um problema parente do problema da linguística cognitiva (que é o problema de não podermos supor uma objectividade directamente simbólica no estudo naturalista da linguagem, o que impõe a necessidade de ir mais atrás, ao entendimento sub-linguístico do mundo pelo organismo para procurar a significatividade da significação). Se levarmos a sério o facto de não começar a haver significação apenas linguisticamente (devemos por exemplo poder afirmar que animais têm estados mentais intencionais e podem ser enganados e tentar enganar), veremos uma série de erros conceptuais a que somos conduzidos por considerarmos como pedra de toque na teoria da significação os estados mentais conscientes, linguísticos, humanos.

Podemos resumir o problema aqui em jogo perguntando quem se sente persuadido pela experiência mental do Quarto Chinês de John Searle<sup>509</sup>. Penso que quem trabalha com linguagem deve considerar esta experiência mental pelo menos uma vez. Quem se sente persuadido pela experiência mental de Searle, considera a consciência como dado inabdicável no pensamento da significação e não aprovará a abordagem teleosemântica, pois esta pretende precisamente desligar o tratamento da intencionalidade (*aboutness*) do tratamento da questão da consciência, considerando que o mental é uma questão de graus e não de tudo ou nada (nesse caso o “tudo” seria determinado pela presença da consciência).

A teleosemântica<sup>510</sup> parte da constatação de que a *aboutness* pode ser cumprida sem ser transparente aos organismos que albergam os “ícones intencionais” (um termo da filósofa americana Ruth Millikan). No exemplo fetiche de Millikan, o exemplo da dança-de-abelhas-que-indica-a-outras-abelhas-onde-está-o-nectar, existem todos os componentes essenciais do significar (*ofness*, Função Própria, produtores e intérpretes em cooperação projectada, e regras de *mapping* em condições Normais) e não existe consciência. Ou antes, não temos que considerar a consciência

<sup>508</sup> Cf. Miguens 1995, p. 424. É claro que estas críticas se ligam também com o conexionismo.

<sup>509</sup> Cf. Searle 1980, *Minds, Brains and Programs*, in *Behavioral and Brain Sciences*, 3, 417-457. Um humano que não compreende uma palavra de chinês (o próprio Searle) está fechado num quarto, rodeado de cestos cheios de símbolos chineses, e com um livro de regras para a combinação e manipulação formal desses símbolos. Em “resposta” às cadeias de símbolos que vêm de fora do quarto, e seguindo as regras do manual, saem do quarto cadeias de símbolos que são tomadas como respostas. O quarto “fala chinês”, mas Searle não percebe uma palavra. Para Searle, este entendimento do sistema-quarto é “como-se”, é intencionalidade atribuída e não intrínseca. O ponto será que a sintaxe não é suficiente para a semântica.

<sup>510</sup> Para uma apresentação mais detalhada dos problemas da teleosemântica, cf. Miguens 1998.

para descrever o conteúdo pois o “Conteúdo” desse Ícone Intencional é descrito dizendo que característica deve existir no ambiente à qual o ícone se aplicaria de acordo com determinadas regras estando o seu intérprete a usá-lo de maneira normal.

Esta constatação de que a *aboutness* pode ser cumprida sem ser transparente aos organismos envolvidos deve-se alargar ao nosso próprio uso significativo de linguagens naturais (o que impedirá, segundo Millikan, o racionalismo da significação, que seria o “3º dogma do empirismo”<sup>511</sup>). Metodologicamente, a ideia é olhar para aquilo que está “abaixo do humano”, i.e., para casos mais simples de *aboutness*, etológicos ou mesmo tecnológicos, à procura da proto-representação, da proto-significação e das suas condições.

Isso far-nos-á ver os erros intelectualistas inerentes a uma abordagem “*top down*” da significação (*top-down* no sentido metafórico de ser feita a partir da significação humana consciente para os outros casos, e não no sentido mais usual de uma abordagem feita, por exemplo, a partir das funções para os neurónios). O erro geral aqui será a demasiada especificação, por projecção daquilo que são as características humanas do significar, a que somos frequentemente conduzidos, e a não visão da indeterminação como condição de possibilidade do significar. Isto é, em suma, o que Dennett e Millikan resumem sob a ideia de “racionalismo da significação” (a tese segundo a qual saber que/o que quero-dizer/penso é um saber a priori determinado ou determinável. Por isso eu não poderia estar enganado acerca daquilo que quero-dizer mesmo que falhe a referir ou a conhecer).

Mas o racionalismo da significação só se sustentaria se nós fossemos *unmeant meaners*, o que não é o caso. Segundo a teleosemântica, o problema da teoria da mente ou psicologia (que como ciência natural é parte da biologia—no caso humano pelo menos) é o prolongamento do problema biológico da função. O “conteúdo”, o semantismo, dos nossos estados mentais é uma *função da função*, i.e., da função biológica de organismos, e não um facto a mais. E para Dennett assim como a função biológica não é qualquer coisa de determinado na sua origem (peças de organismos que não foram seleccionadas para uma função *f* podem ser recrutadas para desempenhá-la), o conteúdo, que é uma função da função, também não é qualquer coisa de determinado na sua origem.

Concretizando: se items como frases têm significação para nós, se nós podemos de certo modo “dar” significação a items físicos como marcas ou sons, é preciso saber como temos essa capacidade, qual é a sua origem. Não temos razão suficiente, de um ponto de vista biológico, para

<sup>511</sup> Numa alusão ao famoso ensaio *Two Dogmas of Empiricism*, de Quine [cf. Quine, 1953, *From a Logical Point of View*, Cambridge MA, Harvard UP].

considerar que o fazemos a partir de zero (e da consciência...) sendo assim *Unmeant Meaners*, na expressão de Dennett, i.e. ‘Intencionadores Não Intencionados’, intencionando *ex nihilo*.

Para Dennett nós não somos *Unmeant Meaners* porque não somos auto-criadores (na medida em que somos organismos), e vermo-nos, enquanto organismos, como uma “máquina de sobrevivência para genes”<sup>512</sup> faz ver o absurdo da posição segundo a qual nenhum artefacto pode ter intencionalidade real ou genuína mas apenas intencionalidade derivada do seu criador. Esta posição corresponde, note-se, à aceitação da intuição à qual faz apelo o Quarto Chinês de Searle. Nós somos esses “artefactos”, na medida em que estamos aqui por desenvolvimento orgânico e não por um *fiat* de que seríamos os autores.

Se a possibilidade de intencionar, e a possibilidade de *mis-take* que acompanha o *take* do intencionar não surgem só com a significação linguística (pense-se por exemplo na dança das abelhas de Millikan: pode haver “erro” no seu funcionamento, se por exemplo retirarmos as flores do local indicado às outras abelhas pela dança), é uma espécie de erro intelectualista esperar que o esclarecimento da significação se faça exclusivamente a partir de cima, da significação linguística de que só os humanos são capazes. Então como ficamos quanto à teoria da linguagem e da mente segundo as directrizes gerais da teleosemântica de Millikan (ou Dennett)<sup>513</sup>?

1- Aceitaremos que a significação não surge *ex nihilo* com a linguagem humana e com a consciência humana. Representar e significar linguisticamente são habilidades que têm “antepassados” nos organismos e especificamente nos cérebros humanos. O cérebro humano não é uma tábua rasa quando a linguagem se instala, mas um órgão altamente especializado em funções cognitivas várias. Há portanto *design* prévio no qual estas habilidades de significar e representar se apoiam e o qual redireccionam. Essas estruturas não são apagadas mesmo quando passa a tratar-se de significação linguística. Há que considerar portanto um apoio da capacidade linguística de significar em habilidades mais básicas e pré-linguísticas como por exemplo identificar, re-identificar, seguir...<sup>514</sup> Neste quadro, a linguagem e o pensamento são categorias naturais, categorias biológicas e portanto, como categorias biológicas, categorias de funções (as coisas que são classificadas sob categorias de funções por serem projectadas – *designed* – para essa função).

<sup>512</sup> Como R. Dawkins propõe no seu livro “O Gene Egoísta”.

<sup>513</sup> Há diferenças profundas entre as posições de Millikan e Dennett; referirei aqui apenas as razões que os fazem partilhar a defesa de uma teleosemântica.

<sup>514</sup> Cf. Miguens 1998, p.501 para o problema da identificação, e para um comentário a um recente artigo de Millikan cobrindo estes temas [A Common Structure for Concepts of Individuals, Stuffs and Real Kinds, More Mama, More Milk, More Mouse] publicado na revista *Behavioral and Brain Sciences* 21-1, 1998.

2 - Veremos o problema da significação em teoria da mente em continuidade com o problema da filosofia da biologia que é o da natureza da função. Assim, como aliás acontece sempre que estão em causa categorias biológicas de funções, verificar-se-á a necessidade de uma noção de normal, e da consideração de uma assimetria entre o normal e o desvio. Normal e desvio são noções comuns na biologia e na medicina (pense-se em corações ou membros normais ou não normais) que nos convêm também na teoria da significação. Isto dá-nos uma teoria da natureza do verdadeiro, muito clássica: uma teoria do verdadeiro como correspondência, *mapping*, e a constatação de que o verdadeiro é “assimétrico” em relação ao falso, i.e., que o verdadeiro é o caso Normal do significar.

3 - A conclusão da teleosemântica é que o externalismo<sup>515</sup> é a única boa posição acerca da significação mas que esse externalismo perturba o racionalismo acerca da significação: *meanings ain't in the head* (como quer Hilary Putnam<sup>516</sup>) mas por isso mesmo eles não são determináveis a priori. Não somos nós que agenciamos a nossa capacidade de significar, ela não nos é naturalmente transparente. Esta remissão externalista ao exterior e à história para a teoria da significação faz com que por exemplo por exemplo, na prática, olhar para padrões neuronais não nos possa dar nunca o que está a ser intencionado. Como diz Dennett frequentemente, não haverá esse tipo de *mindreading*.

O imperativo geral da teleosemântica é portanto o de uma abordagem evolucionista, logo histórica e adaptacionista, das questões da mente e da significação. Como todas as coisas biológicas o pensamento e a linguagem devem ser abordados de um ponto de vista evolucionista, na medida em que a evolução é a luz sob a qual a biologia faz sentido.

Note-se que este “teleofuncionalismo” é diferente do funcionalismo com que comecei: o primeiro – ligado aos nomes de Putnam e Fodor, etc — é uma tese acerca da relação entre hardware e software, o segundo — ligado aos nomes de R. Millikan, D. Dennett, D. Papineau, F. Dretske – é uma tese acerca da relação entre função e finalidade, reintroduzindo portanto a natureza biológica da psicologia e do mental (excluída no funcionalismo-1) como indiciadora da natureza da *aboutness*. Ambas são teses filosóficas acerca da natureza do mental surgidas no pensamento sobre os fundamentos de uma psicologia materialista.

Dennett constata<sup>517</sup> que pensar da maneira naturalista que a teleosemântica indica é muito pouco natural para um filósofo. É aliás pouco natural

<sup>515</sup> O externalismo em filosofia da mente é a posição segundo a qual aquilo que fixa aquilo que é pensado ou dito são aspectos exteriores à mente.

<sup>516</sup> Cf. Putnam 1975.

<sup>517</sup> Cf. Miguens 1998, p. 469.

também para Chomsky, que é um teórico de papel e lápis, que está mais próximo na sua maneira de trabalhar de um matemático do que de um biólogo. Mas o que resulta destas ideias é a necessidade de uma investigação da selecção natural da linguagem natural. Isto implica especular acerca de funções, acerca da história natural da mente e da linguagem, o que é complicado, não só porque o tecido cerebral não fossiliza, mas sobretudo porque se a linguagem é software, descrição de um nível funcional, estaremos a falar de uma ‘arqueologia do software’ (na expressão de Dennett em *Brainchildren*, 1998<sup>518</sup>). É difícil imaginar como poderia tal arqueologia ser praticada—mas isso não justifica uma ocultação da sua necessidade, se é essa necessidade que a teoria nos conduz. Será preciso portanto uma narrativa histórica evolucionista acerca das funções da mente e da linguagem, e esta só pode ser uma explicação *top-down* (i.e. partindo das supostas funções para os seus mecanismos físicos, por exemplo em cérebros ou não veremos literalmente nada—inclusive pela razões exterioristas atrás apontadas. Só se poderá conhecer no sistema nervoso mecanismos e localizações para funções que forem primeiro conhecidas no comportamento – esta é uma intuição comum a Dennett e Millikan).

Aliás, nessa narrativa histórica, nessa história natural da mente e da linguagem, arriscamo-nos a encontrar um dos mistérios de Chomsky: a consciência. Dennett<sup>519</sup> relaciona a presença de linguagem em cérebros humanos com aquilo a que chama a *maquina virtual* para a consciência. O hardware cerebral humano é provavelmente o mesmo desde há milhares de anos, mas esta máquina virtual<sup>520</sup> para a consciência, em grande parte resultante da instalação da linguagem natural em cérebros humanos, seria provavelmente responsável pela constituição de uma *inwardness* natural, progressivamente mais sofisticada, via a possibilidade de autoexortação, importante na passagem de si a si do cérebro que é necessária para essa consciência. Já vimos como a unidade da nossa consciência e identidade pessoal pode ser problemática, dada a modularidade. O abismo teórico entre matéria e *inwardness* de um ser consciente seria transposto assim, por esta narrativa evolucionista. Se uma hipótese deste género fôr defensável, um dos mistérios de Chomsky, e por ser mistério supostamente afastado do estudo empírico, a consciência, poderia vir a revelar estar muito proximamente ligado com os problemas da linguística, i.e., com a instalação e uso da linguagem natural em cérebros humanos.<sup>521</sup>

<sup>518</sup> Dennett 1998, p. 121.

<sup>519</sup> Dennett 1991.

<sup>520</sup> “Máquina virtual” é um termo para uma estrutura feita de regras mais do que materiais (como o programa Word neste computador), um termo que Dennett vai buscar à linguagem de computadores

<sup>521</sup> Segundo Dennett, esta é uma ideia de Nietzsche. Nietzsche afirma que a consciência se desenvolveu sob a pressão da necessidade de comunicação. Cf. Dennett 1991, p.227.

Curiosamente, quando chegamos à consideração evolucionista da faculdade de linguagem chegamos a um dos limites de Chomsky<sup>522</sup>: ele não quer falar de biologia, preferiria falar de física (ou de propriedades físicas de células mas não de evolução<sup>523</sup>) e deixar para depois... Nomeadamente, Chomsky não quer, aparentemente, falar da evolução e selecção natural da linguagem natural. Chega a afirmar que nem sabe como se há-de começar a imaginar como é que a selecção natural poderia ter originado sistemas com asas ou com linguagem. Porque é que Chomsky diz isto? A análise que vou resumir é feita por Dennett em *Darwin's Dangerous Idea* (1995).

Como cientista, Chomsky é um materialista mas por vezes parece ter receio de que, no seguimento do tipo de estudo que iniciou, venha a revelar-se qualquer coisa de pouco digno acerca da mente humana, como por exemplo que as nossas mentes são colecções de *gadgets* maquinais, de “macros”, i.e. de sequências mecanizadas de instruções e tarefas. Esta possibilidade torna-se muito clara numa perspectiva evolucionista-comparativa de tipos de mentes: por exemplo uma neuropsicóloga, Lynn Waterhouse, definiu, perante Dennett, as mentes animais como “quilts (entrelaçamentos) de IRMs (*Innate Release Mechanisms*) e FAPs (*Fixed Action Patterns*)<sup>524</sup>”...

Talvez em parte devido à possibilidade de algo de semelhante poder vir a revelar-se verdadeiro acerca das nossas mentes, Chomsky desconfia bastante da pertinência do evolucionismo e adaptacionismo no estudo da mente e da linguagem. Este é um ponto muito curioso para a filosofia da mente e é preciso afirmar que Chomsky não está sozinho ao pensar que considerações evolucionistas não nos dirão nada sobre o pensamento. Este é um assunto filosoficamente quente, saber se a história evolutiva da mente é ou não é relevante para aquilo que o pensamento pode pensar e para uma teoria da existência e limites do pensamento (por exemplo T. Nagel, em *The View From Nowhere*, dá uma resposta negativa<sup>525</sup>).

Mas porque é que o *Language Acquisition Device* seria diferente de outras partes do mundo biológico, uma vez que Chomsky tão convincentemente o instalou como tal, não sendo resultado de uma evolução por selecção natural? Curiosamente, Chomsky diz muitas vezes que o órgão mental da linguagem é inato, mas nunca diz que ele é resultado da evolução

<sup>522</sup> Cf. Dennett 1995, capítulos 13 e 14 e Miguens 1999a e 1999b.

<sup>523</sup> “Nalgum período remoto uma mutação teve lugar que deu origem à propriedade da infinitude discreta talvez por razões que tenham a ver com a biologia das células, a serem explicadas por mecanismos físicos ainda desconhecidos” [Chomsky 1988, *Language and the Problems of Knowledge, The Managua Lectures*, Cambridge, MIT Press, cit. in Dennett 1995, p. 389].

<sup>524</sup> Cf. Dennett 1991, p. 259.

<sup>525</sup> Nagel 1986, pp.78-82. Para Nagel a existência da razão requer uma explicação não diacrónica [por exemplo evolucionista] mas intemporal. O darwinismo explica a selecção entre possibilidades geradas, não “a possibilidade dessas possibilidades”.

como tudo o que é inato, na exclusão de qualquer imaterialismo. Uma razão pode ser o seu anti-spencerismo<sup>526</sup>. Chomsky alia-se até no imaginário académico, frequentemente, ao paleontologista Stephen Jay Gould como crítico do adaptacionismo (o imperativo dos críticos do adaptacionismo é não ver funções e adaptações sem necessidade e por toda a parte: essa é uma má interpretação da teoria da selecção natural, pois oculta que esta admite *by products* e recrutamentos). Pode-se considerar neste caso que a linguagem seria um *by product* da selecção para qualquer outra coisa, e que portanto considerações funcionais e adaptativas quanto à linguagem seriam especulativas, no sentido pejorativo do termo.

Mas, para Dennett, esta postura de Chomsky relaciona-se antes com um desgosto geral pelas teorias impuras do mental. É como se Chomsky pensasse: para fazer teoria da mente só há dois caminhos possíveis, ou a lógica, a matemática, a física (o caminho “puro”) ou Jane Austen (o caminho “moral”) ... Quem lê Chomsky conhece as suas considerações acerca dos limites da teoria científica naturalista da mente e o seu constante louvor aos romances, ou a Shakespeare. Chomsky pensa que para sempre se aprenderá mais psicologia lendo literatura do que fazendo ciência natural.

Marvin Minsky comentou a certa altura<sup>527</sup>, perante esta bifurcação, que só um professor de humanidades se esqueceria tão ostensivamente da terceira possibilidade: a possibilidade de a psicologia vir a revelar ser, não como a física nem como a literatura, mas como a engenharia. A psicologia (e a biologia) seriam práticas de engenharia invertida (*reverse engineering*). A hipótese de Dennett é que esta perspectiva da teoria da mente-cérebro como engenharia parece vagamente repulsiva a Chomsky (embora em *The Minimalist Program* Chomsky fale explicitamente de soluções de engenharia). Aliás, para reforçar esta interpretação, note-se que Chomsky se tem assumido frequentemente como crítico da IA, que é uma abordagem declaradamente impura, pela engenharia, do mental possível<sup>528</sup>.

Para Dennett, as rejeições da IA e do evolucionismo em teoria da mente têm como origem comum uma intuição quanto ao carácter especial da mente humana consciente, que nunca poderia ser explicável a partir de “baixo” (a partir da física, ou pior, da biologia e da IA). Pelo contrário para Dennett, se as mentes humanas são produtos não miraculosos da evolução

<sup>526</sup> Esta análise é de Daniel Dennett. *Spencerismo*, assim nomeado a partir de H. Spencer, será a consideração segundo a qual toda a complexidade do organismo é função da complexidade do ambiente [corresponderia a um adaptacionismo exagerado]. Note-se no entanto que todo o behaviorismo e todo o empirismo são formas de “spencerismo” – e pode haver boas razões para defendê-los.

<sup>527</sup> Cf. Dennett 1995, p. 387. A situação passou-se num encontro da Society for Philosophy and Psychology na Universidade de Tufts, em 1978.

<sup>528</sup> Por exemplo, Chomsky quase literalmente “perseguiu” Roger Schank, o cientista de computadores, devido à sua descrença quanto às possibilidades de abordagem do entendimento de linguagens naturais pela IA (a área de trabalho de Roger Schank).



elas são artefactos e os seus poderes têm que ter explicações mecânicas. Portanto temos que admitir que nós descendemos de macros (que mais seriam os seres unicelulares antepassados de todos os vivos senão precisamente seqüências de instruções?) e que as nossas mentes são “feitas de macros”. É um facto que nós somos a espécie mais inteligente e a única espécie com linguagem articulada sofisticada. Mas a origem das nossas habilidades mentais não é pura. É pelo contrário muito provável que os mecanismos *wired-in* que constituem hoje as nossas mentes fossem, nos cérebros dos nossos antepassados para outras coisas que não a consciência e a linguagem, tendo depois vindo a ser recrutados para essas funções, e os restos e impurezas ficam. Como Dennett nota, o cérebro pensa melhor e mais depressa quando “pensa” sozinho – por exemplo fazendo-nos ver ou mover—do que quando pensamos voluntariamente, esforçadamente, conscientemente<sup>529</sup>.

Chomsky não está ele próprio interessado em seguir teoricamente a selecção natural da linguagem natural, mas há linguistas, como Steven Pinker<sup>530</sup>, que levam este programa de teoria evolucionista da mente e da linguagem por diante, partindo do princípio de que o *design* do software mental é resultado da evolução biológica e da selecção natural.

Queria só notar que reecontramos aqui uma das condições de que falava no início deste texto. Se o *design* da mente foi seleccionado para resolver problemas de engenharia, quando tratamos a mente segundo estas especificações, damos, como Pinker faz notar, ao mesmo tempo as “*design specs*” (specifications) para construir um robot. Eis-nos de volta ao ponto de partida da exposição das ideias fulcrais das ciências cognitivas: a indistinção entre natural e artificial e o “fantasma” da IA forte (e do seu novo platonismo materialista), que persegue a investigação científica e filosófica das questões da mente.

#### 4. Nota sobre a especificidade do problema filosófico da mente

A minha hipótese até aqui foi que o trabalho dos linguistas e o trabalho dos filósofos têm em comum o facto de ter como objecto a mente. Mas o trabalho dos linguistas, como o de outros cientistas cognitivos, é um trabalho empírico e o dos filósofos não é, nem pretende sê-lo. A questão empírica acerca da mente tem a forma “*Como funciona o cérebro ou qualquer outro dispositivo cognitivo de modo a que a percepção e a acção sejam possíveis no organismo ou outro sistema?*”, enquanto a questão filosófica

<sup>529</sup> Cf. Dennett 1991.

<sup>530</sup> Cf. Pinker 1994.

tem a forma “Qual é a relação entre o mental e o mundo - e especificamente como podem existir o verdadeiro e o falso nessa relação?” A primeira questão pretende uma resposta em terceira pessoa, uma resposta naturalista, a segunda não poderá ter uma resposta em terceira pessoa, pois é argumentável que não poderemos nunca vir a pensar o pensamento a partir de fora<sup>531</sup>.

A primeira questão, a questão empírica acerca da mente, é acerca de categorias psicológicas. Estas dizem respeito à mente como máquina sintáctica: o comportamento do organismo ou sistema tem que ser determinado puramente em termos de noções internas à mente. As representações mentais dos cientistas cognitivos são noções inteiramente não intencionais e não faz qualquer sentido falar da verdade ou falsidade das representações mentais postuladas pelos teóricos para a linguagem, a visão, etc. Elas descrevem um espaço combinatório estruturado, em vários níveis, desde periféricos a centrais, espaço esse que será “estimulado” via transdução a partir dos órgãos sensoriais do organismo. É esta situação que justifica o solipsismo metodológico como método de investigação em teoria da mente<sup>532</sup> ou mesmo a afirmação “o cérebro é cego, surdo e mudo”<sup>533</sup> (pois o cérebro não tem de facto acesso directo ao mundo físico externo ao organismo nem o que nele é representado é o mundo exterior em si na riqueza das suas determinações).

Especificamente, o trabalho dos linguistas é sobre representações mentais neste sentido de categorias psicológicas, presentes com a linguagem natural na mente-cérebro dos humanos. Aquilo que se procura é, como Chomsky diria, uma descrição computacional-representacional da maneira como o cérebro resolveu problemas de engenharia (acrescente-se: através da evolução) constituindo assim esta matéria “formal” da mente-cérebro natural humana a ser usada em pensamentos.

Mas a investigação empírica dos mecanismos da linguagem como módulo da mente não ocupa o mesmo lugar que a teoria filosófica. Esta terá que cobrir pelo menos dois outros aspectos: ela será por um lado, uma investigação epistemológica das condições de possibilidade das ciências do mental, a qual estará ligada a uma metafísica da mente, nomeadamente a uma metafísica da consciência (foi sobretudo da investigação epistemológica

<sup>531</sup> Como nota Jackendoff [cf. Jackendoff 1992, p.158], certos pensadores constituem um lugar privilegiado de observação desta tensão entre psicologia e filosofia na teoria da mente: o exemplo de Jackendoff é Jerry Fodor. Como diz Fodor, o problema é que verdade, referência, etc, não são categorias psicológicas. Serão modos do Dasein?, semi-ironiza Fodor. Cf. Fodor 1980, *Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology, Behavioral and Brain Sciences* 3, 63-73.

<sup>532</sup> Cf. Fodor 1980.

<sup>533</sup> Cf. Jackendoff 1992.

que falei até aqui), e por outro lado, dedicar-se-á àquilo que é indiciado no caso do indivíduo isolado pelo que Chomsky chama os *systems of thought* como sistemas de performance<sup>534</sup>, i.e. a uma teoria do pensamento em acto, envolvendo a explicitação da possibilidade de existência de significação e verdade. Isto porque, se o solipsismo metodológico é acertado num estudo naturalista ele não permite no entanto, o tratamento de noções semânticas: as noções semânticas não são categorias psicológicas e, ao contrário destas não podem ser tratadas de forma solipsista<sup>535</sup>.

Vou procurar mostrar alguma coisa acerca deste segundo aspecto através de algumas ideias sobre significação e verdade do filósofo britânico contemporâneo Michael Dummett, ideias que evidenciam uma grande influência de Frege e do segundo Wittgenstein (conviria mais, talvez, falar de apropriação...), culminando numa posição anti-realista mas não solipsista.

Até aqui tenho passado por alguns problemas colocados pela natureza da linguagem e tem sido aparente que a questão da linguagem não é só uma questão para linguistas; ela é inabdicável na filosofia e especificamente na filosofia analítica. Mas qual é exactamente o lugar da linguagem no estudo do pensamento? Para Dummett, a ideia da filosofia analítica desde os seus inícios com Frege, Russell e Wittgenstein é que o pensamento só é analisável através da linguagem. Dummett fala mesmo do 'teorema fundamental da filosofia analítica'<sup>536</sup>, que seria a afirmação da prioridade da linguagem sobre o pensamento. A tese da prioridade erguer-se-ia contra a ingenuidade da presunção da acessibilidade directa (e solipsista) do pensamento. Como afirma Dummett<sup>537</sup>, a linguagem pode ser um espelho distorcido, mas é o único que temos. A tese da prioridade corresponde portanto a uma des-psicologização, a uma extrusão do pensamento em relação à mente individual<sup>538</sup>; para Dummett é sobretudo devido ao carácter 'comunitário' da linguagem<sup>539</sup> que deve ser defendida a tese da prioridade.

A rejeição do psicologismo (que é aliás comum à filosofia analítica e à fenomenologia, e resultou em parte da necessidade histórica que a filosofia teve de precisar a sua diferença em relação à psicologia nascente no princípio do século XX) consiste em considerar que os pensamentos não devem ser analisados em termos das operações mentais dos indivíduos

<sup>534</sup> Cf. Chomsky 1995.

<sup>535</sup> *A verdade ou a significação não são teorizáveis no âmbito de um cérebro humano isolado*

<sup>536</sup> Dummett 1997, p.37.

<sup>537</sup> Dummett 1993, p. 6.

<sup>538</sup> Cf. Dummett 1993, *Origins of Analytic Philosophy*, p. 22. Dummett considera aliás que foi também este o imperativo director da fenomenologia que contemporaneamente nascia. As figuras teóricas de Frege e Husserl são objecto de comparação neste livro.

<sup>539</sup> Como vimos, este era um ponto filosoficamente complicado para Chomsky, o ponto em que ele reconhece o problema de Wittgenstein, mas este problema não diz respeito apenas ao modelo os *Princípios e Parâmetros*; é um problema geral da teoria do conhecimento.

que os pensamos, e que é um erro defender sem mais que os pensamentos fazem parte do fluxo da consciência do indivíduo. Nem tudo aquilo que é preciso tratar quando se trata o pensamento (nomeadamente a significação e a verdade) aparece no estudo do indivíduo isolado.

Os filósofos analíticos vieram portanto a pensar que os pensamentos devem ser analisados em termos de linguagem. Para Dummett, a tónica do pensamento de Frege, autor em que é possível radicar a origem desta tese da prioridade, era já a comunicabilidade dos pensamentos (*Gedanken*) por oposição às representações (*Vorstellungen*). A ideia wittgensteiniana segundo a qual “o significado é o uso”, i.e. a ideia de um carácter social da significação é uma exploração desta intuição de Frege<sup>540</sup>.

Por curiosidade, para evidenciar este ponto (embora aqui esteja a conceder uma posição realista, a de Frege, que é, precisamente, contestável), Frege pensava que a tese da prioridade era relativa a nós, humanos<sup>541</sup>, i.e. que é concebível que seres sem linguagem possam apreender os mesmos pensamentos que nós. O pensável não é simplesmente identificável com o humanamente pensável.

Com isto queria mostrar que a questão da teoria do pensamento, mesmo se o pensamento só fôr analisável através da linguagem, não é apenas a questão empiricamente tratável da formulação linguística do pensamento, mas sobretudo a questão da significação e a verdade. Este é aliás um problema para os filósofos que são ao mesmo tempo cientistas cognitivos e que se converteram à maneira empírica de trabalhar. A alguns desses filósofos conviria perguntar se a questão da mente será simplesmente identificável com a questão empírica, psicológica, e se não está a dar-se, com a filosofia da mente, uma regressão, uma nova *naiveté*, em relação ao estado de “extrusão do pensamento em relação à mente” conseguido pela filosofia analítica enquanto filosofia da linguagem. Dummett, precisamente, vê na filosofia da mente que hoje (sobretudo nos Estados Unidos) substituiu a filosofia da linguagem no centro do “palco filosófico” os perigos de psicologismo, solipsismo e cientismo<sup>542</sup>...

Não encontraremos portanto na teoria empírica da mente uma teoria do significação e da verdade. Como se sabe, Frege relacionou estas duas questões. Considerou que compreender o significado de uma frase é saber as suas condições de verdade, e que as condições de verdade são as condições que o mundo tem que satisfazer para a frase ser verdadeira. Mas Frege, segundo Dummett, evitou dizer que as condições de verdade têm

<sup>540</sup> Cf. Dummett 1997, pp.25-26. É claro que esta interpretação de Dummett é muito particular e contestável.

<sup>541</sup> Dummett 1997, p. 51.

<sup>542</sup> É esse o resumo das observações de Dummett em Dummett 1997.

a ver com o modo como nós atribuímos valores de verdade, considerou antes que existe uma determinação objectiva destes, mesmo que nós não sejamos capazes de a fazer. Para Dummett, é Wittgenstein quem formula a questão da determinabilidade nos termos daquilo que, nós humanos, somos capazes de fazer. E é isso que Dummett quer fazer também, formular a questão da significação e da verdade nos termos do que *nós* podemos fazer. Não pensando assim, ignorando mais ou menos conscientemente que as condições de verdade têm a ver com o modo como nós humanos atribuímos os valores de verdade caímos na ilusão realista: “A ilusão vem da imagem que nós temos de uma realidade bem determinada que podemos não ser capazes de observar mas que no entanto tem que estar fixada num sentido ou noutro. Não é falso? Bem, então tem que ser verdadeiro. É simples, é uma imagem a que é psicologicamente difícil de resistir e que nós usamos”<sup>543</sup>.

É à consideração desta questão que chega a questão filosófica sobre a mente e ela constitui a “arena” da diferença da filosofia em relação a abordagens empíricas da cognição. As questões da verdade e da significação conduzem portanto aquele que é para Dummett um dos mais incontornáveis problemas da filosofia, o problema do realismo versus anti-realismo. Se Dummett pensa, como Wittgenstein, que a significação e a verdade devem ser pensadas em termos daquilo que nós podemos fazer, há no entanto uma grande diferença entre o seu pensamento e o de Wittgenstein, uma diferença que é também acerca do que é a filosofia. Wittgenstein pensa que a prática à qual se reportam a significação e o pensamento não necessita de justificação nem pode ser criticada, mesmo contradições da prática deveriam ser aceites como factos. Pensa também que a filosofia não poderia intervir em nada, mas apenas descrever. Pelo contrário para Dummett, se nas nossas práticas de pensamento somos levados a asserir coisas para as quais não temos justificação<sup>544</sup>, é possível e desejável avaliar e justificar como se pensa, não nos limitando a constatar que se pensa assim. Segundo Dummett é um mau caminho aquele que nos conduz a pensar que nas formas de inferência nada necessita de justificação, as inferências não estão separadas do resto da prática linguística e conduzem a asserções que podem conter constantes lógicas. Temos portanto que saber se estamos justificados nessa asserções.

<sup>543</sup> Dummett 1997.

<sup>544</sup> Dummett 1997, p. 30. Para Dummett, o significado de uma asserção é dado pelo tipo de justificação que poderia sustentá-la. A justificação de qualquer asserção envolve não apenas experiência mas também inferência, logo as condições de justificação não poderão ser formuladas atomisticamente. É claro que a noção de justificação substitui a palavra mais *demodée* ‘verificação’, mas como diz Dummett “Se a teoria da significação ser verificacionista significa que ela é feita em termos da maneira como nós estabelecemos o valor de verdade de uma asserção, então a minha teoria é verificacionista” (idem, p.43).

Esta posição de Dummett acerca de asserção e justificação vai mesmo conduzi-lo a defender uma revisão das leis da lógica clássica (no que ele considera ser ainda uma consequência do lema wittgensteiniano segundo o qual “o significado é o uso”<sup>545</sup>).

### Nota final

Em suma, são hoje muitos os pontos de comunicação entre o trabalho dos linguistas e dos filósofos em teoria da mente. Procurei defender (1) que a criação desta arena comum deriva precisamente da emergência histórica recente de um campo disciplinar em torno do mental. Procurei mostrar também que, mesmo sendo a linguística entendida tal como Chomsky deseja, como uma teoria da mente e como prática empírica naturalista, não poderá deixar de encontrar a certo ponto problemas da mente e da linguagem que não são problemas empíricos (2, 3). A questão empírica acerca da mente não tem a mesma forma que a questão filosófica (4). Isto significa que o problema mente-corpo, o problema das ciências cognitivas, não é composto somente por problemas empíricos: é argumentável que não são problemas empíricos os problemas dos *qualia*, da *aboutness*, da consciência, da identidade pessoal, da existência de significação e verdade, do realismo/anti-realismo, da avaliação das práticas de pensamento. Esses problemas não empíricos são problemas epistemológicos, metafísicos, éticos, i.e. problemas filosóficos.

### Referências

Chalmers, David, *The Conscious Mind*, Oxford, Oxford University Press, 1996.

Chomsky, Noam, *Syntactic Structures*, The Hague, Mouton, 1957.

Chomsky, Noam, *Aspects of the Theory of Syntax*, Cambridge, MA, MIT Press, 1965.

Chomsky, Noam, *Rules and Representations*, New York, Columbia University Press, 1980.

Chomsky, Noam, *Knowledge of Language*, New York, Praeger, 1986.

Chomsky, Noam, *The Minimalist Program*, Cambridge MA, MIT Press, 1995.

Chomsky, Noam, 1994, “Chomsky”, in Dancy, J., *The Blackwell Companion to the Philosophy of Mind*, Oxford, Blackwell, 1994.

<sup>545</sup> Dummett 1997, p.27. Cf tb. *The Philosophical Basis of Intuitionistic Logic*, in Dummett 1978.

Dennett, Daniel, *Consciousness Explained*, New York Little, Brown and Comoany, 1991.

Dennett, Daniel, *Darwin's Dangerous Idea*, New York, Simon and Schuster, 1995.

Dennett, Daniel, *Brainchildren*, London, Penguin, 1998 .

Descombes, Vincent, *La Denrée Mentale*, Paris, Minuit, 1995.

Dummett, Michael, *Truth and Other Enigmas*, Cambridge MA, Harvard University Press, 1978.

Dummett, Michael (entrevistado por Fabrice Pataut), Uma Perspectiva Anti-realista sobre a Linguagem, o Pensamento, a Lógica e a História da Filosofia Analítica, *disputatio*, 3, 1997.

Dummett, Michael, *Origins of Analytical Philosophy*, Cambridge, MA, Harvard University Press, 1993.

Fodor, Jerry, Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology, *Behavioral and Brain Sciences* 3, 63-73, 1980.

Fodor, Jerry, *Psychosemantics*, Cambridge, MA, MIT Press, 1987.

Fodor, Jerry, 1994, "Fodor", in Dancy, J., *The Blackwell Companion to the Philosophy of Mind*, Oxford, Blackwell, 1994.

Gardner, Howard, *The Mind's New Science - A History of the Cognitive Revolution*, New York, Basic Books, 1985.

Jackendoff, Ray, *Consciousness and the Computational Mind*, Cambridge MA, The MIT Press, 1987.

Jackendoff, Ray, *Languages of the Mind*, Cambridge MA, The MIT Press, 1992.

Kim, Jaegwon, *Philosophy of Mind*, Oxford, Westview, 1996.

Kripke, Saul, *Wittgenstein on Rules and Private Languages*, Cambridge, MA, Harvard UP, 1982.

Lakoff, George e Johnson, Mark, *Metaphors We Live By*, Chicago, University of Chicago Press, 1980.

Lakoff, George, *Women, Fire and Dangerous Things - What Categories Tell Us about the Mind*, Chicago, University of Chicago Press, 1987.

Langacker, Ronald, *Foundations of Cognitive Grammar*, Stanford, California, Stanford University Press, 1986.

Miguens, Sofia, *As Ciências Cognitivas e a Naturalização do Simbólico*, Tese de Mestrado, Universidade Nova de Lisboa, 1995.

Miguens, Sofia, Dennett, Millikan e o Teleofuncionalismo, *Revista Portuguesa de Filosofia*, LIV, 3-4, 1998.

Miguens, Sofia, *Daniel Dennett e a Inteligência Artificial*, Intelectu, ([www.terravista.pt/Guincho/2644/colaboradores](http://www.terravista.pt/Guincho/2644/colaboradores)), 1999a. Também neste volume.

Miguens, Sofia, *A mente segundo Chomsky*, Intelectu, ([www.terravis-](http://www.terravis-)

ta.pt /Guincho/2644/colaboradores), 1999b. Também neste volume.

Moura, Vítor, Wittgenstein, A possibilidade do jogo transcendental, in *Diacrítica* 9, 1994.

Pinker, S., *The Language Instinct*, London, Penguin, 1994

Putnam, H., *Minds and Machines* (1960), in S. Hook, ed., *Dimensions of Mind*, New York UP, 1960

Putnam, H., *The Nature of Psychological Predicates* (1967) in Putnam, Hilary, *Mind, Language and Reality, Philosophical Papers*, vol. 2, Cambridge, CUP, 1975.

Putnam, H., *The Meaning of Meaning*, 1975, in Putnam, Hilary, *Mind, Language and Reality, Philosophical Papers*, vol. 2, Cambridge, CUP, 1975.

Pylyshyn, Zenon, *Computation and Cognition*, Cambridge MA, MIT Press, 1984.

Searle, John, *A Redescoberta da Mente*, Lisboa, Instituto Piaget, 1998.

Simon, Herbert, *The Sciences of the Artificial*, Cambridge MA, The MIT Press, 1969.

Talmy, Leonard, *Force Dynamics in Language and Cognition*, *Cognitive Science*, 12, 49-100, 1988.

Wittgenstein, Ludwig, *Investigações Filosóficas*, Lisboa, Fundação Calouste Gulbenkian, 1987.

Zilhão, António, *Linguagem da Filosofia, Filosofia da Linguagem- estudos sobre Wittgenstein*, Lisboa, Colibri, 1993.





*LANGUAGE AND THOUGHT (THE NATURE OF MIND FROM G. FREGE AND J. FODOR TO COGNITIVE LINGUISTICS)*<sup>546</sup>

Abstract: In trying to decide how to go about studying the relation between language and thought we face several options. We may, for instance, do work in logic, or investigate mental models associated with natural languages, or set out to compare animal and human minds. Still, whichever option we decide for, we will not necessarily have to make our commitments concerning the nature of mind and the place of mind in nature explicit. Philosophers of mind, though, are especially interested in making such commitments explicit. In this article I will try, from the viewpoint of the philosophy of mind, to analyse the commitments about the nature of mind that can be found in Frege, Fodor and cognitive linguistics.

### **Introduction**

In trying to decide how to go about studying the relation between language and thought we face several options. We may for instance, do work in logic, or investigate mental models associated with natural languages, or set out to compare animal and human minds. Still, whatever option we decide for, we will not necessarily have to make our commitments concerning the nature of mind and the place of mind in nature explicit. Philosophers of mind, though, are especially interested in making such commitments explicit, and this may involve distinctions among close notions, such as those of thought and cognition. That is the perspective I try to explore in this article.

I will consider on the one hand two philosophers – two analytical philosophers one century apart, G. Frege e J. Fodor, and on the other hand cognitive linguists, some of them present in this conference. I will explain why I chose Frege and Fodor for this comparison, which is fact also a comparison of the somewhat different ways linguists and philosophers engage in the study of the relation between thought and language. This is the reason. Both Frege and Fodor worried about the issue at stake – language

<sup>546</sup> I thank Professor George Lakoff for the many valuable suggestions he gave me in the Braga Conference (2003) in which I gave this talk, on how to develop the ideas defended in the last part of the paper.

and thought – and both proclaimed the need to replace, at least in some theoretical contexts, natural languages for something different. This was, in Frege’s case, logic, and in Fodor’s case the language of thought, considered as a design feature of cognitive systems. What in my opinion makes them relevant is the fact that they can help us clarify what we mean by ‘the problem of the relation between language and thought’. For instance, and more specifically, is there something wrong, when we’re dealing with the problem, in not identifying ‘language’ with natural language? This is, in fact, not uncommon in philosophy.

I want to assess the legitimacy of such a shift of focus, from natural language to something else, as the one we find in Frege and Fodor. I must say I am not persuaded that Frege or Fodor are misguided, although what I will call their formalist commitments definitely seem bound to a head-on collision with some of cognitive linguistics central tenets. But I also think cognitive linguistics gives us ground to argue against overlooking natural languages when working on general questions, such as those concerning the relation between language and thought or the nature of mind.

What I will try to do here then is first (i) to give some information about the way these two philosophers have dealt with the problem of language and thought. In doing that I will (ii) try to underline the specificity of the philosophical perspective and to point out some of the reasons in favour of what I call, borrowing the term from G. Lakoff (Lakoff 1987), ‘formalist perspectives’. I will also try to (iii) defend the need for a distinction between thought and cognition. Only then will it be possible to (iv) locate the contribution of cognitive linguistics in the context of a wider discussion about thought and language and look for cognitive linguistics to provide us with arguments against overlooking natural languages when we are trying to understand the nature of mind and the place of human mind in nature.

### **1. Cognitive linguistics and formalist perspectives**

Let me begin by spelling out what I think brings Frege and Fodor together: I will call it a formalist perspective on the problem of language and thought. Actually, we may look at Frege and Fodor as examples of what George Lakoff (Lakoff 1987) called ‘the metaphor of formal languages for natural languages’. When Lakoff targets this metaphor (the very core of the formalist approach he wants to replace with his ‘experientialist view of thought’ and of embodied, imaginative reason), he aims not only at Chomsky but at the omnipresence of logic in the study of language, and

so also at much contemporary philosophy. Cognitive linguists are perfectly aware of how widespread formalist convictions about the nature of language are and how different from cognitive linguistics central tenets they are. For instance, G. Lakoff (Lakoff 1987: 58) says: «It is by no means obvious that language makes use of our general cognitive apparatus» (*this is of course a central thesis of cognitive linguistics*) «In fact, the most widely accepted views of language within both linguistics and the philosophy of language make the opposite assumption: that language is a separate modular system independent of the rest of cognition. The independence of grammar from the rest of cognition is perhaps the most fundamental assumption on which Noam Chomsky's theory of language rests (...) the very idea that language is a formal system (in the technical mathematical sense used by Chomsky and many other linguistic theorists) requires the assumption that language is independent from the rest of cognition».

## 2. Philosophical characters

Let me now present the philosophical characters I choose and try to explain why I think they are important for us here. I didn't choose them in vain, they represent trends, methodological orientations, which are very widely spread, at least in contemporary analytical philosophy. Frege was a German mathematician and philosopher, whose main works were published approximately 100-120 years ago. He is usually regarded as the 'founding father' of three things: analytical philosophy, formal logic and philosophy of language. He represents here the intimate relation of contemporary analytical philosophy with logic. Jerry Fodor is one of the most important contemporary philosophers of mind. He has also worked in linguistics and was especially close to N. Chomsky, theoretically and spatially speaking, in the 60s (when they were both at the MIT). What he represents here is an ontological view of cognition, which we may call cognitivism.

### 2. 1. Frege

Let us start with Frege. From the point of view of philosophy, Frege's work marks an historical moment, the moment when formal logic changes the way philosophers approach the problem of the structure of thought. Let us consider a thought with the structure 'A is B'. If we look at classical philosophy, this problem was for instance for the empiricist D. Hume a question of association of ideas, for Kant a question of judgments, somehow originated by the mind through the synthetic unity of apperception, that is, consciousness. But to Frege this question – what makes concepts

hang together in a thought, so that we may entertain it – has a logical nature, is not a psychological or transcendental question. This new approach to the problem, the fact that Frege comes up with the apparatus of modern formal logic to deal with it, is the origin of so called analytical philosophy.

Analytical philosophy starts from the idea according to which the object of philosophy is the analysis of the structure of thoughts. And that is done by means of an analysis of language. In other words, philosophy's epistemological-ontological inquiry (how do we know? what is there in the world?) is to proceed by means of logical-linguistic investigations. This is the so called linguistic turn in philosophy, a conception which brings together philosophers as different as Frege, Wittgenstein, Quine, Davidson, etc.

Now, language as it is considered here – when our aim is to investigate the nature of thought and ultimately the nature of reality – is not necessarily natural language. To Frege, namely, and that's one reason why it is slightly ironical to consider him the initiator of philosophy of language, natural language is plainly an obstacle. A great deal of the work of a philosopher as Frege conceived of it is to get rid of natural language, more specifically to get rid the vagueness and ambiguity which characterize it. Frege thinks natural languages resist any coherent systematization, namely any systematization of the laws of thought. That's why Frege creates an artificial language, a *Begriffsschrift*, or conceptual notation, which is a fundamental element of his philosophical outlook, and which is supposed to provide more precise means to express thought and to formulate the principles of thought explicitly. If we consider the way Frege meant to put his *Begriffsschrift* (Frege 1879) to use – namely in his investigation of the logical foundations of arithmetic – we get a clear example of Frege's approach to the problem of the relation between language and thought.

So, recapitulating, according to Frege vagueness will forever plague natural language thinking and we can do better than that. One thing is the pragmatic skill, which is psychologically basic, essential to each of us, natural language thinking, another is the best theory we can develop in a certain domain and for the conceptual analysis of those theories artificial languages will help us best. But still, why the focus on language if what we are interested in are ontological and epistemological investigations? Frege's point is that language may be a distorting mirror but it is the only one we possess. What does he mean? He means that other beings, other kinds of minds, might perhaps do without language but not humans. Our thought is linguistic (articulate, inferential, not intuitive). This is why focusing on language when we want to analyze thought is unavoidable. Another reason for focusing our epistemological and ontological inquiries on language is

the fact that this allows for a de-mentalized, de-psychologized investigation of thought. In other words, it allows us to keep cognitive questions (questions about minds, which have thoughts) apart from the questions the philosopher-logician is interested in (questions concerning the structure of thought and theories (and reality)).

The characteristic of thought which makes this ‘separation’, this division of labour, possible is thought’s intentionality. I will explain what’s usually meant by it. Philosophers usually make a distinction between *qualia*, the very private feeling of thinking, what it is like to be a mind, and intentionality, aboutness, what the thought is about (numbers, chairs, chemical elements, galaxies). My *quale* of red is – maybe – different from yours (I wouldn’t know, given the privacy of minds) but we both use the word ‘red’ to identify something wordly, something supposedly common. We may then say that it is language that makes this objectivity possible. In other words, it is language that makes room for an objectivity of thought, it is language that makes room for the independence of thought from mental processes, from individual minds. So, to Frege, philosophy consists in logical investigations, and he thinks of these logical investigations as being about thought and not about mental processes or cognition. He thinks the objectivity of human thought is due to its linguistic nature. That’s why he is a philosopher of language.

I will give a brief example, an historically important one, of the way Frege puts his conceptual notation to use. What Frege proposes to do in the *Foundations of Arithmetic* (Frege 1884) is the first example of analytical philosophy in the sense I have been trying to sketch. What we have there is the program for a logical investigation, an investigation of the concept of number. The investigation proceeds by asking how the meaning of sentences which comprise terms for numbers is determined. Now, what Frege says he wants to do in *Foundations of Arithmetic* depends on his Begriffsschrift. He is sketching a program of use for the technical means he had devised, the conceptual notation, and he wants to use this conceptual notation to help him go beyond the vagueness of natural language in thinking about the nature of numbers.

In fact, Frege’s philosophical agenda is to show that truths of arithmetic can be derived from logic alone. That philosophical project is known as logicism and its failure is not particularly important for us here. What we were interested in are the reasons why Frege developed an artificial language, a Begriffsschrift, to approach the problem of language and thought. And we could formulate them thus: he wants to overcome the vagueness and ambiguity of natural language, still capture the objectivity of thought,

and also to do something that is not possible by means of natural language only: to make the nature of sound inference explicit.

Frege's example is still very influential in philosophy, as a model of how we go about thinking about thought. We may say that when Frege speaks of thought, what he is considering are conscious, explicit, justified beliefs, which constitute our best theories in whatever domain. It is not common cognitive tasks which interest him but the best theories available, and what those theories make us think about the way the world is (as for instance, in the example I gave, number theory – it is not hard to see that the foundations of arithmetic are not a problem for the individual cognizer)). This is what he proposes formal logic for. Logic is essential for philosophy's endeavour in that epistemological-ontological investigations cannot be done by means of natural languages only.

It is also very important to notice that he does not want take his artificial language back and apply it to natural languages. Of course history shows us that formal logic, created by Frege for the analysis of thought was projected back, throughout the 20th century, to modeling natural languages. But Frege did not conceive logic as modeling real reasoning, psychology or natural language thinking. Fregean thought is not mind or cognition. Frege was not interested in cognition. If we consider the way he describes the analysis of thought by means of an analysis of language we see that for him it is ultimately unimportant to consider the physical and psychological nature of cognition. This guarantees on the one hand a claim of universality – the logical-linguistic investigation of the concept of number is supposed to be good for every thinking human, it is the essence of thought, according to Frege that it is transferable, communicable between minds, without a residue – but on the other hand this poses a problem: fregean thought has no place in nature (nor need it have one, at least in this context).

So, what we get from Frege is: thought is not individual cognition (the psychological going-about of categorization, object identification, memory, processing, etc). Thought is what our species elaborates by means of theoretical, scientific inquiry, about the nature of reality. Mind, cognition, is only a part of that reality (at least for those who profess a physicalist theory of mind; I wouldn't bring Frege in here). Anyway, that's why he is a philosopher of thought.

## 2. 2. Fodor

Fodor is definitely something different. Fodor is a philosopher of cognition: he is interested in the nature of mental states and processes and

in their relation to the brain and to the physical world. In terms of philosophy, much and almost one century stands between Frege and Fodor – some people would speak of a tendency towards psychologizing within analytical philosophy (this is supposed to be a very bad thing). The fact is, in the last decades philosophy of mind has somehow replaced philosophy of language in terms of importance. Arguably, this is due to cognitive science and to the fact that it represents such a challenge to philosophy.

From Frege to Fodor, then, we go from thought, as defined, to cognition, as defined, but there is something in common: language still occupies centre stage in philosophical investigations and natural language is still considered somehow secondary. Why is that so? Language is very important to Fodor in that he basically thinks mind or cognition exists in nature because there is language. But the language Fodor thinks of, the language of thought, is not natural language. To Fodor, natural languages come second, as an interface. This is the so called ‘communicative conception of the relation between language and thought’ (Carruthers 1996). What does Fodor mean by this? He means that for there to be minds (i.e. intentionality as representation and rationality as representation transformation, use of representations to guide action) there must be symbols and use of symbols. More precisely, there must be the possibility of articulation and productivity of representations, and only a language can guarantee that.

In his 1975 book, *The Language of Thought*, Fodor claimed that no cognitive scientist who takes psychological explanation to be computational, can refuse to admit a medium for computation, an internal system of representations. And this is what the language of thought is, and this is not natural language. As Fodor puts it, it may seem crude, offensive and unbiological to suppose – in order to be able to explain representation and rationality – that people have sentences in their heads<sup>547</sup>, but it is cognitive theory that compels us to accept it. The Language of Thought Hypothesis is then an engineering hypothesis according to which all cognitive systems must share one design trait. They must have an internal representation system, a language of thought. What we call mental processes should then be conceived as computations of these representations. Representations themselves are instantiations of symbols in physical systems. And as Fodor puts it himself: «No representations, no computations. No computations, no mind»<sup>548</sup>. We must be careful here: the Representational Theory of Mind associated with Fodor’s Language of Thought Hypothesis is not a

<sup>547</sup> Fodor 1990a, *Fodor’s Guide to Mental Representations*.

<sup>548</sup> Fodor 1975: 31. The centrality of this idea to cognitivism is particularly clear in Pylyshyn 1984. In the Preface, the author explicitly thanks Fodor, whom he takes to be responsible for having showed him what philosophy is for.



theory of content or meaning. That is another big problem for Fodor, but a different one. RTM is a syntactic theory of the nature of certain entities, which make cognition and mind possible, a theory of, I would say, empty syntax, not semantics.

So, everything Fodor says when he defends the Language of Thought Hypothesis is about cognition, not about thought in the Fregean sense. It is the problem of the nature of cognition Fodor wants to deal with, and he sees it as a hardware-software problem in a biological system such as the human brain. The Language of Thought is a software for cognition, something which is 'below' natural language thinking. Natural language thinking would presumably involve consciousness, something Fodor doesn't even touch here.

### 3. Cognitive Linguistics

I agree with Lakoff – and that's why in the beginning of this article I brought in a quotation of *Women, Fire and Dangerous Things* – that a great deal of contemporary approaches of language within philosophy, whether they are more connected with Frege's lineage, more epistemological-ontological, or more connected with Fodor's, more psychological and cognitive, are in fact somehow committed to the metaphor of formal languages for natural languages. That commitment is not bad *per se*. In fact if we look at what Frege's main issue was, we understand it concerns the structure of theories, the soundness of inference, not cognition. Arguably, formal languages have advantages in helping us there. I think the formalist commitment brings us directly into problems only when what we are dealing with is the nature of mind and cognition.

Could it be then that cognitive linguistics has something to propose here? And in case it does, which one of our philosophical characters would it oppose more directly? Frege or Fodor? Fodor, of course, and this is why I said Frege and Fodor could show us different aspects of the so called problem of the relation between thought and language. Cognitive linguistics' principles stand opposed to Fodor's because they're both dealing with the nature of cognition, not with thought in Frege's sense.

What would cognitive linguistics positions about the place of mind in nature then be, knowing they are contrary to Fodor's? I will of course consider only very general principles, but I will try to be systematic. I've been searching for traces of this discussion in the writings of Langacker, Lakoff and Talmy (Langacker 1978, Lakoff 1987, Talmy 1988). In *Women, Fire and Dangerous Things* (Lakoff 1987) George Lakoff points very directly at what opposes cognitive linguistics to formalist perspectives

of language: a formalist perspective considers that minds are symbol processors, formal software, running on whichever – biological, artificial – hardware. Now the kind of linguistic phenomena cognitive linguists have always been interested in – grammar as imagery, typicality in categorization, metaphor, force dynamics – give us evidence to build a case against that conception of mind and mental processes. What do we get then as an alternative to formalism? Embodiment of mind and meaning, cognitive linguists say. But what does it mean?

Let's put our philosophical characters to use here. It is this embodiment of mind both Frege and Fodor overlook. They do it in different ways, though. Given Frege's aims, it is not unreasonable to consider the physical nature of symbolism indifferent. But given the fact that Fodor's cognitivism is a self-professed physicalism or materialism, he does have to deal with the physical nature of symbolism, its implementation and processing. And there he faces a problem: Fodor's cognitivist ontology is committed to a dualism of physics and symbolism. In fact, only this dualism makes his formalism sustainable.

#### 4. Conclusion

I will finish on a practical note (as practical as possible, coming from philosophy...): I think that if we spell out the principles of Fodor's approach to the problem of the nature of mind and representation, we will get, by contrast, cognitive linguistics philosophical principles. Those positions are what makes cognitive linguistics very interesting from the point of view of the philosophy of mind.

So, where on the side of Fodor's Representational-Computational theory of mind we have the idea that only a language can make for the existence of cognition and the isolation of language from the rest of cognition, on the side of cognitive linguistics we have the idea that language is not independent from the rest of cognition. Now, language here is natural language, seen in its specificity and not through the lens of formal languages, and thus involving (i) a relation of linguistic categories to perception, (ii) processing of space/time relations, characteristics which are obviously absent in formal languages. This intimate relation of what we call mental with perception and with space-time is the context for the well known thesis of the meaningfulness of syntax, syntax as imagery, continuity between grammatical and lexical categories, pervasiveness of metaphor, force dynamics phenomena, etc.

Again the metaphor of formal languages makes the cognitivist side conceive of language processing as consisting in 'Rules and Representations',

on the model of logics. On the other side we will more likely have neuroscience as a model and language categories seen as patterns, becoming entrenched (Langacker 1987). More generally, it means that on the one hand we have a cognitive system seen as a formal system implemented, on the other hand a cognitive system seen as a dynamic system. These oppositions are common enough in cognitive science. What do they translate in, in philosophical terms, besides the centrality of perception and space-time in the nature of mind (as opposed to the centrality of logic)? I think that we have two forms of physicalism, one that is dualistic, leaving us with the problem of the bridge from physical to symbolic, as Zenon Pylyshyn calls it (Pylyshyn 1984), the other is monist, involving a continuity from the physical nature of cognitive systems to perception to language. What else do we have, considering, namely, the ongoing philosophical discussion about the place of consciousness in nature (whether it is a basic building block of the world or not)? We have on the one hand unconscious cognitive (algorithmic) processes and eventually an abrupt distinction between consciousness and non-consciousness, on the other hand continuity from representation to consciousness (and so consciousness as a form of representation).

Concerning cognition and not thought – through Frege I think I showed what cognitive linguistics is *not* talking about – I would say that what we have here is a confrontation, within physicalism, of mind embodied versus mind implemented. We may now see that cognitive linguistics basic tenets – that language is not independent from the rest of cognition, the embodiment of meaning and mind – represent a general orientation opposed to Fodor's cognitivist ontology, which is rather widespread in cognitive science. So cognitive linguistics, seen from the perspective of philosophers interested in the nature of mind, can be invaluable in providing evidence and data that allows us to think about the several dimensions of the embodiment of mind, namely the continuity of physical nature of cognitive systems and symbol instantiation and processing, and the importance of perception in what we call mind. This is why, I think cognitive linguistics can be so important persuading us that a formalist perspective, which leads us to overlook natural languages, is a bad choice when we are facing the problem of the nature of mind.

### References

- Carruthers, Peter, 1996 *Language, Thought and Consciousness – An essay in philosophical psychology*. Cambridge: Cambridge University Press.
- Dummett, Michael, 1981, *Frege – Philosophy of Language*. Cambridge MA: Harvard University Press.

Frege, Gottlob (1879) 1967, *Begriffsschrift – a formula language, modeled upon that of arithmetic, for pure thought*. In Jean van Heijenoort (ed), 1967, *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931* Cambridge MA: Harvard University Press.

Frege, Gottlob (1884) 1992, *Fundamentos da Aritmética*. Lisboa: INCM.

Fodor, Jerry, 1975 *The Language of Thought*, Cambridge MA: Harvard University Press.

Lakoff, George & Johnson, Mark, 1980, *Metaphors We Live By*, Chicago, Chicago University Press.

Lakoff, George, 1987, *Women, Fire and Dangerous Things – What Categories Tell us About the Mind*, Chicago: Chicago University Press.

Langacker, Ronald, 1987, *Foundations of Cognitive Grammar*, vol. 1, Stanford, California, Stanford University Press.

Miguens, Sofia, 1997, As Ciências Cognitivas e a Naturalização do Simbólico – a mente computacional e a mente fenomenológica. *Revista da Faculdade de Letras – Série de Filosofia*, 14 : 385-427.

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

Pylyshyn, Zenon, 1984, *Computation and Cognition*, Cambridge MA, MIT Press.

Talmy, Leonard, 1988, Force Dynamics in Language and Cognition, *Cognitive Science* 12.

Weiner, Joan, 1999, *Frege*, Oxford, Oxford University Press.



# RACIONALIDADE E FILOSOFIA CONTINENTAL



## UM CERTO DESDÉM DE HABERMAS FACE A DERRIDA

“The sad truth is that poems don’t have presence, unity, form or meaning... What does a poem possess or create? Alas, a poem has nothing and creates nothing. Its presence is a promise, part of the substance of things hoped for, the evidence of things not seen. Its unity is in the good will of the reader... its meaning is just that there is, or rather, was, another poem.”

H. Bloom, *Kabbalah and Criticism*

### 1. Contradições performativas

No *Discurso Filosófico da Modernidade* (1985<sup>549</sup>), um claro e polémico panorama das filiações e destinos teóricos no pensamento europeu pós-kantiano, Habermas acusa um conjunto de críticos da razão e da ambição iluminista moderna— por exemplo Heidegger, Adorno, Bataille e Foucault, mas também, no caso que aqui me interessará, Derrida— de uma *contradição performativa*, i.e. acusa-os de fazerem aquilo que não podem, por não poderem justificar aquilo que fazem. Nomeadamente, estes autores continuam a produzir discursos servindo-se do próprio meio que é entretanto acusado e desmascarado, pretendem produzir discursos especiais, situados fora do horizonte estrito da razão cognitiva e argumentativa, mas sem serem totalmente irracionais. Segundo Habermas, esta situação provém de Nietzsche, cuja herança ele vê e pretende fazer ver em todos os autores referidos.

<sup>549</sup> As referências dizem respeito à edição portuguesa [1990] do *Discurso Filosófico da Modernidade*. Habermas define o discurso da modernidade como aquele que desde o fim do séc. XVIII tem tido um único tema sob títulos sempre novos: o enfraquecimento dos laços sociais e as deformações da praxis quotidiana devidas à racionalização, deformações essas que evocariam a necessidade de qualquer coisa de equivalente ao poder unificador perdido da religião. O problema é portanto saber se pode haver uma orientação racional normativa da “consciência moderna”, se a razão é capaz de se dar uma finalidade uma vez afastada a legitimação transcendente, uma vez que a autonomização dos discursos [científico, jurídico, artístico] e dos sistemas sociais [estado, economia], característica da modernidade, pode ser experienciada como “alienação”. Ao contrário da religião que a antecedeu, a razão não teria o poder de ligar os indivíduos numa totalidade ética. Hegel foi o primeiro filósofo para quem a modernidade na sua relação com a racionalidade e com o possível alheamento dos indivíduos em relação aos funcionamentos racionais auto-suficientes da sociedade se tornou o problema da filosofia



As soluções encontradas pelos vários críticos da modernidade para a contradição performativa inerente à auto-crítica da razão centrada no sujeito, a razão “autora” do projecto civilizacional iluminista da Modernidade, são diversas. Segundo a descrição de Habermas, Heidegger escapa-se furtivamente, recolhendo-se à “silenciosa contemplação do místico”, “assumindo num gesto de visionário a autoridade do iniciado”<sup>550</sup>. Adorno, que não partilha do desprezo elitista de Heidegger pelo pensamento discursivo mas pensa que as premissas do Iluminismo morreram, e só as suas consequências (i.e. as sociedades contemporâneas) permanecem perante nós como um cadáver vazio de espírito, crê que “andamos à deriva no discurso como no exílio” e que o pensamento só pode ser “um constante auto-dementido”. Trata-se de um “fetichismo do desencantamento”, diz Habermas<sup>551</sup>. Bataille procura no domínio do sagrado e do erótico o “excesso da subjectividade que se transgride”, mas a partir de uma “ideia do sagrado rigorosamente atea”<sup>552</sup>. Foucault elabora uma historiografia arqueológica das ciências humanas a partir de uma teoria — nietschiana e não sociológica—do poder. Há no entanto algo de comum à diversidade das soluções: nenhum destes trajectos traduz um desejo iluminista de emancipação, e, segundo Habermas, estes discursos, sendo “interessantes no fundamental, permanecem indiferenciados nos resultados”<sup>553</sup>.

Como se sabe, é pela introdução de um *paradigma comunicacional de racionalidade* que Habermas pretende evitar no seu próprio discurso as aporias de que acusa os críticos da razão e da modernidade. Seria assim contornada a contradição performativa inevitável em qualquer tentativa de destituição da filosofia da consciência pelos próprios meios dessa filosofia da consciência (cf. *Discurso Filosófico da Modernidade*, Uma outra saída para a filosofia do sujeito<sup>554</sup>) e mantido aquilo que Habermas considera vivo na tradição da modernidade, o carácter emancipatório da racionalidade. Esta escolha afasta-o dos autores referidos, que não distinguiam na racionalização social característica da modernidade os aspectos emancipadores dos aspectos repressivos — uma vez que o seu nietschianismo lhes desvia o olhar da praxis quotidiana e dos ganhos de autonomia que nesta aconteceram nas sociedades modernas<sup>555</sup> — e aproxima-o de Marx, a figura de referência da Teoria Crítica de Frankfurt.

<sup>550</sup> Habermas 1990, p.178.

<sup>551</sup> Cf. Habermas 1990, p.179.

<sup>552</sup> Habermas 1990, p. 204.

<sup>553</sup> Habermas 1990, p. 310.

<sup>554</sup> Habermas 1990, p. 275.

<sup>555</sup> Para Habermas os críticos nietschianos da racionalidade são insensíveis ao conteúdo ambivalente da modernidade social e cultural: apenas vêem (e bem) a reificação e o controle envolvidos no processo moderno de racionalização, ignorando porém o que é para Habermas um simultâneo potencial utópico

Aliás, Habermas pensa que a concepção comunicacional de racionalidade lhe permitirá também evitar as aporias da teoria social crítica de linhagem marxiana que parte da acção social entendida como trabalho, em ligação com uma concepção da racionalidade centrada na consciência (Habermas pensa em Lukàcs, Adorno, Horkheimer).

A *Teoria da Acção Comunicacional* (1981)<sup>556</sup>, a obra na qual Habermas apoia o seu diagnóstico acerca do discurso filosófico da modernidade, além de ser uma investigação dos fundamentos da teoria social e uma reconstrução das ideias dos teóricos clássicos da sociedade (Marx, Weber, Durkheim, Mead, Parsons, Lukàcs, Adorno, Horkheimer) pretende ser também uma história da recepção no marxismo ocidental do diagnóstico de Max Weber sobre as sociedades modernas racionalizadas, assim como uma recuperação do projecto ético-político iluminista.

Este diagnóstico de M. Weber sobre “os tempos”, i.e. sobre a modernidade como racionalização, que para Habermas continua a ser pertinente e é central no seu projecto teórico, incide sobre a perda da capacidade da cultura de reconciliar os interesses particulares dos indivíduos através de convicções, uma *perda de sentido* que M. Weber liga à substituição social da religião por uma “razão restringida à dimensão cognitivo-instrumental (...) colocada ao serviço de uma auto-asserção meramente subjectiva”<sup>557</sup> e sobre a *perda de liberdade* no mundo burocratizado, administrado da modernidade. Habermas quer reapropriar-se deste diagnóstico dentro de uma filosofia da praxis<sup>558</sup>, que substitua o já não pertinente conceito de acção social de que Marx e toda a geração anterior da Escola de Frankfurt partem (i.e. o conceito de praxis como trabalho) pelo conceito de acção comunicacional. Assim, o anti-iluminismo de Weber, a sua identificação da razão socialmente triunfante com a técnica, o cálculo, a organização e a administração seria rebatido por meio do potencial emancipatório contido na acção comunicacional.

A caracterização da acção humana como linguisticidade ou da racionalidade como acção linguística, que permitirá a Habermas a sua leitura estratégica do Discurso Filosófico dos Modernos é explicitada na Teoria da Acção Comunicacional e do Mundo da Vida. A inspiração filosófica da teoria é múltipla – vai desde a teoria dos *speech acts* de Austin e Searle, às reflexões wittgensteinianas sobre a certeza, à ideia husserliana de *Lebenswelt*. Interessa-me aqui apenas aquilo que Habermas julga fazer com ela. Habermas pensa que faz uma reconstrução – à maneira de Chomsky – uma

<sup>556</sup> Habermas, *Theory of Communicative Action*, 1981.

<sup>557</sup> Habermas 1984, vol. II, p.301.

<sup>558</sup> Isto porque as patologias sociais visadas por Marx, as “deformações da praxis”, como lhes chama Habermas, não desapareceram.

evidenciação daquilo que está sempre já em acto, sempre presente em qualquer discurso, inscrito em qualquer uso humano de linguagem, nomeadamente uma competência comunicacional que estabelece a possibilidade de um acordo gerado por motivação racional. A linguisticidade da razão ou da acção racional tem o seu ponto exemplar, o seu elemento regulador, na concepção da *Ideel Sprachsituation*, a situação ideal de fala: i.e. a comunicação racional não constrangida é para Habermas uma possibilidade constitutiva inerente a cada situação de fala.

Habermas pensa que “atingir o entendimento é o *telos* inerente da fala humana”<sup>559</sup> e que a capacidade de dar razões por ou contra as pretensões de validade avançadas é o próprio núcleo da noção de racionalidade. Esta excede assim a construção de verdades proposicionais, uma vez que as pretensões de validade susceptíveis de serem sustentadas por razões são de várias dimensões (dá-se razões não apenas para sustentar pretensões à verdade proposicional mas também à rectidão normativa e à sinceridade expressiva). Segundo Habermas, não encontramos esta estrutura racional inerente à acção humana num mundo inteligível ou numa análise intuitiva da consciência de si: antes, temos diante dos olhos as formas, mesmo que distorcidas e fragmentárias, de incorporação histórica e social dessa razão. A reconstrução das condições de racionalidade não é desligada da situação empírica.

A ideia de uma constitutiva possibilidade de acordo gerado por motivação racional em toda a acção comunicacional, e portanto em toda a acção humana, é o ponto de referência normativo da teoria habermasiana da racionalidade, da ética, da sociedade e da modernidade.

Ora, apenas no contexto do projecto descortinável através das principais teses do *Discurso Filosófico da Modernidade* e da *Teoria da Acção Comunicacional* que sustenta essas teses, é compreensível a disjunção habermasiana entre retórica e argumentação na consideração do carácter linguístico da acção comunicacional, e portanto na consideração da racionalidade. Esta distinção de princípio é usada como arma de batalha num sub-capítulo do *Discurso Filosófico da Modernidade*, o *Excurso sobre o nivelamento da diferença genérica entre filosofia e literatura*<sup>560</sup>, escrito contra o filósofo francês Jacques Derrida, a pretexto do capítulo em que Habermas trata o pensamento deste (o capítulo intitulado “*Exacerbação da filosofia temporalizada da origem: A crítica de Derrida ao fonocentrismo*”<sup>561</sup>). Ela serve, creio, para lançar luz sobre as posições de Habermas acerca de moralidade, política e civilização.

<sup>559</sup> Habermas 1984, Vol. I, p.287.

<sup>560</sup> Habermas 1981, p.178.

<sup>561</sup> Habermas 1981, p.157.

Habermas pensa que a racionalidade enquanto acção linguística é argumentativa mas apenas superficialmente retórica, e isto permitir-lhe-á delimitar fronteiras claras entre uma linguagem normal na qual podemos esperar um tribunal de recurso argumentativo para os nosso diferendos e o discurso ficcional de qualquer tipo, ou, mais especificamente ainda, permite-lhe distinguir textos racionalmente prototípicos contra a consideração geral de todos os textos como texto (que é aquilo que faz Derrida, segundo Habermas).

A ausência desta delimitação clara permitir-nos-ia o passo em falso de Derrida, levar-nos-ia a supor o nivelamento da diferença genérica entre filosofia e literatura, e impedir-nos-ia de considerar certos textos “melhores” do que outros textos. Este movimento conduziria assim àquilo que é péssimo para a filosofia aos olhos de Habermas: uma *Literarisation*<sup>562</sup>, um literarizar-se, tornar-se literatura, que é ao mesmo tempo um afastamento das pretensões de racionalidade e por isso uma perda do potencial emancipatório ético-político.

Para pesar este veredicto, justo ou injusto, de Habermas, que interpreta Derrida e a desconstrução como reduzindo indevida e inevitavelmente a filosofia à literatura — um exemplo demasiado óbvio seria a desconstrução da oposição entre metáfora e conceito<sup>563</sup> — considere-se a comparação que R. Rorty faz em *Contingência, Ironia, Solidariedade* (1990), para medir as suas próprias posições filosóficas relativamente às de Foucault e Habermas. Poder-se-á assim ver melhor como aparece aquele que enuncia o veredicto.

Rorty utiliza como factor de comparação a figura do *ironista liberal* — que é afinal a proposta ética da sua filosofia<sup>564</sup> — e afirma: o primeiro

<sup>562</sup> Cf. Critchley & Honneth 1998, “Philosophy in Germany”, *Radical Philosophy* 89, May/June 1998, onde é discutido sob este nome o “perigo francês” em filosofia, apontado nomeadamente por Dieter Henrich e Jürgen Habermas, os dois grandes nomes da filosofia alemã contemporânea.

<sup>563</sup> Cf. “A Mitologia Branca”, em *Margens da Filosofia* (1972) acerca do infundado da distinção absoluta entre conceptualização e metaforicidade, distinção que é obviamente pressuposta na presunção de corte claro entre discursos racionais (científicos, filosóficos) e discursos artísticos, literários. Daí a mitologia branca, nome para a nossa racionalidade que assim se presume incontaminada. Note-se que a pretensão de Derrida não é assumir um estatuto artístico para o seu próprio discurso, mas, como diz Bennington (cf. Bennington & Derrida 1991, Jacques Derrida, p.119), mostrar a inapelável cumplicidade dos dois campos. Conceber o metafórico como apenas desvio do qual sempre se retorna seria a marca do metafísico, segundo Derrida (sob o título de ‘metafísica’ cairia qualquer presunção de uma racionalidade auto-assertiva, transparente a si). Por isso “Metáfora” é um conceito metafísico, é uma das instâncias da hierarquização entre originário e secundário/cópia, insustentável, para Derrida. Esses pares da originaridade versus secundaridade, como o par conceito-metáfora, são o próprio “des-construto” da desconstrução, como se verá à frente.

<sup>564</sup> Um ironista é para Rorty alguém que encara as suas próprias crenças e desejos como contingentes, que nunca pode ser absolutamente sério acerca de si próprio porque tem consciência da contingência da sua própria auto-descrição: “O ironista [encara] a possibilidade de ter sido iniciado na tribo errada, ensinado a jogar o jogo de linguagem errado”, cf. Rorty 1990, *Contingência, Ironia, Solidariedade* [CIS], p.75. Um liberal é aquele que pensa que a crueldade é a pior coisa que podemos praticar e vê esse como o primeiro princípio de uma razão pública. A figura do ironista liberal procura conjugar a possível, aliás provável, irrelevância privada do pública e politicamente benéfico [os exemplos de Rorty são as ideias políticas de J.Stuart Mill, de Rawls, de Habermas...] e a irrelevância pública/política do que é precioso para a auto-interpretação privada [por exemplo o pensamento de Nietzsche, de Foucault, de Derrida].

(Foucault) é um ironista que não está disposto a ser liberal, o segundo (Habermas) é um liberal que não está disposto a ser um ironista. Ou, a partir de um outro ângulo, considerando as posições contrastantes de Foucault e Habermas face a Nietzsche: Foucault é um devedor e continuador de Nietzsche, e Habermas considera Nietzsche como o guia para um caminho sem saída, não obstante o mérito da sua oposição à filosofia da subjectividade.

Aos olhos de Habermas, com Nietzsche a (positiva) crítica da Modernidade deixou pela primeira vez de ter qualquer conteúdo emancipatório e para Habermas, que é assumidamente marxiano ao equacionar o mérito filosófico com o mérito das implicações políticas de uma filosofia<sup>565</sup>, esta é uma consequência inaceitável. Para Habermas, o teor político de um pensamento, a sua aposta numa normatividade, é a pedra de toque que nos permitirá distinguir aquilo que é filosofia daquilo que é ou tende a ser outro tipo de discurso, artístico ou científico, por exemplo. E a ausência de conteúdo emancipatório da crítica da modernidade, que foi o legado de Nietzsche aos seus herdeiros (Heidegger, Adorno, Bataille, Derrida, Foucault), foi e continua a ser politicamente desastrosa. Mesmo Rorty, cuja visão destes autores é muito diferente da de Habermas, subscreveria o diagnóstico do politicamente desastroso, ou pelo menos politicamente inútil. Rorty diria que estes são “filósofos privados”, no sentido em que “O tipo de autonomia que ironistas como Nietzsche, Derrida ou Foucault procuram é o tipo de coisa que não poderá ser incorporada em instituições sociais”<sup>566</sup>. Para Rorty, a “autonomia” não é alguma coisa que todos os seres humanos tenham dentro de si e que a sociedade possa soltar deixando de reprimi-los: é uma coisa que certos seres humanos particulares desejam atingir pela auto-criação e que alguns (poucos) de facto atingem. É exactamente por isso o desejo de ser autónomo não pode ser relevante para o desejo liberal de *evitar a crueldade* (é a definição de liberal que Rorty retoma de Judith Shklar). Rorty pensa que talvez não consigamos nunca conjugar aquilo que queremos em termos sociais e políticos com aquilo que pensamos e queremos privadamente<sup>567</sup> (pensa que devemos até, seguindo Isaiah Berlin<sup>568</sup>, deixar de acreditar que os todos valores que têm regido vidas humanas poderiam ser em última análise compatíveis e deriváveis entre si). Sobretudo a suspeita de que o ironismo em filosofia não ajudou em nada, pelo contrário, o liberalismo como filosofia política desejável, é perfeitamente fundada para Rorty<sup>569</sup>.

<sup>565</sup> Cf. Rorty 1990, p. 83.

<sup>566</sup> Rorty 1990, p.85.

<sup>567</sup> Cf. “Private irony and liberal hope” e “The contingency of a liberal community”, in Rorty 1990.

<sup>568</sup> Cf. Berlin 1969, *Four Essays on Liberty*.

<sup>569</sup> Por isso a preocupação de Rorty é mostrar que o ironista não será necessariamente um anti-liberal. No entanto há problemas “elementares”, quase problemas de forma para um ironista liberal: “Hoje em dia muitas

Em suma, estão em causas as relações entre ironia, moralidade e política, e neste quadro se Rorty pode compreender o que Derrida faz, Habermas não pode, ou não deve fazê-lo, porque isso invalidaria todo o seu intuito, que é o de uma particular recuperação do projecto Moderno de Ilustração e Emancipação mesmo após a crítica da filosofia da consciência que lhe esteve historicamente ligada.

Uma vez que evoquei Rorty para falar de Habermas, dando agora a voz a Habermas, é preciso notar que este diria, acerca do perfil de Rorty como filósofo, que Rorty permitiu que “os sóbrios *insights* do pragmatismo fossem enevoados por um *pathos* nietzschiano de *Lebensphilosophie* que cumpriu a viragem linguística”<sup>570</sup>...

## 2. Como pode uma concepção argumentativa e por isso conflitual de razão supor um acordo?

Segundo Habermas, a vontade e a identidade individuais formam-se naturalmente numa comunidade de comunicação sob obrigação de cooperação e portanto apenas são concebíveis mediante a acção social. É assim, através de uma acção social que é sempre já racionalidade de forma comunicacional e que está “primeiro”, que se configura a alternativa às concepções teóricas de sujeito que desagradam a Habermas: o ou/ou entre sujeito-observador e sujeito-acontecido, entre posição extra-mundana de um eu transcendental e posição intra-mundana de um eu empírico, é substituído pela intersubjectividade linguisticamente gerada, na acção comunicacional. “Acção Comunicacional” é portanto o nome da acção social coordenada através dos actos de fala nos quais e pelos quais os falantes pretendem à verdade, rectidão e sinceridade<sup>571</sup>.

É dos pressupostos desta prática argumentativa, que num sentido geral cobre toda a acção social, que Habermas se propõe derivar a peça central do seu projecto ético-político, o qual é, na verdade, uma reformulação do Imperativo Categórico Kantiano. Se o imperativo kantiano afirma “age de modo a que a tua acção seja regida por um princípio universalizável a todos os seres racionais”, a versão habermasiana será qualquer coisa como: “age de modo a que a acção seja regida por um princípio obtido por

*peçoas tomam como certo que um gosto pela desconstrução [...] é um bom sinal de falta de responsabilidade moral. Assumem que a marca do intelectual moralmente confiável é uma espécie de prosa transparente, limpa, não demasiado consciente de si – exactamente o tipo de prosa que nenhum ironista auto-criador gostaria de escrever”, Rorty 1990, p. 89.*

<sup>570</sup> Cf. a referência do próprio Rorty, em Rorty 1990, p. 66.

<sup>571</sup> Esta descrição não é propriamente uma profissão de fé ingénua de Habermas, antes pretende ser a explicitação de uma normatividade em acto no uso de linguagem, normatividade no sentido em que, por exemplo, podemos afirmar que não poderia haver “mentira” se não houvesse “verdade”.

um consenso inter-subjectivo, consenso cujas condições de possibilidade estão sempre já inscritas na acção comunicacional”. I.e. para Habermas só podem pretender validade asserções que pudessem contar com o assentimento de todos os seres racionais afectados se estes participassem numa deliberação.

Habermas interpreta portanto o imperativo categórico kantiano não apenas como uma máxima de acção para um indivíduo isolado deliberando moralmente, mas como um princípio de justificação, que nos permite chamar válido àquilo que todos poderiam pensar ou querer.

Em suma, Habermas tem que poder chegar a um ponto de vista a partir do qual o uso de linguagem orientado para o entendimento aparece como um modo principal, originário do uso de linguagem e todos os outros aparecem como derivados, parasitários. Encontra-o na análise deste “ponto de vista moral” inscrito nas estruturas da interacção.

A existência de uma fundamentação discursiva transcendental<sup>572</sup> é a chave da ética habermasiana. Esta será uma investigação racional de argumentos, um procedimentalismo, que supõe uma prioridade do justo sobre o bom, i.e. do argumentável sobre a virtude, uma ética formalista que estabelece regras ou procedimentos que delimitam o modo como um conflito prático moralmente relevante pode ser julgado imparcialmente, de um ponto de vista moral sem se pronunciar quanto a conteúdos. Por isso mesmo, muitas das objecções que comunitaristas e feministas lançam à teoria ética e política de J.Rawls – e que são afinal contestações contextualistas ao apelo à universalidade e ao formalismo kantiano em ética – podem ser também lançadas a Habermas. Ambos apresentam projectos universalistas que pretendem assegurar a possibilidade de legitimação ética e política e, na ausência de um acesso directo a um centro pensante, a um “todos”, a uma “humanidade”, pode sempre ser lançada a desconfiança, pode ser feita uma hipótese sociológica sobre o verdadeiro autor dessa pretensa “universalidade”. I.e. face a um apelo universalista, pode-se sempre perguntar “Universal, mas de quem?”

A ética discursiva habermasiana, mesmo sendo uma ética mínima<sup>573</sup>, é um projecto de Ilustração (o que claro, justifica os diferendos em que Ha-

<sup>572</sup> Vou ignorar os problemas que Habermas tem com esta palavra, pois ela é a que melhor descreve as suas intenções.

<sup>573</sup> No sentido em se afirma que, se bem que possamos investigar argumentos morais, só as questões de justiça e não as questões de conteúdo são racionalizáveis (cf. J. Rawls, *A Theory of Justice*, para o lugar de referência de posições deste tipo). Neste resíduo está todo o problema: as questões de justiça podem supostamente fazer apelo a um princípio de universalização, as outras, avaliativas, devem ser [só podem ser] ponderadas de acordo com a tradição de uma forma de vida concreta. Este conceito de ética-quase-ciência não pode afectar, pesar em ou determinar qualquer conceito de vida boa. Evidentemente que o pano de fundo aqui presente é a questão “o que fazer com o pluralismo?” (tomado como dado nas sociedades contemporâneas) e sobretudo o que fazer com as concepções de mundo, de entre as várias concepções irreduzíveis, que não são auto-irónicas no sentido de Rorty. Note-se que a actual posição de Rawls não é tão simplesmente

bermas se envolveu há alguns anos com defensores da pós-modernidade, como J. F. Lyotard) e é correlativa, em termos de pensamento político, de uma não abdicação do projecto da modernidade como racionalização e autonomização, como Ilustração e Emancipação individual e social, maugrado as “reificações”, os usos não livres de constrangimento da comunicação, sempre existentes lateralmente.

O fundamento das pretensões universalistas de Habermas é portanto o estado de coisas pragmático universal, que a teoria da linguagem e da racionalidade descobre na acção humana e se encarregará de reconstruir e ilucidar. O estado de coisas pragmático universal é o princípio da ética discursiva<sup>574</sup>, e culmina numa teoria da sociedade, da racionalidade e da modernidade, que se vê a si própria na sequência da Teoria Crítica de Frankfurt mas não se permite a meditação civilizacional pessimista que caracteriza por exemplo Adorno e Horkheimer, os quais, rendidos ao diagnóstico de Weber, vêem as sociedades modernas racionalizadas como totalmente administradas e nada ilustradas, dominadas pela “Racionalidade Instrumental”, tendo-se assim gorado as expectativas de Marx quanto ao potencial emancipatório do progresso tecnológico, que teria antes tido como efeito irónico a imobilização das forças sociais disruptivas que supostamente libertaria.

O desvio habermasiano em relação ao pessimismo contemplativo dos anteriores representantes da Escola de Frankfurt implicará nomeadamente um afastamento das categorias do trabalho e da reificação como instrumentos teóricos básicos numa teoria da sociedade: como se disse, com Habermas, a comunicação substitui o trabalho como acção humana arquetípica e possibilidade de fundamento da teoria social (Cf. o *Excursão sobre o envelhecimento do paradigma da produção* - i.e. na filosofia da praxis)<sup>575</sup>. É portanto a Teoria da Acção Comunicacional que fundamenta simultaneamente os discursos das ciências sociais e a solução habermasiana para as aporias do discurso filosófico da modernidade, quer dos críticos da razão com o olhar voltado para o que Habermas chama o ‘extra-quotidiano’, como Heidegger, Bataille, Derrida, quer daqueles que continuam a ter como preocupação central uma teoria da sociedade e da modernidade, como Adorno ou Marx.

*universalizante ou idealizante como a de Habermas. Veja-se nomeadamente a noção de overlapping consensus, que supõe a inexistência de um fundamento comum, o acordo sobre a existência de desacordo. Encontra-se evidente aqui o próprio dilema central da filosofia moral e política contemporânea: o dilema do universalismo versus contextualismo.*

<sup>574</sup> Assim chamada na medida em que a ética se vê como disciplina reconstrutiva das condições de possibilidade dos procedimentos gerais de discurso, encontrando os parâmetros do “consenso motivado racionalmente” e da “Situação Ideal de Fala” (i.e. a comunicação racional livre de coerção, entendida como elemento regulador).

<sup>575</sup> Habermas 1984, p.78.



Precisando os termos desta Teoria da Acção Comunicacional, Habermas pensa que quando um falante e um ouvinte se entendem acerca de algo no mundo eles se movem dentro do horizonte do Mundo da Vida comum, “um pano de fundo intuitivamente conhecido, não problemático, indesmestrável e holístico”<sup>576</sup>. Cada situação de discurso, da qual se extrairá a normatividade acima referida, é um excerto desse Mundo de Vida que é o contexto e fornece os recursos para a compreensão. O conceito de Mundo de Vida, de origem fenomenológica, e também inspirado no pensamento de Wittgenstein acerca da certeza, é um complemento essencial ao conceito de Acção Comunicacional. O conceito de Mundo da Vida é de resto um bom exemplo do ecletismo de meios teóricos, filosóficos e sociológicos, de Habermas, na medida em que é um autêntico mosaico: tanto é usado para nomear a ideia de solo inexplicitável-na-totalidade de qualquer saber explícito, ideia que é reportada ao pragmatismo, a Wittgenstein e a Husserl, como é usado para nomear aquilo que deve ser irreduzível aos efeitos de distorção da modernidade como racionalização, apoiando-se então na *Sittlichkeit* (eticidade) hegeliana. Para falar deste último perigo, Habermas refere-se à “colonização do Mundo da Vida pelos sistemas, dentro daquilo que considera ser um necessário conceito duplo de sociedade: as sociedades são Mundos da Vida e Sistemas.”<sup>577</sup> Os “sistemas” são domínios de acção auto-suficientes formalmente organizados, tais como o estado e a economia. É uma característica da modernidade o ganho de autonomia dessas organizações e essa autonomia redundante, na linguagem de Habermas, numa indiferença à sociedade, cultura, personalidade (características do Mundo da Vida) e numa desvinculação de quaisquer constrangimentos normativos. I.e. a diferenciação dos sistemas em relação ao Mundo da Vida supõe qualquer coisa como uma ‘neutralização ética’ do funcionamento racionalizado.

Habermas enquanto sociólogo – e Habermas pretende ser acima de tudo um teórico das sociedades contemporâneas – tem necessidade da “Acção Comunicacional” como meio através do qual o Mundo da Vida se reproduz. Essa reprodução nomeia funções que transcendem a perspectiva do actor e que devem poder ser conceptualizadas: propagação de tradições, integração de grupos por normas e valores, socialização das gerações vindouras.

<sup>576</sup> Habermas 1984, p. 278.

<sup>577</sup> Estas duas perspectivas do estudo das sociedades, a do Mundo da Vida e a dos Sistemas, seriam, se consideradas cada uma isoladamente, estratégias unilaterais, uma internalista, cega à auto-suficiência dos sistemas, outra externalista, observadora a partir de fora, cega ao estatuto das fixações sociais como produções normativas, com autor, susceptíveis de interferência e correcção.

A Acção Comunicacional, o Mundo da Vida e a “invasão do Mundo da Vida pelos sistemas”,<sup>578</sup> são conceitos indispensáveis na teorização e fundamento da metodologia das ciências sociais – naquilo que é, traduzido para o campo disciplinar da filosofia e para o domínio das propostas normativas, uma ideia kantiana, hegeliana e marxiana de teoria social, ética e política. A inspiração do pensamento de Habermas é kantiana quando se procura o fundamento do dever-ser numa universalidade formal, numa auto-determinação racional, é hegeliana porque se pretende ultrapassar o “universal abstracto” da moralidade centrada no indivíduo e seguir a realização histórica e social da racionalidade, e é de linhagem marxiana na medida em que pretende que uma análise das patologias sociais, da “reificação” susceptível de ser produzida pela racionalização, conduza à emancipação.

Esquemmatizou-se até aqui o pensamento de Habermas como uma teoria da racionalidade que se estende desde a linguagem à sociedade. Mas o que é exactamente “racionalidade” neste contexto maior da teoria da Acção Comunicacional? Não se trata evidentemente apenas de produção teórica de uma imagem do mundo: chama-se racionalidade de forma comunicacional à disposição por parte dos sujeitos falantes e actuantes de se envolverem em modos de comportamento para os quais há boas razões ou fundamentos, o que implica a capacidade de orientação em relação a exigências de validade que assentam sobre o reconhecimento intersubjectivo. O critério desta racionalidade-de-forma-comunicacional é a existência ou não existência de um procedimento argumentativo de liquidação de exigências ou pretensões de validade, seja essa validade verdade (proposicional), rectidão (moral) ou sinceridade (estética). O critério aponta sempre, finalmente, para o acordo racionalmente motivado via a possibilidade da argumentação.

Esta posição de Habermas supõe, para se sustentar, que exista algo como uma ‘motivação racional’ –o que é problemático—mas pode conceder-se que, se a definição se sustenta, a teoria da racionalidade sai definitivamente da esfera da consciência, não faz apelo a uma interioridade transparentemente auto-acedida, a-histórica e a-social. A teoria da significação e da racionalidade é constitutivamente alargada à pragmática, torna-se teoria da significação e da racionalidade “socializadas”, ultrapassando

<sup>578</sup> *A invasão do Mundo da Vida pelos sistemas é traduzível como o perigo para a Eticidade das racionalizações “reificantes”, por exemplo visíveis na distância entre a prática quotidiana e os discursos científicos, jurídicos, políticos, de crítica de arte, i.e. os discursos racionais cuja autonomia é o adquirido da Modernidade, mas que podem ser desligados dessa “totalidade de nível superior ao indivíduo” a que Hegel chamava Eticidade. É essa eticidade hegeliana, o dever-ser des-subjectivado, o que é o núcleo do conceito de Mundo de Vida. Como foi dito, não é a única acepção deste conceito, que Habermas faz passar por várias metamorfoses, na tentativa de cobrir teoricamente a natureza de uma racionalidade social. Cf. Redondo 1991, p.17.*

a fixação na função de língua que consiste em construir factos. É muito importante para Habermas mostrar que a validade não é redutível legitimamente (e Habermas nunca tem problemas com apelos à legitimidade) ao valor de verdade. Cada acto de fala elementar pode ser contestado nos diferentes aspectos de validade: a contestação pode dirigir-se não apenas à verdade, como à rectidão, como à veracidade da pretensão. Habermas só quer pois falar do uso que é feito de enunciados em asserções e nunca de enunciados “despidos”: se de um enunciado se pode discutir a verdade ou falsidade quanto a uma asserção deverá sempre falar-se de justificabilidade nos vários sentidos apontados.

Em suma, embora partilhando, nas suas afirmações explícitas, das críticas da Escola de Frankfurt à Racionalidade Instrumental como patologia social caracteristicamente moderna, Habermas mantém no entanto e pretende defender a todo o custo uma confiança *sui generis*: a confiança na necessidade e possibilidade de legitimar racionalmente. Em termos da história da reflexão social de linhagem marxiana da Escola de Frankfurt, Habermas é o ponto em que, porque a comunicação substituiu o trabalho como categoria central para a reflexão social, apesar de toda a *Entzauberung*<sup>579</sup> da fundamentação da ordem social a razão albergada na linguagem fornece intacto um novo lugar para a legitimação.

A própria existência da ética discursiva mostra que é possível uma passagem à reflexividade da acção social que explicita o potencial racional em qualquer situação de discurso (pretende-se que seja esse o estatuto da própria Teoria da Acção Comunicacional) e que é concebível a ‘pesagem’ de uma sociedade como superior a outra nesses termos. É suposta portanto uma filosofia da história, a defesa da existência de um trajecto, de um progresso de racionalização nas práticas humanas, de uma aquisição, de um mais. Este progresso é o progresso da racionalização na história e é um, tem um sentido específico.

Em suma, o projecto Habermasiano pode esquematizar-se como:

Uma teoria da linguagem e da racionalidade → uma ética → uma teoria da sociedade e da modernidade

Este esquema permite a Habermas acreditar inabalavelmente que há discursos melhores do que outros discursos, por exemplo a filosofia seria supostamente “melhor” do que a literatura, ou pelo menos teria outras pretensões racionais, ou não seria condenável o nivelamento da sua diferença genérica

Mas esta teoria da linguagem e da racionalidade – i.e. a Teoria da Acção Comunicacional – que evidencia segundo Habermas uma normativi-

<sup>579</sup> Desencantamento [termo de M. Weber].

dade sempre já presente nas práticas e pretende usá-la como fundamento para a teoria ético-política, sustentar-se-á de modo a poder sustentar todo o edifício? É isso que se pretende analisar no confronto das posições de Habermas com as de Derrida.

### 3. Universalidade ou diferença incontornável?

Vou considerar em seguida as linhas gerais do pensamento de Derrida sobre o que é pensar e usar linguagem. Estas constituem uma alternativa incontornável às posições de Habermas, que como se viu se fundamentam também elas numa teoria da linguagem. A minha intenção é compreender como pretendo Habermas ter neutralizado as posições de Derrida no *Excurso sobre o nivelamento da diferença genérica* a partir das teses centrais do próprio *Discurso Filosófico da Modernidade*.

Quem afirma que o nivelamento da diferença genérica entre filosofia e literatura não deve ser feito está a sustentar simultaneamente que essa diferença genérica existe e que o discurso da filosofia é melhor, mais importante, mais valioso do que o da literatura ou da crítica literária. Com que pressupostos o faz? Em função de que razões se pode afirmar assim que certos discursos “pretendem mais” – que por exemplo as ciências naturais, concebidas não enquanto modo positivista de conhecer mas enquanto “campo exemplar da fixação não constrangida das crenças”, para falar a linguagem de Habermas, são melhores do que uma arte? Para mantermos clara uma referência, o neo-pragmatismo de linhagem rortyana não responderia a esta pergunta do mesmo modo que Habermas<sup>580</sup>. E mesmo admitindo a diferença de pretensões haverá razão para supor uma divisão de terreno clara e distinta que nos permita afirmar “isto é apenas arte” mas “isto já é racional e portanto mais importante”?

A posição de Habermas baseia-se num pressuposto que Rorty não aceita e a que chama em *A Filosofia e o Espelho da Natureza* o pressuposto da comensurabilidade de todos os discursos (artísticos, éticos, científicos,...). É preciso portanto, antes de mais, não apagar a questão subjacente ao carácter bem ou mal fundado do apagamento de “diferenças genéricas”: estamos ou não estamos preparados para trazer para debaixo de um mesmo conjunto de regras a física, a literatura, a religião, etc, de modo a ordenar e hierarquizar as suas práticas e discursos? Isso suporia segundo Rorty poder enunciar esse conjunto de regras, para podermos dizer que sabemos para onde vamos, o que queremos, o que somos e que a resposta

<sup>580</sup> Cf. Rorty 1988, *A Filosofia e o Espelho da Natureza* [capítulo *Epistemologia e Hermenêutica*] e Rorty 1990, *Contingência, Ironia, Solidariedade*

é *uma*. Então poderíamos afirmar que há uma razão, uma civilização, um progresso. Teríamos, em suma, uma *filosofia da história*. Há quem pense que a temos, há quem pense que não a temos: esta é uma maneira de nomear o diferendo entre Habermas e Derrida.

As teses em jogo no *Excurso* são então as seguintes: Habermas afirma que Derrida, “no trabalho revolucionário geral da desconstrução, que visa a destruição das hierarquias, a derrocada dos contextos de fundação e das relações conceptuais de dominação”<sup>581</sup> nivela a diferença genérica entre filosofia e literatura, e para isso tem um interesse particular em inverter a primazia aristotélica da lógica face à retórica. Isto porque Derrida ambiciona o “acesso esotérico à verdade como Heidegger, mas não o quer como privilégio, seja de que instância fôr”<sup>582</sup>. “Não haverá nome único, nem que seja o nome do ser”, diz Derrida<sup>583</sup>. Então, se apenas se pode falar em contradição à luz de exigências (lógicas) de consistência, se a prioridade do lógico face ao retórico desaparecer, Derrida, que não quer ultrapassar sobranceiramente como faz Heidegger, a objecção de inconsistência pragmática, poderá torná-la inútil. É assim que Derrida se permite tratar obras de filosofia como se fossem literatura, e faz os textos falarem contra ou à margem dos seus autores à custa do excedente retórico. A deslocação para o campo da retórica seria a saída derridiana da aporia da auto-referência, e seria mais um exemplo, para Habermas, da “hostilidade à razão, praticada na terra de ninguém entre argumentação, narrativa e ficção”<sup>584</sup>.

Mas Habermas crê que isso é ignorar o fundamento (que é um fundamento quotidiano, prático) da distinção entre fictício e não fictício, entre retórica e argumentação. O discurso quotidiano não é discurso fictício ou retórico, está sempre sujeito a obrigações de comprovação (das quais Habermas extrai a normatividade acima referida). Segundo Habermas, deve-se por isso, associar respectivamente retórico e literário, argumentativo e lógico (ou jurídico). Há que defender a todo o custo de qualquer contaminação a literalidade do literal, a seriedade do sério, a não ficcionalidade do não-ficcional.

Mas o que provem da figura matricial de Nietzsche, da raiz de tantos males teóricos e tanta influência nefasta (para Habermas) sobre os críticos da razão? Nietzsche ensinou-nos que não há na linguagem literalidade a que o figurativo possa ser reconvertido, mostrou-nos a ‘natureza’ retórica da linguagem. Daí a curiosa posição de Nietzsche quanto a razão e linguagem:

<sup>581</sup> Cf. Habermas 1984, p.180.

<sup>582</sup> Habermas 1984, p.180.

<sup>583</sup> Cf. *La Différance* [A Diferença], in Derrida 1972 [as referências de páginas que se seguem dizem respeito à tradução portuguesa indicada na Bibliografia].

<sup>584</sup> Habermas 1984, p.262.

ela permite-nos dizer que a razão é linguagem, possibilidade de ficção, mas nunca no sentido de argumentação, antes sempre e só retórica.

“Nietzsche-contra-Heidegger” é aliás o teor da afirmação acima citada de Derrida: “Não haverá nome único, nem que seja o nome do ser”. E Derrida continua: “É necessário pensá-lo sem nostalgia, isto é, fora do mito da língua puramente maternal ou puramente paternal, da pátria perdida do pensamento. Pelo contrário, é necessário afirmá-lo, no sentido em que Nietzsche põe a afirmação em jogo, num certo riso e num certo passo de dança”<sup>585</sup>. Então só podemos - e é isso que devemos fazer - desconstruir os contextos de fundação, as hierarquias a ela associadas: oral/escrito, retórica/lógica, ficcional/verdadeiro, lúdico/sério, ou em suma, (supostamente) secundário ou cópia versus original, sem presunção de remontar a uma originariedade, que não existiu. É esse o sentido da inscrição da alteridade na mesmidade, de não-posses numa presença a si consciente, que os conceitos derridianos de escrita, rastro, diferença procuram cobrir.

Retomando a citação de H. Bloom em epígrafe, o que é que um poema possui ou cria? Nada, se não houver outro poema. Esta não é uma afirmação restringível à crítica literária entendida disciplinarmente como textos que falam sobre textos: é o núcleo da posição de Derrida quanto a pensamento e sentido.

Mas então como escrever, como fazer, se o que se quer mostrar é a inexistência de um lugar de presença a si originário e originador do fazer, pensar, escrever, historicamente chamado razão e consciência? Derrida-escritor é ele próprio um leitor, os seus textos são textos sobre textos, textos sobre Husserl, sobre Rousseau, sobre Lévinas, sobre Heidegger, sobre Foucault, sobre Genet, Joyce, Blanchot, ... Quando se trata de textos, de pensamentos, são impensáveis a confusão absoluta assim como a compreensão absoluta, o princípio absoluto como o fim absoluto: estamos a meio, entre. A esse “entre” chama-se texto, escrita, tradução não terminável. Um texto nunca está fechado: não sei quais são os meus destinatários futuros, mas o que eu escrever poderá ser lido depois da minha morte. Aliás, cada uma das minhas palavras só pode ser compreendida agora porque é já uma repetição que eu posso supor se continuará a repetir no futuro. Para poder haver pensamento, para poderem existir efeitos de identidade, por exemplo a mesmidade de uma palavra reconhecida, deve haver signo, e o haver signo supõe repetição, e a repetição supõe inscrição, passar por fora da consciência, exterioridade (Derrida diz “escrita”). Nenhuma “fala” seria possível sem essa “escrita”, essa escrita precede a distinção entre língua e fala, é um nome para a própria possibilidade de signo e de pensamento.

<sup>585</sup> Derrida 1972, p. 68.

Seja o meio de exterioridade aquele que fôr — sons, traços, suportes eletrônicos — será exterioridade, alteridade, ausência de um intencionador animado, vivo, humano. Daí a ligação escrita-morte: “A escrita comunica o meu pensamento longe de mim, na minha ausência, mesmo depois da minha morte. Quando ler a minha carta o destinatário sabe que eu posso estar morto, durante o tempo mesmo mínimo entre o momento de terminar a carta e o momento da sua recepção” (...) “Não é necessário que eu esteja morto para que possas ler-me, mas é necessário que sejas capaz de me ler mesmo se eu estiver morto”<sup>586</sup>. Morte é um nome genérico para esta ausência necessária naquilo que é pensado, escrito. E quem quer que fale, quem quer que escreva estará assim, entre texto e texto, inscrito, prometido à repetição, à disseminação e não à restituição. Aliás, eu “leio-me” e “escrevo-me” a mim próprio, no sentido em que os efeitos acima descritos, do diferir, da não presença, se aplicam ao próprio acontecer do meu pensamento, sem necessidade de evocar outro comunicador. Por isso o querer dizer centralizado da hermenêutica suporia a ingenuidade da recuperação, do autor, do originário, da fonte, i.e. a crença na possibilidade absoluta de não violência. Contra isso se coloca a afirmação de Derrida segundo a qual nenhum acontecimento é o acontecimento primeiro, originário, e todos os textos e pensamentos são textos-eventos e nunca monumentos definitivos. A interpretação é interminável e está sempre já aí, a disseminação do sentido é incontrolável.

Estas teses redundam na prática que é comumente coberta pela expressão “crítica da metafísica”, porque: “A metafísica consiste desde logo em excluir a não-presença determinando o suplemento como exterioridade simples, como pura adição ou pura ausência (...) O paradoxo é que se anula a adição, considerando-a como uma pura adição. Aquilo que se acrescenta não é nada, visto que se acrescenta a uma presença plena, à qual é exterior.(...) A cultura vem acrescentar-se à natureza, o mal à inocência, a história à origem, etc (...) É o mito do apagamento do rastro”<sup>587</sup>.

Nos seus primeiros textos mais académicos e sérios (os de 1967<sup>588</sup>, 1972<sup>589</sup>) Derrida preocupou-se com a defesa destas teses através das suas figuras conceptuais da não-presença (o rastro, a disseminação, a diferença, a escrita) num diálogo ainda bem explícito com a tradição filosófica: nomeadamente com Heidegger e a crítica à metafísica da presença<sup>590</sup>, com Husserl e a possibilidade de fenomenologia como descrição das vivências

<sup>586</sup> Bennington & Derrida 1991, p. 51.

<sup>587</sup> Derrida 1967a, *De la Grammatologie*, p. 234.

<sup>588</sup> Nomeadamente *De la Grammatologie, L'Écriture et la Différence, La Voix et le Phénomène*

<sup>589</sup> Por exemplo *Marges de la Philosophie*.

<sup>590</sup> Cf. *La Différance*, em Derrida 1972.

da consciência<sup>591</sup>. Depois, continuando a ler e escrever sobre textos de filósofos (e não só) os seus textos passaram a consistir cada vez mais naquilo a que Rorty chama “alusões privadas”<sup>592</sup>, fruíveis e deslindáveis apenas por quem é já possuído por esses textos, pensamentos, obras, só podendo agir sobre pessoas em quem Heidegger ou Hegel, ou Freud, ou outros, já fizeram alguma coisa.

Voltando a Habermas, ele pensa que Derrida não pode escapar assim tão facilmente, tornando inútil a objecção de inconsistência pragmática através do que Habermas considera ser um desvio para o campo da retórica e assim um nivelamento da diferença genérica entre textos. Habermas gostaria de mostrar que o pensamento de Derrida tem fontes e matizes religiosas, embora, é claro, Derrida não queira pensar religiosamente, nem afirmar Deus. Habermas gostaria de mostrar isto<sup>593</sup> porque quer mostrar que Derrida se aferra a uma cegueira irracional perante a incrustação laica da racionalidade nas instituições, nas práticas, nos discursos, que a modernidade produziu. Derrida não quer, segundo Habermas, ver a razão apesar de tudo existente num mundo irracional. E para vê-la seria preciso ver o que Derrida não vê: que há regras a que estão subordinados apenas certos discursos, que pretendem mais do que outros discursos, precisamente sustentar uma argumentação racional num contexto livre de coerção. E falar de ‘texto’, fazendo como se os textos tivessem perdido as suas diferenças específicas, tão custosamente ganhas na história, e de Escrita para todo os usos de linguagem, inclusivamente os supostamente autónomos (das ciências, moralidade, literatura, artes, áreas cuja autonomia é precisamente um ganho do processo de racionalização da modernidade) é uma estratégia errada, que produz cegueira social e política<sup>594</sup>. É aliás esta estratégia errada que sustenta a entusiástica recepção a Derrida pelos críticos literários na América, nomeadamente os de Yale, como Bloom e de Man. Falar por

<sup>591</sup> Cf. Derrida 1967c, *La Voix et le Phénomène*. Para Derrida a presença do presente seria derivada da repetição e não o inverso.

<sup>592</sup> Rorty 1990, p. 122.

<sup>593</sup> Habermas chega a afirmar explicitamente que “O trabalho da desconstrução serve [...] para a renovação inconfessada de um diálogo com Deus”, cf. Habermas 1984, p. 177.

<sup>594</sup> O outro “grande culpado” da degenerescência culturalista da academia americana [uma vez que Rorty e Derrida têm neste momento pendente sobre si a acusação de responsabilidade pela pobreza intelectual, falta de seriedade e relativismo vazio de grande parte dos estudos culturais nos Estados Unidos] não tem dificuldade nenhuma em ser o primeiro a afirmar este perigo, [cf. Richard Rorty 1998, *Achieving our country - Leftist Thought in Twentieth Century America*.] Fazendo a história da esquerda nos Estados Unidos, Rorty quer fazer uma acusação à “esquerda cultural” actual, que liga à disciplina académica a que normalmente se chama no mundo anglo-saxónico “Teoria” ou “Crítica Literária”, que faz nomeadamente uso do seu próprio pensamento, bem como das obras de Derrida, Nietzsche, Foucault, etc. Por uma vez Rorty acusa o uso académico dos autores que admira, pois pensa que aquilo a que chama a “esquerda cultural” não apenas tornou a esquerda irrelevante para a maior parte da população como joga nas mãos da “direita”, ao ter abandonado as questões económicas, por não querer falar de riqueza e distribuição de riqueza e se dedicar sobretudo a mind-sets, maneiras de pensar, configurações culturais (i.e. toda a área do politicamente correcto). Ao pretender ter assim, i.e. culturalmente, relevância política submeter-se-ia antes à estratégia dos verdadeiros decisores, que seria a de manter os políticos e intelectuais ocupados com questões culturais e nacionais uma vez que a prerrogativa das decisões económicas verdadeiramente relevantes pertence a uma classe superior cosmopolita indiferente a essas questões e que não tem necessidade de qualquer sentido de comunidade.



toda a parte de diferença, disseminação, rastro, e estender ao tratamento de todos os textos, sem diferença de género, a tarefa de desconstrução da história da metafísica, é uma fraude e uma menorização daquilo que é (ou deve ser) filosofia, ética, política.

Isto para Habermas, porque os “seguidores” de Derrida poderão ver a sua prática como uma contra-política à exclusão, repressão e marginalização inscritas nos usos das linguagens, poderão ver a desconstrução como uma prática não terminável de desmontagem da pretensa neutralidade e originariedade de qualquer discurso, poderão ver a desconstrução como uma prática de impossibilitação da *Aufhebung*, da redução da alteridade. Mas talvez falar de seguidores de Derrida seja um contra-senso, talvez só deva haver um Derrida<sup>595</sup>.

É no entanto possível que em vez da menorização de um pensamento ético e político que se vê como só podendo ser iluminista e emancipador, Derrida mostre outra visão, menos susceptível de *happy-end* do que a de Habermas, do que se passa em ética e em política. Uma política concreta como a que os seus defensores anglo-saxónicos gostariam de ver proposta é impossível a Derrida em virtude da sua própria visão da teoria, embora a desconstrução possa ser considerada como “o mais radicalmente ético e político dos discursos”<sup>596</sup>, precisamente por ser uma prática (interminável) de impossibilitação de redução da alteridade. No entanto um ponto de vista que permite apontar certas fraquezas de uma ética e política do tipo habermasiano é possível a partir de Derrida.

É preciso notar que não é sem razão que o feminismo ou a mesmo a teologia se puderam apoderar dos termos teóricos da desconstrução. Nem é vã a aproximação que certos intérpretes do pensamento de Derrida<sup>597</sup> fizeram da sua escrita com a tradição rabínica (Derrida é aliás judeu e aquilo que ele faz já foi interpretado à luz desse facto: aquilo que Derrida afirma acerca de sentido é que a interpretação está sempre já aí e o comentário é infinito e numa certa versão do pensamento judaico o comentário humano é concebido como parte da Torah ela mesma, não há uma revelação, originária, a interpretação humana é parte da revelação divina<sup>598</sup>).

Precisamente os objectos teóricos da teologia, do feminismo, Deus, as mulheres, são uma alteridade que pode supostamente estar a ser reduzida quando é pensada a partir de, e nos termos próprios a, um autor do discurso que é humano e masculino. I.e. esses objectos seriam o que não

<sup>595</sup> Neste sentido não seria possivelmente tão perniciosa a proliferação dos “Habermas” como a dos “Derridas”. Cf. nota anterior e a classificação rortyana de filósofos “privados” e “públicos”.

<sup>596</sup> Bennington & Derrida 1991, *Jaques Derrida*, p. 214.

<sup>597</sup> Nomeadamente Susan Handelman.

<sup>598</sup> Habermas está particularmente interessado nesta interpretação, cf. Habermas 1984, *Exacerbação da filosofia temporalizada da origem*.

tem voz, nem linguagem para tê-la (como falar por exemplo de uma dominação sem o que nos permite apercebê-la como tal?), o que não é uma consciência que a partir de si fala, é “o outro”, o que fica ou ficou de fora do (suposto) consenso de uma racionalidade humana que configura as possibilidades de uma civilização. É neste tipo de situações que o reconhecimento derridiano de um “além” pode ser utilizado (por exemplo alguns dos seus leitores teólogos<sup>599</sup> chegam a afirmar que “sem a fé, Derrida teria razão”).

Se a acusação de Habermas a Derrida consiste em considerar que “o sério é diferente do não-sério, a ficção é diferente da não ficção, a razão é diferente da não razão e Derrida quer ignorar tudo isso”, poder-se-ia perguntar, conhecendo a valorização habermasiana do consenso na teoria da linguagem, da racionalidade, da ética e da política se não será tal pressuposto apenas um preconceito. Mais: não será este um preconceito perigoso? A noção de consenso não estará subterraneamente minada por uma lógica identitária de supressão da alteridade? Não será o consensualismo uma versão *soft* do fundacionalismo?<sup>600</sup> (os filósofos chamam fundacionalismo à possibilidade de apelo a um princípio primeiro auto-subsistente, do qual decorra o edifício teórico, todo a nosso pensamento; esse princípio seria o “começo absoluto”, aquilo que Derrida diz que não existe).

Por exemplo, em possíveis sociedades humanas futuras, quantas *formas de vida racionais* restarão, se forem supostas ser racionais de acordo com a análise de Habermas? Note-se que a modernidade como racionalização, com a sua autonomização de áreas de discurso e a sua independência em relação ao Mundo da Vida de esferas de acção racionalizada, essa modernidade cujo projecto Habermas quer continuar, está muito especificamente ligada a uma particular civilização.

A ideia derridiana de Rastro, a ideia de um “passado que nunca foi presente”<sup>601</sup>, que marca a relação ao outro como um anúncio e uma dissimulação, é um princípio de resposta à presunção de transparência e centralização racional em ética e em política que orienta o pensamento de Habermas: mesmo que a partir de um paradigma de intersubjectividade se procure evitar as armadilhas da filosofia da consciência e do fundacionalismo, há em Habermas um “mito do apagamento do rastro”.

<sup>599</sup> Cf. Loughlin 1998, *Rains for a famished land*, *Times Literary Supplement*, April 10, para uma recensão a livros recentes de Catherine Pickstock e Graham Ward, teólogos da Universidade de Cambridge. É claro que há em Derrida o reconhecimento de um “para além”, a que Derrida não chama deus, mas um *Sim* [só a fé poderá fazê-lo, chamar Deus e não um nada a esse para além, e a leitura dos teólogos cujo projecto é pensar a teologia na pós-modernidade irá precisamente nesse sentido].

<sup>600</sup> Sampaio da Silva 1997, p. 151.

<sup>601</sup> Esta é uma expressão de E. Lévinas para o outro, no sentido de outro-eu, mas utilizada por Derrida para as figuras da não-presença.

Já as objecções de H. G. Gadamer a Habermas<sup>602</sup> no curioso diálogo de textos entre ambos (iniciado em 1967, com a recensão feita por Habermas a *Wahrheit und Methode*, e terminado em 1972), evidenciam essa característica da personalidade teórica de Habermas que é o desejo de “centralização racional”, a pretensão ao não-resíduo, características que justificam a sua aversão a Derrida. Nessa discussão<sup>603</sup>, Gadamer objecta a Habermas que a presunção da possibilidade de uma transparência resultante do processo reflexivo da tomada de consciência e ao mesmo tempo de uma total abdicação da retórica (o mesmo ponto que está em causa no Excurso pretextado pela filosofia de Derrida, note-se) a favor de um diálogo racional livre de coerção é de um irrealismo impressionante. Habermas parece pensar, segundo Gadamer, que, uma vez tornada transparente, a estrutura do preconceito ele já não funciona como preconceito. Gadamer não quer negar o poder contestatário da reflexão reivindicado por Habermas, apenas duvida da existência de um grau zero do preconceito no pensamento.

E a propósito da recorrente consideração por Habermas da psicanálise como modelo da correcção da comunicação distorcida e por isso como modelo para a teoria social, Gadamer objecta a Habermas que essa aproximação é descabida, que ninguém possui ou poderá possuir o estatuto de médico da sociedade, até porque podem ser muitas as formas de vida com direito ao título de “vida boa”.

A ideia habermasiana de psicanálise como modelo para a correcção da comunicação distorcida pode dar-nos um bom ângulo para o que acima se chamou o desejo de centralização racional de Habermas: se tomarmos a psicanálise nos seus próprios termos, esta é uma concepção “demasiado humana” daquilo que está aí em jogo, que é afinal um não-domínio de si por si que é constitutivo, inultrapassável, é uma condição, e não um desvio provisório e corrigível até uma total e transparente auto-determinação na acção e no pensamento.

Sobretudo, e a partir de Derrida, é possível contestar a crença na possibilidade em absoluto de não violência (na interpretação, no pensamento,

<sup>602</sup> Gadamer, o autor de *Wahrheit und Methode* [1960] e, figura central da hermenêutica, tem aliás curiosíssimos debates com ambos, Derrida e Habermas. A discussão com Habermas inicia-se precisamente com a recensão deste a *Wahrheit und Methode*, na qual, louvando aquilo que a hermenêutica de Gadamer oferece à auto-reflexão metodológica das ciências sociais, Habermas a considera no entanto um “idealismo da linguisticidade”, por esquecer que o contexto objectivo da acção social não é apenas linguagem mas linguagem, trabalho e dominação. Cf. para uma descrição pormenorizada, Sampaio da Silva 1997, Heidegger, Gadamer e o Fundacionalismo, capítulo Habermas e Gadamer, p.139. Para os textos do mais recente debate entre Gadamer e Derrida, cf. Palmer & Michelfelder 1989.

<sup>603</sup> A discussão é acerca de hermenêutica, no sentido não apenas de metodologia das ciências sociais e humanas mas de condição humana da compreensão. O livro de Gadamer apresenta um conjunto de instrumentos teóricos – tais como a “fusão de horizontes” e a “*Wirkungsgeschichtlichebewusstsein*” – que são louvados por Habermas. Este critica, no entanto, na hermenêutica o que considera ser um inevitável conservadorismo político, que só poderá ser obviado por uma Teoria Crítica da sociedade dirigida especificamente àquilo que além da linguagem constitui a acção social: o trabalho e a dominação.

na teoria) que está contida no uso habermasiano do conceito de comunicação e de razão comunicacional, com o seu ponto ideal na noção de consenso obtido em situação de deliberação racional livre de coerção. É certo que a noção de consenso pretende des-subjectivar ou des-individualizar isso a que se chama racionalidade, de modo a poder assim ligá-la a formas sociais emancipadoras. No entanto este movimento não livra a racionalidade habermasiana da acusação de pretender ser sem-outro, sem resíduo.

Essa contestação da crença na possibilidade de uma absoluta não-violência no pensamento, Derrida fê-la, como já foi referido atrás, já em *L'Écriture et la Différence* (1967), quando defendeu Heidegger contra as acusações de Lévinas<sup>604</sup>. Lévinas que é própria epítome do apólogo da não violência e da abertura ao outro como guia do pensamento filosófico. Quando Lévinas quer erguer contra Heidegger a “abertura total ao radicalmente outro”, Derrida diz simplesmente que isso é impossível.

Em suma, a noção de comunicação está, como qualquer outra noção comum, contaminada de múltiplas maneiras quando começa o seu uso teórico. Ora quando se pretende utilizá-la para fundamentar transcendentemente uma ética e uma teoria política como faz Habermas, torna-se necessário apontar claramente o que é que ela apaga, sobretudo quando é frequente ouvir deplorada a provocação do desacordo por Derrida sob o signo da diferença, quando seria preciso trabalhar ética e politicamente para o consenso.

Quanto ao uso de uma ingénua noção de comunicação, Derrida diria que é preciso verificar que “O horizonte semântico que comanda habitualmente a noção de comunicação é excedido ou fendido pela intervenção da escrita, quer dizer, de uma disseminação, que não se reduz a uma polissemia. A escrita não dá lugar em última instância a *uma* decifração hermenêutica, a *uma* descodificação de um sentido ou de *uma* verdade”<sup>605</sup>

Quanto ao pretexto que tomei, o *Excurso sobre o nivelamento da diferença genérica entre filosofia e literatura*, penso que ao contrário do que Habermas aí afirma, Derrida não pretende nivelar nem desnivelar filosofia e literatura (trata-se obviamente de um diálogo de surdos entre Derrida e Habermas, o que é sem dúvida um ‘argumento’ favorável a Derrida...) mas antes mostrar a presunção implícita quando se afirma a possibilidade de separação, de in-contaminação entre filosofia e literatura, metáfora e conceito, exterioridade e consciência, não-presença e presença. Essa presunção é a presunção de um apagamento das condições materiais, e o que Derrida mostra e repete é a presença desse “outro” em qualquer efeito de identidade.

<sup>604</sup> Cf. Derrida, *Violence et Métaphysique*, em Derrida 1967b, *L'Écriture et la Différence*.

<sup>605</sup> Cf. *Assinatura, Acontecimento, Contexto em Derrida 1972, Margens da Filosofia*, p. 434 (itálicos acrescentados por mim).

Em suma, em todos os conceitos da não-presença se constata o carácter radicalmente ético daquilo que Derrida faz (se por ética entendermos deixar ser o que é outro, não o reduzir, não supôr possível reconvertê-lo dialecticamente ao que é pensamento, racionalidade, consciência) e a sua afirmação de uma inumanidade essencial dos actos de pensar e escrever. A diferença, a escrita, o rastro são incontrolláveis, dão-se-nos, não podemos presumir captar em consciência e assumir a direcção das condições de possibilidade do nosso pensamento a um ponto tal que possamos ver a possibilidade de um *continuum* entre linguagem-ética-política a que chamaríamos “racionalidade”. Por isso, não haverá nome único, nem que seja o nome de ‘consenso’...

### Referências

Benhabib, Seyla, Jürgen Habermas, The Philosophical Discourse of Modernity: Twelve Lectures, *The Journal of Philosophy*, vol.LXX-XIV,12,1987, pp752-757.

Bennington, Geoffrey e Derrida, Jacques, *Jacques Derrida*, Paris, Seuil, 1991.

Berlin, I., *Four Essays on Liberty*, Oxford, Oxford University Press, 1969.

Bloom, Harold, *A Map of Misreading*, Oxford, Oxford University Press, 1975.

Carrilho, Manuel Maria, *Rhétoriques de la Modernité*, Paris, PUF, 1992.

Critchley, Samuel & Honneth, Axel, Philosophy in Germany, in *Radical Philosophy* 89, May/June 1998.

De Man, Paul, *A Resistência à Teoria*, Lisboa, Ed.70, 1989.

Derrida, Jaques, *De la Grammatologie*, Paris, Minuit, 1967a.

Derrida, Jacques, *L'Écriture et la Différence*, Paris, Seuil, 1967b.

Derrida, Jacques, *La Voix et le Phénomène*, Paris, PUF, 1967c.

Derrida, Jacques, *Marges de la Philosophie*, Paris, Minuit, 1972 (ed. portuguesa: *Margens da Filosofia*, Porto, Rés, s.d.)

Derrida, Jaques, *La carte postale*, Paris, Flammarion, 1990.

Goodin, Robert & Petit, Philip (eds.), *A Companion to Contemporary Political Philosophy*, Oxford, Blackwell, 1993.

Habermas, J., *O Discurso Filosófico da Modernidade*, Lisboa, Dom Quixote, 1990 (ed. original alemã: *Der Philosophische Diskurs der Moderne*, 1985)

Habermas, J., *Theory of Communicative Action*, London, Heinemann, 1984 (ed. original alemã *Theorie des Kommunikativen Handelns*, 1981)

Habermas, J. *Escritos sobre moralidad y eticidad*, Barcelona, Paidós, 1991.

Loughlin, G., Rains for a famished land, *Times Literary Supplement*, April 10, 1998

Moura, Vítor, As muitas probabilidades da injustiça, *Diacrítica* 11 (1996), 737-765

Palmer, Richard, e Michelfelder, Diane, *Dialogue and Deconstruction, The Gadamer-Derrida Encounter*, New York, SUNY Press, 1989.

Putnam, Hilary, A politics of hope (Richard Rorty, Achieving Our Country), *Times Literary Supplement*, May 22, 1998.

Ralws, John, *A Theory of Justice*, Cambridge, Belknap Press, 1971.

Redondo, Manuel Jiménez, *Kant y Hegel en el pensamiento de Habermas, Introducción*, in *Escritos sobre Moralidad y Eticidad*, Barcelona, Paidós, 1991.

Rorty, Richard, *A Filosofia e o Espelho da Natureza*, Lisboa, Dom Quixote, 1988.

Rorty, Richard, *Contingency, Irony, Solidarity*, Cambridge, Cambridge University Press, 1990.

Rorty, Richard, Posties (*Der Philosophische Diskurs der Moderne* by Jürgen Habermas), *London Review of Books*, 3 September 1987.

Rorty, Richard, *Achieving our country*, Cambridge MA, Harvard University Press, 1998.

Sampaio da Silva, Rui Jorge, *Heidegger, Gadamer e o Fundacionalismo*, Tese de Mestrado, Universidade Nova de Lisboa, FCSH, 1997.







