

Sofia Miguens e Susana Cadilha

Acção e Ética

Conversas sobre racionalidade prática

Lisboa

Colibri

2011

ÍNDICE

INTRODUÇÃO

Sofia Miguens, “ <i>Conversations on Practical Rationality and Human Action</i> ” (2007-2011)	4
1. Origens, relações e continuidades.....	5
2. Um projecto prático sobre racionalidade prática – o <i>script</i> da entrevista.....	11
3. O domínio do projecto: A. Mele, H. McCann, M. Bratman, G. Ainslie, D. Hausman, J. Knobe.....	15
3.1 Alfred Mele.....	15
3.2 Hugh McCann.....	22
3.3 Michael Bratman.....	27
3.4 George Ainslie.....	32
3.5 Daniel Hausman.....	38
3.6 Joshua Knobe.....	42
4. Um pouco mais de espaço no domínio: J. Broome, M. Smith e J. Elster.....	47
4.1 John Broome.....	47
4.2 Michael Smith.....	52
4.3 Jon Elster.....	54

PARTE I

Filosofia da acção e racionalidade prática.....	65
Susana Cadilha e Sofia Miguens, Filosofia da acção – uma introdução.....	66
António Zilhão, Carlos Mauro, Susana Cadilha, Acção, decisão e explicação da acção – Conversa com António Zilhão.....	91
Ricardo Santos, Susana Cadilha e Sofia Miguens, Da filosofia antiga à filosofia da acção – Conversa com Ricardo Santos.....	138
Vasco Correia e Sofia Miguens, Limites da racionalidade, racionalidade motivada e emoções – Conversa com Vasco Correia.....	145
João Alberto Pinto, A concepção inferencial da acção – Uma comparação entre a interpretação causal e a interpretação normativista, Respostas a Sofia Miguens.....	190

PARTE II

Ética, Juízos e Objectividade.....	193
Sofia Miguens, ‘ <i>What Some Philosophers Wouldn’t Dream of Counting as Part of Their Job</i> ’ – Cora Diamond, literatura e filosofia moral.....	194

Sofia Miguens, Hilary Putnam acerca de ética – respostas a Susana Cadilha.....	207
Susana Cadilha, Em defesa da meta-ética.....	232
APÊNDICE: Referências bibliográficas do Projecto “<i>Conversations on Practical Rationality and Human Action</i>”	263

Introdução

O Projecto “*Conversations on Practical Rationality and Human Action* (2007-2011)

Sofia Miguens

1. Origens, relações e continuidades

O projecto de investigação *Conversations on Practical Rationality and Human Action* foi um projecto do MLAG (Mind, Language and Action Group) financiado pela Fundação Luso-Americana para o Desenvolvimento entre 2007 e 2010¹. Foi concebido como um projecto prático e sobre racionalidade prática, tendo no seu núcleo a filosofia da acção e a filosofia moral, sucessor do Projecto RBD2 (*Rationality, Belief, Desire II – from cognitive science to philosophy – POCI/FIL/55555/2004*), mais direccionado para a teoria da mente. A equipa era reduzida: Sofia Miguens, Carlos Mauro e Susana Cadilha, e os trabalhos envolvidos de uma certa envergadura, nomeadamente a realização de entrevistas a autores em várias partes do mundo, durante a quase totalidade da duração do projecto, e em várias fases. A filosofia da acção estava no seu núcleo; tínhamos por isso que começar por esclarecer para nós próprios o que se entende contemporaneamente por filosofia da acção – o artigo que abre este volume (S. Cadilha e S. Miguens, *Filosofia da acção*) disso dá conta². Uma forma simples de nos localizarmos no domínio da filosofia da acção é perguntar pelos temas centrais desta – e aí encontramos questões relativas à natureza da acção e à explicação de acções, e consequentemente questões relativas a intenção, deliberação, decisão, razões e causas de acções, acrasia, etc – e pelos autores de referência: a ideia razoavelmente consensual de que partimos foi que estes seriam Aristóteles, Elizabeth Anscombe e Donald Davidson³.

¹ De facto, o projecto prolongou-se até 2011, com a preparação das publicações correspondentes e com a organização de cursos resultantes das actividades e materiais recolhidos (nomeadamente o seminário ‘Filosofia da Acção’ no Curso de Mestrado em Filosofia, Ramo de Filosofia Contemporânea, na Faculdade de Letras do Porto, a ter a sua primeira edição em 2011/2012).

² Uma versão (a versão original) do artigo é um capítulo de uma obra coordenada por Pedro Galvão, *A Filosofia por disciplinas* (publicação resultante de um projecto do Centro de Filosofia da Universidade de Lisboa).

³ Centralmente o trabalho de Aristóteles sobre racionalidade prática (cf. Aristóteles [2009], *Ética a Nicómaco*), incluindo as suas ideias sobre acrasia, o trabalho de Anscombe sobre intenção (cf. Anscombe 1957, *Intention*) e o trabalho de Davidson sobre razões e causas (cf. Davidson 1963,

Mas se a filosofia da acção é a sua origem e a sua referência, o projecto teve desde o início um horizonte mais complexo, que é marcado pelo título *Acção e Ética* do presente volume: interessava-nos compreender a forma como concepções de acção, de racionalidade prática e de racionalidade na acção, que segundo algumas perspectivas podem ser formuladas de forma neutra, se reflectiam em concepções *substanciais* daquilo ‘que se deve fazer’. Noutras palavras, interessavam-nos as repercussões, que se estendem desde a ética à filosofia política, à filosofia da economia e ao desenho de políticas públicas, da abordagem das questões relativas a acção, agência e racionalidade. Um tal horizonte significava antes de mais que a própria possibilidade de uma disciplina pura de filosofia da acção, centrada em agentes e na racionalidade prática destes, e ‘isolável’ de questões substanciais e normativas acerca do que se ‘deve fazer’, do ‘bem’ e da ‘justiça’ constituía para nós uma questão em aberto. Alguns autores que participaram no projecto concordam com uma separação – por exemplo António Zilhão, numa das entrevistas incluídas no presente volume, responde a uma das nossas perguntas dizendo: “creio que a filosofia da acção pode ser separada da filosofia moral”⁴. Outros autores discordam, ou pelo menos vêm o tratamento dos tópicos da filosofia da acção como estando inevitavelmente próximo da filosofia moral; é essa a opinião de, por exemplo, Alfred Mele: “Há muitas conexões entre a filosofia da acção e a filosofia moral. Na verdade, alguns dos tópicos sobre os quais os filósofos da acção e os filósofos morais escrevem são os mesmos: por exemplo responsabilidade moral, livre-arbítrio, fraqueza da vontade, natureza das razões para agir, natureza da motivação (incluindo motivação moral) e explicação de acções. A fertilização mútua pode ser muito produtiva”⁵. É de qualquer forma inegável que compreender a natureza de agentes, acções e explicação de acções é essencial para se pensar de forma fundamentada nas questões substanciais e normativas acerca do que devemos fazer – sem compreensão da natureza de agentes, acções e motivação para a acção não poderia nomeadamente haver *ética*, como quer que definamos esta.

Actions, Reasons and Causes). Estabelecendo um fio histórico entre os três, Davidson afirmou acerca do livro *Intention*, de Anscombe, que este era o ‘trabalho mais importante sobre acção desde Aristóteles’.

⁴ Cf. Conversa com A. Zilhão, mais à frente neste volume.

⁵ Cf. Alfred Mele, entrevista ao Projecto *Conversations*. Versão (não-final) on-line em <http://mlag.up.pt>.

É verdade que a separação disciplinar ‘canónica’ da filosofia da acção em relação à ética e à filosofia moral, na filosofia analítica contemporânea, parece marcar a possibilidade de se manter uma distinção clara. A esta distinção vêm de resto juntar-se outras distinções e classificações relativas à própria ética: introduzindo esta como campo disciplinar, é frequente (aliás, quase ‘canónico’) na filosofia analítica, separar a *meta-ética* da *ética normativa* e da *ética aplicada* assumindo assim a possibilidade de separar aquelas a que podemos chamar questões ‘teóricas’ (metafísicas, semânticas) da racionalidade prática, que giram em torno da natureza de propriedades, entidades e juízos morais, e que são atribuídas à meta-ética, das questões ‘práticas’ da racionalidade prática, que estão envolvidas em respostas substanciais específicas à questão ‘O que devo fazer?’. Estas últimas fariam convergir desde as discussões em torno de concepções alternativas dos princípios que regem a acção (como por exemplo a deontologia e o consequencialismo), a argumentos em torno de posições em ética prática perante questões como os direitos dos animais ou a pobreza e a globalização ou outros, por exemplo de entre os variados tópicos específicos da bioética⁶. Todos estes tópicos seriam reportados à ética normativa.

Nos inícios do presente projecto estava entre outras coisas a intuição de que talvez a situação fosse mais complexa do que isso. Tais separações e classificações cumprem sem dúvida uma função didáctica e de iniciação, mas os tópicos eles próprios não se deixam dominar assim tão facilmente. Se perguntamos por exemplo ‘O que é agir moralmente?’⁷ – tomada esta questão como uma formulação da questão ainda mais geral da racionalidade prática e que seria ‘O que devo fazer?’ – não temos certamente que aceitar como resposta algo como ‘*Devo agir moralmente e agir moralmente é agir de forma perfeitamente racional, como qualquer agente racional agiria nas mesmas circunstâncias*’. Esta resposta, em traços gerais kantiana, é sem dúvida muito difundida e muito importante na nossa tradição filosófica, mas a verdade é que ao subscrevê-la não apenas identificamos a moralidade com a própria estrutura da agência racional como o fazemos como se soubéssemos perfeitamente de que falamos quando falamos de racionalidade, e de racionalidade prática em particular, no tipo de agentes que nós somos. Estamos nomeadamente a assumir que no tipo de agentes que

⁶ Tais como o aborto, a eutanásia, a reprodução assistida, a clonagem, etc. Cf., para estas distinções, Galvão (no prelo) e Guisán (2010).

⁷ Esta pergunta é uma citação e uma homenagem – é este o título (pelo menos inicial) da dissertação de doutoramento da co-coordenadora do presente volume, Susana Cadilha.

nós somos pode haver motivação para a acção perfeitamente imparcial, desapaixonada, desinteressada. Kant assumia-o (ou pelo menos a imagem simplificada de Kant usualmente evocada na filosofia moral retrata-o dessa forma). Mas será realmente esse o caso? Algum agente humano agirá alguma vez assim? Que as coisas poderiam ser mais complexas não foi uma intuição que nos tenha caído do céu: foi muito importante para isso o contacto que tivemos no contexto do projecto *Rationality, Belief, Desire II* com críticas várias a concepções ‘puras’ da racionalidade prática e da racionalidade na acção, não apenas a concepção kantiana já referida, mas também a concepção instrumental de racionalidade e a teoria da decisão racional⁸. A crítica ao uso desta última para conceber a interacção social e económica de agentes por autores como Amartya Sen⁹ e Jon Elster¹⁰, por exemplo, com o recuo que essa crítica representa perante concepções ‘mainstream’ de racionalidade nas ciências económicas e sociais, tinha já sido uma referência para nós no passado. As concepções instrumental e de maximização da utilidade esperada (da teoria da decisão racional) são, não resta qualquer dúvida, incontornáveis e centrais como respostas à pergunta acerca da natureza da racionalidade prática. Elas dizem-nos, respectivamente, que devemos (i.e. que é *racional*) agir de forma a mobilizar os meios adequados à prossecução dos nossos fins, e que devemos escolher de forma a maximizar a utilidade esperada. Mas o verdadeiro trabalho começa com as dúvidas que se colocam em seguida: será aquilo de que falamos quando falamos de racionalidade prática ou racionalidade na acção de facto, em última análise, nada mais do que relações entre meios e consequências, i.e. racionalidade instrumental, e

⁸ Para incursões anteriores na temática da racionalidade, cf. Miguens 2002, Miguens 2004 e Miguens & Mauro 2006.

⁹ Economista indiano (n. 1933), Prémio Nobel da Economia em 1998, pelos seus contributos para a teoria da decisão social e do *welfare state*. Ensinou em Delhi, Londres (London School of Economics), Oxford e Harvard, e foi também Reitor (Dean) na Universidade de Cambridge. É um dos fundadores do ONU-WIDER (World Institute for Development Economics Research (cf. <http://www.wider.unu.edu/>), baseado em Helsínquia. Alguns dos seus livros mais importantes são "*On Economic Inequality*" (1973, edição expandida e revista 1997), "*Poverty and Famines: an essay on entitlement and deprivation*" (1981) e "*On Ethics and Economics*" (1987) e "*Development as Freedom*" (1999). Defendeu que o desenvolvimento de um país está essencialmente ligado às oportunidades que este oferece à sua população de fazer escolhas e de exercer a cidadania, o que inclui não apenas a garantia dos direitos sociais básicos, tais como a saúde e a educação, mas também segurança, liberdade, habitação e cultura. Proponente de uma *capabilities approach* como indicador de desenvolvimento e qualidade de vida (cf. Sen 1979, *Equality of What?*), Sen foi um dos criadores, em 1993, com Mahbub ul Haq, do Índice de Desenvolvimento Humano usado desde então pelo Programa das Nações Unidas para o Desenvolvimento.

¹⁰ Cientista social e filósofo norueguês, cf. apresentação mais à frente neste artigo e referências bibliográficas no Apêndice final.

maximização da utilidade esperada tal como esta é caracterizada pela teoria da decisão racional? Para começar a decompor os problemas, podemos desde já notar que a nossa forma de falar de racionalidade e irracionalidade é ela própria múltipla, alguns diriam ambígua: afinal de que é que nós afirmamos a racionalidade e a irracionalidade – das nossas crenças? dos nossos desejos? das nossas acções? dos agentes? das decisões destes? das suas preferências? das suas intenções? A verdade é que na nossa forma comum de falar, falamos de ‘racionalidade’ para caracterizar condições que vão desde a consistência de crenças (pense-se num contexto de epistemologia), até à eficiência de processos sociais e administrativos (pense-se na forma como o filósofo e sociólogo alemão Max Weber¹¹ moldou o uso do termo ‘racionalização’), até à autonomia de indivíduos na agência plenamente moral (pense-se em Kant). Estaremos sempre a falar da mesma coisa? Como poderemos orientar-nos para responder a esta pergunta?

Além de autores vindos das ciências sociais e da economia, como Elster e Sen, também filósofos como por exemplo Jürgen Habermas (tendo Max Weber, precisamente, no horizonte) e Hilary Putnam¹², se interessaram na sua obra pela racionalidade prática e mais especificamente pela forma como concepções da natureza da racionalidade prática subjazem a posições em ética e filosofia política. Putnam, no seu livro *The Collapse of the Fact-Value Dichotomy*¹³, chama a atenção para o que considera serem as perniciosas consequências práticas da aceitação, para tantas pessoas perfeitamente natural, da ideia segundo a qual, quando consideramos a acção, de um lado está a razão, com o seu potencial de universalidade e do outro lado

¹¹ Max Weber é um dos fundadores da moderna sociologia e é também um autor influente na economia, na filosofia, no direito, na ciência política e na teoria da administração. Começou a sua carreira académica na Humboldt Universität, em Berlim e, posteriormente, trabalhou na Albert Ludwigs Universität, de Freiburg, e ainda nas Universidades de Heidelberg, Viena e Munique. Foi uma personagem influente na política alemã da época, consultor dos negociadores alemães no Tratado de Versalhes (1919) e da Comissão encarregada de redigir a Constituição de Weimar. Dedicou-se a estudar aquilo a que chamou o processo de racionalização e desencantamento (*Entzauberung*) característico das sociedades capitalistas modernas. O seu livro *A Ética Protestante e o Espírito do Capitalismo* (*Die protestantische Ethik und das ‘Geist’ des Kapitalismus*, 1920) é célebre. Aí investiga a importância da religião no desenvolvimento tão diverso das culturas do Ocidente e do Oriente, salientando nomeadamente a importância que teriam tido características específicas do protestantismo (como um certo ascetismo), no nascimento do capitalismo, da burocracia e do estado racional e legal nos países ocidentais. Foi ele que em *Política como vocação* (*Politik als Beruf*, 1919), definiu o Estado como uma entidade que reivindica o monopólio do uso legítimo da força física, uma definição muitas vezes repetida nos estudos políticos.

¹² Nos nossos estudos em torno de questões de mente e linguagem no MLAG Putnam já nos interessara muito; neste projecto foram as suas incursões mais recentes na ética a interessar-nos.

¹³ Putnam 2002.

juízos meramente subjectivos¹⁴. Recuemos um pouco para compreender o que está aqui em causa. Existem juízos morais, e juízos morais exprimem aquilo que se deve fazer: por exemplo ‘Esta mulher é adúltera – ela deve ser apedrejada’ ou então ‘Esta mulher é adúltera – não a convidarei para o meu *cocktail party*’. Mas qual é a natureza destes juízos? É para dar resposta a esta questão que é razoavelmente frequente afirmar que juízos morais são juízos de valor e que os juízos de valor são subjectivos. Acrescenta-se então que *valores ou fins não podem ser racionalmente discutidos* (uma ideia que de acordo com muitas pessoas se encontraria por exemplo num filósofo tão central para pensar nas questões da racionalidade prática como David Hume); serão no entanto estas ideias defensáveis? Segundo Putnam, é extremamente importante compreender de um ponto de vista filosófico, e da história da ética filosófica, que determinadas concepções de racionalidade geraram separações e ordenações disciplinares que não são de todo eternas e incontestáveis e que podemos mesmo considerar perniciosas. Putnam tem nomeadamente em mente a concepção de racionalidade que permite aos economistas expurgarem as questões éticas da sua disciplina, ao contrário do que foi o caso nas origens desta, com Adam Smith. Um passo como este pode por exemplo ter como consequência que se considere menor, ou de alguma forma ilegítima, ou teoricamente não fundamentada, qualquer tentativa de *welfare economics*. Na verdade, porque numa área como a economia se trata frequentemente de políticas públicas e de aconselhar governos e organizações não governamentais na sua actuação, a convicção de que de um lado está a razão e do outro ‘juízos subjectivos’¹⁵ tem consequências práticas notórias: como diz Putnam, estão em jogo questões de vida e morte. Uma figura inspiradora central dos escritos de Putnam neste domínio é de resto o atrás referido economista Amartya Sen, colega de Putnam em Harvard durante largos anos e reconhecido crítico de alguns pontos da teoria económica *mainstream* e sobretudo alguém que se dedicou a explorar as relações teóricas da economia com a ética. De qualquer forma, se a racionalidade instrumental e uma concepção abstracta de decisão racional como maximização da utilidade não esgotam a racionalidade prática, uma investigação daquilo que rege essa agência – chamemos-lhe razões, fins, valores – torna-se

¹⁴ Estas questões são retomadas mais à frente neste livro em Miguens, *Hilary Putnam acerca de ética – respostas a Susana Cadilha*.

¹⁵ O exemplo de Putnam em *The Collapse of the Fact-Value Distinction*, é Lionel Robins (economista britânico, da London School of Economics, 1898-1984).

incontornável numa concepção de agentes e agência.

São precisamente estas as preocupações que estão no centro da área Acção, Agência e Racionalidade do MLAG. Sendo todas elas formuláveis como questões para uma teoria da racionalidade prática e como dúvidas acerca da possibilidade uma teoria *pura* da racionalidade prática (quer seja centrada numa concepção instrumental de racionalidade e na teoria da decisão racional, quer seja kantianamente identificada com a agência moral como agência racional, de qualquer agente racional nas mesmas circunstâncias), elas faziam já parte dos projectos sobre racionalidade que decorreram no Instituto de Filosofia do Porto desde 2003. O projecto *Conversations on Practical Rationality and Human Action* veio assim dar corpo ao nosso interesse a longo prazo pela racionalidade prática na sua extensão a disciplinas várias e a áreas várias da filosofia, sendo esse interesse corporizado numa área temática permanente, a área do MLAG a que chamámos **AAR** (Acção, Agência e Racionalidade). Esta área existe desde 2008, e nela, partindo do tópico da racionalidade na acção, pretendemos chegar a tratar questões que se estendem desde a filosofia da acção à ética, à filosofia política e filosofia da economia, incluindo nesta as questões da economia do bem-estar (*welfare economics*), da economia do desenvolvimento (*development economics*) e das políticas públicas¹⁶. O Projecto *Conversations* foi o seu primeiro projecto de maior dimensão.

2. Um projecto prático sobre racionalidade prática – o *script* da entrevista

¹⁶ A área é dirigida por Carlos Mauro, que fez uma tese de doutoramento intitulada *Uma Teoria deflacionista da racionalidade na acção* e tem de momento em curso um trabalho de pós-doutoramento intitulado *Filosofia da economia e políticas públicas – contributos de uma abordagem filosófica da acção, da intenção e da racionalidade* (SFRH/BPD/64367/2009, co-supervisão: Sofia Miguens e Alberto Pinto, Universidade do Porto, e Joshua Knobe, Universidade de Yale, USA). A *acrasia*, ou fraqueza da vontade, um fenómeno central de irracionalidade prática, em que um agente não age de acordo com o seu próprio melhor juízo, que é um dos temas centrais do projecto “Conversations” dada a nossa intenção de compreender a natureza da agência, foi um tópico central do trabalho de doutoramento de Carlos Mauro (na dissertação Mauro pretendeu defender que não existe *acrasia*). Algumas das sub-áreas do domínio Acção Agência Racionalidade têm tido expressão nos Ciclos de Conferências *Direito e Filosofia* e *Filosofia e Políticas Públicas*, iniciados em 2010 na FLUP.

Como já tinha dito, o projecto *Conversations* foi concebido como um projecto prático, o que significou antes de mais que foi concebido em torno de entrevistas e conversas (*'conversations'*) sobre racionalidade prática. Um objectivo maior era organizar um conjunto de publicações em torno dessas entrevistas e conversas, nas quais, tendo no centro a filosofia da acção e estudos da racionalidade prática de proveniência vária, fossem exploradas as ligações destes com, nomeadamente, a filosofia moral, a psicologia moral e a filosofia da economia. O volume principal estaria centrado em entrevistas com figuras relevantes destes campos na cena internacional. As seis questões introdutórias gerais de que partimos foram as seguintes:

- 1) Na sua opinião, quais são os mais centrais, ou mais importantes, problemas da filosofia da acção?
- 2) Para alguns ou todos os problemas seguintes – acção, agência, agente – com o que é que eles contrastam de forma mais significativa?
- 3) Quais são susceptíveis de ser racionais/irracionais?
- 4) Em que sentido é que *a coisa a fazer* deve ser decidida por aquilo que é racional?
- 5) O que explica a acção, e como?
- 6) Como é que a akrasia é possível (se é que o é)?

A questão específica seria simplesmente:

- 7) Na sua opinião, de que forma é que o seu próprio trabalho contribui para o campo de estudos? Quais considera serem as suas contribuições mais importantes? Quais são os seus planos para investigação futura?¹⁷

¹⁷ Uma vez que trabalhámos em inglês, o script era mais exactamente o seguinte: The interview-script: “General questions:1) In your view, what are the most central (or important) problems in the

A ideia era que este *script* gerasse a exploração do pensamento de cada autor, colocando a esse autor as seis questões gerais e a questão específica, prolongando em seguida as entrevistas em função do seu curso natural. A escolha dos autores a entrevistar manifestava por si a nossa intenção de provocar uma justaposição entre a filosofia da acção puramente conceptual e as questões em torno de racionalidade que se erguem a partir da economia e da psicologia cognitiva – i.e. a partir de investigação em áreas empíricas e aplicadas em que está em causa a acção e a racionalidade prática. Este era um diálogo que acreditávamos poder ser frutífero, sem no entanto distorcer o nosso propósito central de ter como ponto de partida e como referência o estado da arte na filosofia da acção. De facto, essa justaposição serviria ainda um outro propósito do projecto, que era o de estreitar as relações entre pessoas que trabalham em Filosofia, em Psicologia e em Economia em torno de questões de racionalidade prática¹⁸.

Começámos por seleccionar autores contemporâneos a quem nos interessaria fazer as perguntas do *script* da entrevista: eles foram Alfred Mele, John Broome, Michael Bratman, Michael Smith, Joshua Knobe, Daniel Hausman, Hugh McCann, George Ainslie e Jon Elster. A lista não coincide com os entrevistados que chegaram ao fim do processo, após os três anos do projecto, mas define o domínio deste¹⁹. De entre esses autores alguns são filósofos, outros provêm de outras disciplinas, outros são filósofos ligados muito de perto a outras disciplinas - uma intenção adicional do projecto era de qualquer forma a de abrir, em Português, caminho numa área a que no

philosophy of action?, 2) For some or all of the problems below – action, agency, agent – what do they contrast with most significantly?, 3) Which of these are liable to be rational/irrational?, 4) In what sense is the thing to do to be decided by what is rational? Are there limits of rationality?, 5) What explains action and how? What is the role of deliberation in rationality?, 6) How is akrasia possible (if you think it is)? / Questions directed at each author’s own work: 7) How do you think your own work has contributed to the field? What do you think are your most important contributions? What are your plans for future research?”

¹⁸ Pensámos ainda numa forma de tornar este livro apelativo (para além do interesse teórico natural dos tópicos tratados, como é óbvio). Decidimos então que existiria um momento do processo editorial em que os entrevistados teriam acesso às respostas uns dos outros e seriam encorajados a utilizar um blogue, para comentários mútuos, objecções, perguntas e respostas. Este foi um blogue privado, ao qual apenas coordenadores e autores tiveram acesso. Foi criado e moderado por Carlos Mauro.

¹⁹ Cf. Apêndice de referências bibliográficas no final do presente volume, com o qual pretendemos caracterizar tal domínio. Versões não finais das entrevistas principais encontram-se disponíveis em <http://mlag.up.pt>.

mundo anglo-saxónico se chama *Philosophy, Politics and Economics*, que consideramos muito importante como um contexto possível para os estudos de Filosofia²⁰, e na qual as questões da acção, agência e racionalidade são fulcrais.

Se nem todos os contactos com os autores chegaram ao fim como entrevistas – no volume em inglês, que consideramos um dos resultados principais do projecto, reunimos entrevistas com Alfred Mele, Michael Bratman, George Ainslie, Daniel Hausman, Hugh McCann e Joshua Knobe –, todos os autores foram no entanto, e entretanto, estudados, no âmbito de um mapeamento que constituía também uma das nossas intenções com o projecto²¹. Uma outra consequência prática da natural progressão do projecto e ligada ao nome de um dos entrevistados, Joshua Knobe, foi a descoberta, ou o contacto mais próximo de alguns membros da equipa (especificamente Carlos Mauro) com um movimento que tem despertado bastante polémica na filosofia analítica contemporânea: a filosofia experimental. Desse contacto seguiu-se a criação, por Carlos Mauro, do X-Phi Lab, Laboratório de Filosofia Experimental, na Universidade do Porto²².

Se num dos extremos das consequências práticas do projecto, a filosofia experimental, se encontra afinal, um ‘regresso’ à ciência cognitiva, que sempre foi um dos horizontes da investigação do MLAG, no extremo oposto temos um ‘regresso à ética’ – a segunda parte do presente volume, *Ética, Juízos e Objectividade* marca tal direcção. Da mesma forma que a ciência cognitiva, a ética não constitui propriamente uma novidade no MLAG: vários trabalhos de doutoramento em torno de questões éticas foram levados a cabo nos últimos anos no Porto em torno do MLAG,

²⁰ O programa original de PPE (Philosophy, Politics and Economics) é o da Universidade de Oxford: cf. http://www.ox.ac.uk/admissions/undergraduate_courses/courses/philosophy_politics_and_economics/

²¹ Uma das nossas intenções com o projecto era ter repercussões em termos do ensino na Universidade do Porto, e foi por essa razão que foi criada em 2010, no Mestrado em Filosofia – Filosofia Contemporânea, uma disciplina de Filosofia da Acção.

²² Cf. *Porto X-Phi Lab – Studies on Agency and Moral Psychology*, em <http://portoxphilab.org/>. O laboratório tem neste momento vários projectos, entre eles dois projectos (financiados pela Fundação Bial e pela Reitoria da UP) sobre ‘*Judgement of moral wrongdoings and emotions*’. A história do movimento da filosofia experimental pode ser, sem problemas, reconstituída usando o blogue Filosofia Experimental em <http://filosofiaexperimental.wordpress.com/2010/08/03/porto-x-phi-lab/>. Vários dos autores com quem trabalhamos comentam nas entrevistas a ideia de filosofia experimental (por exemplo, neste volume, António Zilhão).

nomeadamente os de Susana Cadilha, Miguel Ámen e Rui Vieira da Cunha²³ (cf. From an Ethical Point of View, em Miguens, Pinto & Teles 2010, *Aspects of Judgment: Proceedings of the C-MLAG Annual International Conference 2009 & 2010*) e parte do meu próprio trabalho foi-se dirigindo para questões éticas formuladas a partir de uma postura de alguma forma crítica à maneira ‘canónica’ de encarar a ética na filosofia analítica, a partir de dentro da própria filosofia analítica²⁴.

3. O domínio do projecto: A. Mele, H. McCann, M. Bratman, G. Ainslie, D. Hausman, J. Knobe

Passo então a descrições breves dos autores que nos serviram como ponto de partida no projecto “*Conversations on Practical Rationality and Human Action*”²⁵, utilizando por vezes extractos das entrevistas feitas com aqueles com quem esse processo chegou ao fim. As descrições mapeiam a área AAR (Acção, Agência, Racionalidade) do MLAG, área da qual este foi o primeiro projecto de investigação.

3.1 Alfred Mele

Alfred Mele é, desde 2000, William H. and Lucyle T. Werkmeister Professor of Philosophy na Florida State University²⁶. Trabalha em filosofia da mente e da acção, metafísica e filosofia grega. Considera o comportamento humano o denominador comum dos seus interesses filosóficos variados e vê a filosofia da acção, objecto da entrevista que lhe fizemos²⁷, como uma parte da filosofia da mente com inúmeras conexões com a filosofia moral, considerando nomeadamente o interesse comum, já

²³ Respectivamente em torno de *O que é agir moralmente?*, *Livre arbítrio e The Concept of Person* (todos os trabalhos são orientados por Sofia Miguens, no último caso com co-orientação de Mikko Yrjönsuuri (Universidade de Jyväskylä, Finlândia)).

²⁴ Cf. neste volume Miguens, “What Some Philosophers Wouldn’t Dream of Counting as Part of their Job”. Cf. também o trabalho em curso em torno da objectividade dos juízos morais, no Projecto *The Bounds of Judgment – Task: Judging Morally*, partindo do pensamento de John McDowell.

²⁵ Muita da informação prática e institucional aqui presente foi retirada de sites pelos quais os autores são responsáveis directos ou indirectos (páginas pessoais e páginas de projectos de investigação em universidades, nomeadamente).

²⁶ Alfred Mele já esteve em Portugal tendo nomeadamente sido então entrevistado por Ricardo Santos, para a Revista da Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa, *Cadernos de Filosofia* 4 (1998): 53-84.

²⁷ Disponível em <http://mlag.up.pt>, Projectos ⇒ Conversations on Practical Rationality and Human Action (2007-2011).

atrás referido, da filosofia da acção e da filosofia moral em tópicos como o livre-arbítrio, a responsabilidade moral, a acrasia, as razões para agir, a motivação para a acção e em geral a explicação das acções.

No seu primeiro livro, intitulado *Irrationality – An Essay on Akrasia, Self-Deception, and Self-Control* (1987), Mele abordava problemas vários em torno de duas formas de comportamento irracional: o auto-engano e a fraqueza da vontade ou acrasia. No seu segundo livro *Springs of Action – Understanding Intentional Behavior* (1992) dedicou-se a procurar compreender o que explica as acções, mais especificamente o papel de crenças, desejos e intenções na produção de acções humanas. No terceiro livro, *Autonomous Agents – From Self-Control to Autonomy* (1995), dedicou-se a analisar a possibilidade de existência de acção livre ou autónoma²⁸. No seu quarto livro *Self-Deception Unmasked* (2001), regressou a um dos temas do primeiro, o auto-engano, analisando trabalho empírico que na sua opinião apoia a posição acerca de auto-engano defendida em *Irrationality*. O quinto livro, *Motivation and Agency* (2003), desenvolve uma teoria do lugar da motivação na vida de agentes inteligentes. O sexto livro, *Free Will and Luck* (2006), enfrenta os desafios teóricos maiores que se colocam à tese de que por vezes agimos livremente, em particular os desafios colocados pela sorte e pela manipulação. No seu último livro até à data, *Effective Intentions: The Power of Conscious Will* (2009) examina alegadas provas científicas a favor da tese, recentemente bastante discutida, segundo a qual a vontade livre, ou livre-arbítrio, seria uma ilusão. Esta ideia seria apoiada pelo facto de o nosso cérebro ‘decidir’ o que nós vamos fazer antes de nós sequer termos tomado consciência de tal decisão²⁹; haveria uma minúscula ‘janela temporal para o livre-arbítrio’, de cerca de 100 ms, e tudo o que nesse intervalo de tempo seria possível fazer seria ‘vetar’ decisões ou intenções. Desta forma, as nossas intenções nunca teriam qualquer papel na produção das acções correspondentes (por exemplo a minha intenção de levantar o meu braço não teria qualquer papel no estar agora a levantar-se do meu braço); a nossa ideia de que somos agentes livres, e enquanto tal merecedores de elogio ou culpa pelas nossas acções, seria assim totalmente ilusória. A ideia segundo a qual pensarmos que somos livres é uma ilusão não é obviamente nova na história da

²⁸ Cf. uma sinopse em <http://www.ucl.ac.uk/~uctytho/dfwVariousMele.html>.

²⁹ Filósofos da mente como Daniel Dennett e Patricia Churchland têm interessantes discussões deste tópico – elas travam-se nomeadamente em torno do chamado ‘Problema de Libet’, em referência ao neurocientista Benjamin Libet. Cf. acerca do diferendo entre Dennett e Churchland em torno deste tópico, Miguens 2002: 330-334.

filosofia: pense-se por exemplo em Espinosa ou em Nietzsche e na forma como estes desconstroem a ‘fenomenologia do querer’. A experiência do querer (fazendo-nos experienciar o fazer-acontecer) na sua ligação com aquilo que acontece seria enganosa por nos fazer pensar que somos ‘senhores em nossa própria casa’, sendo tal casa a nossa mente, incluindo a nossa vontade (esse ‘fantasma’ que moveria ‘coisas’³⁰). A ideia segundo a qual pensar que somos livres é uma ilusão é tornada mais premente hoje pela explosão e aprofundamento da investigação em torno da ‘unidade central de controlo’ do nosso corpo, o cérebro, erguendo-se assim inevitavelmente como desafio à filosofia da acção: será que esta nossa experiência do querer como um fazer-acontecer tem alguma relação directa com a realidade causal do mundo? E se a ‘fenomenologia do querer’ e a causação forem completamente independentes e separadas, ou mesmo ‘contraditórias’? O que é que isso significaria para nós? Repare-se, como auxílio à nossa imaginação neste ponto, que o cenário do Génio Maligno não se coloca apenas para a representação de um mundo exterior ao pensamento mas também para a vontade: um ‘neurocientista maligno’ poderia fazer-me acreditar que sou eu que faço as coisas que faço quando de facto nada é feito por mim, tudo é comandado por ele.

Voltando ao trabalho de Alfred Mele, ele é ainda o coordenador de obras de referência no domínio do nosso projecto, tais como *Free Will and Consciousness: How Might They Work?* (com R. Baumeister & K. Vohs, 2010), *Rationality and the Good* (com M. Timmons e J. Greco, 2007), *The Oxford Handbook of Rationality* (com J. P. Rawling, 2004), *The Philosophy of Action* (1997), *Mental Causation* (com J. Heil, 1993) bem como de mais de uma centena de artigos³¹. Mele dirige actualmente um projecto de grande dimensão sobre livre arbítrio, chamado *Big Questions In Free Will Project* (2010-2013)³². O projecto é financiado pela John Templeton Foundation, uma instituição devotada à ‘exploração da natureza e finalidade última da realidade humana’, e que nas suas áreas de patrocínio inclui a área de conexão entre a teologia e a filosofia (essa é mesmo uma das áreas nucleares patrocinadas, já que, claramente, questões relativas à acção e ao livre arbítrio são incontornáveis numa antropologia de inspiração religiosa, que está no horizonte da

³⁰ Cf. Nietzsche, O Crepúsculo dos Ídolos. Um fantasma meramente acompanha eventos, não causa coisa alguma - o ponto de Nietzsche é que não há causas mentais.

³¹ Cf. Apêndice a esta introdução.

³² Cf. <http://www.freewillandscience.com/wp/>

fundação). Mele continua actualmente a escrever sobre livre-arbítrio, fraqueza da vontade, acção intencional, e tópicos próximos.

Na entrevista que lhe fizemos para o Projecto “*Conversations*”, Mele, remetendo para os seus trabalhos acima referidos, analisa cuidadosamente o que entende por ‘acção’. Na sua opinião ‘acção’ é, das três categorias que propomos na entrevista para considerar o ‘ponto de aplicação’ da racionalidade ou irracionalidade (acção, agência, agente), aquela que é a mais básica, aquela a partir da qual um ‘agente’ pode ser definido como uma ‘entidade que age’ e a ‘agência’ como a ‘propriedade de ser um agente’. Considerando a expressão ‘agentes humanos’ a primeira coisa que Mele faz notar é que com ela nos referimos a humanos que agem, e que é muito importante ter em conta que *os humanos nem sempre agem*. Tomando alguns exemplos seus na entrevista, se um neurocientista perverso utiliza o cérebro de Sam como calculadora para fazer multiplicações, não é Sam ele próprio quem está a calcular quando os resultados de operações lhe ocorrem, esses resultados não são os seus pensamentos, pensados por si; Sam é um humano que não está a ser um agente, é alguém em quem aquilo que poderiam ser acções está a ocorrer, mas não enquanto acções de Sam. Se queremos compreender o que é uma acção, Mele pensa que é importante compreender por que devemos dizer que alguém como Sam é um ‘cenário’ de eventos, e não um agente. Na entrevista Mele contrasta em seguida uma situação em que uma outra pessoa, Uma, calcula a conta do restaurante, para a pagar, usando as suas fórmulas habituais de multiplicação, com a situação em que essa mesma pessoa, Uma, faz compulsivamente cálculos de cabeça sem razão nenhuma – com este exemplo Mele quer sublinhar que a compulsão, no segundo caso, altera significativamente o que queremos dizer acerca de Uma como agente.

Mas o que é que há aqui que nos permita captar a forma como acções contrastam com não-acções? A ideia é que algo que parece uma acção – movimentos do corpo próprio, sequências aparentemente intencionadas de eventos – não é realmente uma acção quando há uma falha do agente em participar na suposta “acção” que em si e por si acontece. O que seria essa falha? A ideia de Mele é que para acções serem acções elas têm que ser causadas da maneira certa e esta é uma questão histórica acerca daquilo que de facto as causou: acções têm de ser causadas por crenças, desejos e intenções *do agente* ou não serão acções e os casos acima pretendem precisamente ilustrar tal ponto. Esse ponto pode ser formulado como uma tese geral *causalista*, que é subscrita por Mele: acções intencionais só podem ser

adequadamente explicadas em termos de crenças, desejos e intenções, tais crenças, desejos e intenções são as crenças, desejos e intenções de agentes, e as explicações que os evocam são explicações causais.

Uma vez existindo agentes (entidades que agem em função de crenças, desejos e intenções), coloca-se a questão da responsabilidade desses agentes por (algumas) acções: para Mele muitos agentes não são responsáveis por nada (por exemplo agentes como trutas ou castores) e para um agente ser responsável não basta que a acção seja sua e intencional (Mele dá o exemplo do homem que lava as mãos compulsivamente inúmeras vezes ao dia): apenas agentes humanos são moralmente responsáveis, e ainda assim apenas por algumas das suas acções. O auto-controlo, tão importante na teoria da agência de Mele, é uma condição necessária mas não uma condição suficiente para a existência de acção livre e responsável.

Mele aborda ainda questões relativas a irracionalidade e racionalidade de agentes em relação com as razões para agir, ou razões para a acção. Razões para a acção são razões para os agentes fazerem coisas (por exemplo atiro o vaso da varanda quando o meu inimigo está a passar na rua porque tenho o desejo de matar o meu inimigo e acredito que fazendo aquele vaso cair sobre a sua cabeça o matarei). Na entrevista Mele nota que o interesse pelo tópico das razões para a acção está hoje muitíssimo difundido na filosofia, e que diferentes preocupações podem conduzir-nos a ele, preocupações que focalizam o nosso interesse sobre ângulos diferentes: no contexto da necessidade de explicar acções intencionais, as razões para a acção tendem a ser, como no exemplo que acabei de dar, consideradas como *estados da mente do agente* (combinações crenças-desejos à maneira de Davidson), enquanto que no contexto da avaliação de acções intencionais e seus agentes as razões para a acção usualmente remetem para *estados do mundo externo ao agente* (por exemplo se o preço dos títulos financeiros R está a baixar hoje e vai subir muito amanhã essa seria uma razão para os comprar hoje). Em ambos os casos Mele admite que as razões para a acção podem ser inconscientes. Mas como é que um agente vem a ter razões para agir de determinada maneira? Os caminhos para aí chegar podem ser vários, e um caminho particularmente importante para as questões da racionalidade prática é a *deliberação*. Numa primeira caracterização informal, podemos dizer que deliberar é ponderar o que se deve fazer; usualmente pensamos na deliberação como conduzindo à formação de uma intenção, por contraste com uma posição em que eu prevejo o que vou fazer, ou talvez devêssemos aqui dizer aquilo que ‘me vai acontecer enquanto corpo’ – como

no caso em que eu me atiro de pára-quedas d um avião e o pára-quedas não abre. A *deliberação* decide o problema prático acerca do que fazer e pode ser vista à maneira de um processo de inferência conduzindo a uma conclusão valorativa; a deliberação decide ‘*a coisa a fazer*’, para citar a expressão do *script*-base das nossas entrevistas. Tudo isto é discutível, e era intenção do projecto discuti-lo (por exemplo: devemos dizer que as conclusões de uma deliberação são intenções? devemos dizer que são crenças? ou devemos dizer que a conclusão é a própria acção?) A posição de Mele é a seguinte: a *deliberação* decide o problema prático que o agente coloca a si próprio acerca do que fazer e pode ser vista à maneira de um processo de inferência conduzindo a uma conclusão valorativa (Mele nota entretanto que um agente pode deliberar racional ou irracionalmente; não constitui condição necessária para o agente agir racionalmente que a acção em causa seja um produto directo de uma deliberação). Mas se eu, agente, delibero, e julgo que A é a melhor coisa a fazer, adquirei então (por *default*, diz Mele, como numa rotina em programação) a intenção de A. Isto é assim, é claro, quando não existe *oposição por parte da motivação* do agente – esta oposição será muito importante na forma como Mele lidará com outras questões relativas a agência e agentes. Decisões práticas, i.e. decisões acerca do que fazer, são então para Mele “acções mentais momentâneas de formação de intenções”, e intenções são para Mele “atitudes executivas em relação a planos” (mais especificamente Mele fala de uma acção de assentimento executivo a um plano de acção pertinente em primeira pessoa, ‘*an action of executive assent to a pertinent first-person plan of action*’). Nem todas as intenções, sublinha Mele, são adquiridas ‘por acção’, activamente por mim (*actionally*): se eu abro a porta do meu gabinete todos os dias, quando me aproximo hoje dessa mesma porta posso formar a intenção de a abrir ‘*nonactionally*’, de forma não activamente adquirida – podem, assim, existir intenções que não decorrem de decisões acerca do que fazer. Eu abro a porta, tencionei abrir a porta e abri-a, mas não houve nunca uma decisão explícita de fazer tal coisa.

Até aqui descrevi a forma como Mele vê as coisas correrem ‘por defeito’ no que respeita a deliberação e intenção em agentes. São, no entanto, possíveis disparidades entre processos de formação de intenções como os descritos e a motivação do agente, e é isto que nos conduz a outra questão central do nosso *script*, a acrasia³³. Mele

³³ Este tópico sempre foi particularmente importante para um dos membros da equipa, Carlos Mauro.

expôs-nos a sua posição perante a questão: na sua opinião, a acrasia é auto-controle deficiente do agente – o auto-controle é, por seu lado, “o contrário da acrasia”. A tese é explicada com apoio noutras duas teses, que Mele avança e defende: (1) que “a força motivacional dos desejos não está sempre ‘alinhada’ com a avaliação, pelo agente, dos objectos de desejos”, e que (2) “tipicamente, melhores juízos decisivos são formados em grande parte com base na nossa avaliação dos objectos dos nossos desejos”. Se as teses 1 e 2 são verdadeiras, e Mele pensa que é esse o caso, não deverá ser surpreendente que possam existir casos em que embora pensemos que é melhor fazer A do que B, estamos mais fortemente motivados para fazer B do que para fazer A. Fenómenos de acrasia surgirão aqui. Para Mele eles são portanto possíveis, existem e têm esta explicação. Quanto ao auto-engano, de acordo com a concepção dita “motivacional” do auto-engano de Mele, é de forma *involuntária* e *inconsciente* que os agentes formam crenças irracionais: devido ao impacto das nossas emoções e dos nossos desejos sobre a maneira como tratamos a informação, nós acabamos por adoptar crenças que são contrárias àquilo que a informação de que dispomos parece sugerir.

Considerarei acima a forma como Mele vê a relação do melhor juízo do agente com a motivação deste. Convém no entanto esclarecer que Mele rejeita a identificação das decisões práticas humanas com o melhor juízo do agente: é perfeitamente possível decidir fazer alguma coisa que não se pensa ser a melhor coisa a fazer e é possível julgar que alguma coisa é a melhor coisa a fazer e não decidir fazê-la. De resto e mesmo nos casos típicos de acção acrática, Mele não crê que a crença de que é melhor fazer A seja apagada: a crença persiste, embora o agente pratique uma acção intencional e não constrangida que lhe é contrária.

Considerados mecanismos de crenças e desejos, deliberação, intenção, decisão e motivação do agente, coloca-se ainda uma questão mais, uma questão por assim dizer ‘global’: será o não-constrangimento (do agente a agir) *liberdade*? Esta é uma questão clássica da filosofia, e saber se agentes humanos podem ser livres é algo que também ocupa Mele na sua obra (embora não discuta directamente o tópico na entrevista realizada connosco). De certa forma, é a essa questão que conduz todo o seu minucioso trabalho de análise sobre comportamento humano e racionalidade prática: é apenas levando em conta a panóplia de considerações relativas a desejos, motivação, razões, deliberação, algumas das quais acabei de descrever, que a questão do livre arbítrio deverá ser tratada. Mele afirma mesmo: “*I myself have no special use*

for the noun “will.” [Eu próprio não tenho nenhum uso para a palavra ‘vontade’]. Nisto Mele contrasta com outros autores com quem trabalhamos, tais como George Ainslie ou Michael Bratman, a quem não repugna considerar que ao lidarem com racionalidade prática estão em última análise interessados em propor uma *teoria da vontade*. Mas para Mele, não é falando de ‘vontade’, e considerando que o problema do livre-arbítrio diz respeito a esta, que se fala de liberdade, mas sim considerando todos estes funcionamentos do agente, e a possibilidade de autonomia e auto-controle que eles estabelecem.

Que o primeiro entrevistado do nosso projecto tenha sido Alfred Mele não foi casual – como ele próprio nos diz no fim da sua entrevista, a sua obra tem boas pretensões a oferecer algo como uma ‘filosofia geral da acção’. Partindo de questões muito específicas acerca de acrasia, auto-controlo, papel de crenças, desejos e intenções na produção de acções, peso dos desejos, relação entre decisões e intenções, natureza de motivações para agir, Mele chega a uma concepção do agente livre, e assim a um terreno da discussão muitas vezes reservado a quem trabalha entre a metafísica e a ética: o terreno das clássicas discussões entre compatibilistas e incompatibilistas em torno do livre-arbítrio e da vontade, com o determinismo e o indeterminismo como pano de fundo.

3.2 Hugh McCann

Hugh McCann é Professor no Departamento de Filosofia da Texas A&M University. Interessa-se por filosofia da acção e por áreas da filosofia da mente, da metafísica e da filosofia da linguagem relacionadas com esta, em particular questões relativas a actos mentais, ontologia dos eventos e da mudança [*change*], causalidade e tempo. Interessa-se também por filosofia da religião, em particular por tópicos relacionados com a acção divina, a criação e a eternidade. É autor de *The Works of Agency: On Human Action, Will and Freedom*, bem como de numerosos artigos sobre tópicos muito distintos relacionados com a natureza da acção. Olhemos para alguns dos

artigos de McCann³⁴, de forma a compreender o seu leque de interesses e a sua posição de alguma forma ‘contra-corrente’ relativamente ao naturalismo dominante: ele escreveu por exemplo “Is Raising One’s Arm a Basic Action?”, “Volition and Basic Action”, “Divine Sovereignty and the Freedom of the Will”, “Creation and Conservation”, “The Author of Sin?”; “Intentional Action and Intending: Recent Empirical Studies.” “Resisting Naturalism: The Case of Free Will”, “The Will and the Good”, “Pointless Suffering: How to Make the Problem of Evil Sufficiently Serious”, e também o artigo intitulado “Divine Providence” da *Stanford Encyclopedia of Philosophy*. Em suma: na obra de McCann encontramos discussões em torno da volição, da vontade, do bem e do mal, da soberania divina, da providência divina e da criação – ora, estes são claramente tópicos (e vocabulário) correspondentes a uma abordagem dos agentes e das suas acções bastante distinta da do naturalismo dominante.

O interesse de McCann pela filosofia da acção por um lado e pela filosofia da religião por outro é por si próprio revelador mas não é de todo estranho: na verdade ao longo da história da filosofia muitos dos tópicos hoje tratados pelos filósofos da acção foram abordados num contexto teológico ou filosófico-teológico. Basta considerar que compreender o que explica acções é fundamental para, por exemplo, abordar questões de pecado e responsabilidade, ou que considerar a temporalidade na sua relação com o conhecimento disponível é fundamental para pensar na decisão divina e humana ou para discutir a relação entre liberdade (humana) e presciência e onisciência divinas. Discussões sobre agência estiveram assim sempre presentes no próprio seio da tradição filosófica em torno de questões filosófico-teológicas maiores – desde a liberdade humana, à soberania divina, ao problema do mal³⁵. Por isso uma forma de abordar a obra de McCann, de acordo com a sua própria sugestão na entrevista que nos deu, será procurar compreender a forma como tópicos como volições e acções básicas, formação e execução de intenções como actividades de raciocínio prático se relacionam com questões acerca de vontade e liberdade humanas e soberania divina, ou por exemplo perguntarmo-nos como pode uma noção de vontade livre, a desenvolver no âmbito da filosofia da acção, ter um papel importante a desempenhar na tentativa de resolver o problema filosófico-teológico do mal.

³⁴ As referências completas podem ser encontradas no final do presente volume.

³⁵ O problema é gerado pelo facto de a onipotência, onisciência e omnibenevolência divinas não parecerem consistentes com a existência de mal no mundo.

O conceito emblemático de McCann como filósofo da acção, que lhe permite formular a sua oposição ao causalismo formulado em termos de crenças e desejos e lhe permite conceber a forma como a acção difere de outros eventos, e que virá ainda a ser relacionado com vontade livre e responsabilidade, é o conceito de *volição*. Volição é de acordo com McCann a “actividade interna pela qual a agência é exercida quando levamos a cabo acções manifestas”, e não é “um acto momentâneo” mas sim “uma actividade de querer” (querer exercer aquilo por que a acção se cumpre)³⁶. Para percebermos a necessidade deste conceito para McCann, consideremos um exemplo seu na entrevista que nos deu: um paciente, numa enfermaria, sofrendo de uma determinada condição neurológica, quer levantar o braço, está a querer levantar o braço, mas não tem qualquer sucesso, o seu braço não está a levantar-se. McCann pensa que há que dar conta desta possibilidade na nossa teoria da acção: é perfeitamente correcto afirmar que uma acção levada a cabo por uma razão é intencional, não há nada de errado em tratar as crenças e desejos do agente como razões para a acção, mas isso não deve fechar os nossos olhos ao facto de *agir ser sobretudo querer agir*, querer que algo (por mim) aconteça: a acção não está nos acontecimentos mas no facto de os produzir. Crenças e desejos podem ser elementos necessários para que a acção tenha lugar, no entanto apenas o acto volitivo é condição suficiente³⁷. Para McCann o aspecto fenomenológico aqui envolvido é muito importante: estar envolvido em volição é sentir-se *espontaneamente activo*, não é alguma coisa que nos aconteça, ou à qual demos por nós submetidos. Mas antes de chegar à volição há que compreender as relações entre volição e intenção e entre crenças, desejos e intenções. Antes de mais, a ideia de McCann é que intenções não podem ser simplesmente combinações de crenças e desejos; num outro exemplo que McCann usa na entrevista, eu posso desejar ver Roma, e acredito que posso fazê-lo se for de férias para Itália este Verão. Mas daí dificilmente se segue que eu tenha a intenção de ir a Itália. O que falta? A intenção requer que eu tenha decidido, que eu esteja comprometido com fazer F e com formar um plano eficaz para tal, comprometido ainda com levar esse plano a cabo em tempo apropriado. Precisa-se da intenção, e no entanto segundo McCann, se nos referimos apenas a intenções, referir-

³⁶ McCann sublinha na entrevista que muitas pessoas, desde que Gilbert Ryle tornou a palavra *volition* quase tabu, falam de “occurrent intention” ou “intention in action”, mas na sua opinião referem-se aquilo a que ele chama volição.

³⁷ Parafrazeio aqui Cadilha e Miguens, *Filosofia da Acção*, mais à frente neste volume.

nos-emos a um estado, não a eventos, enquanto que as acções são eventos. Por outro lado um movimento tal como o movimento de um dos nossos membros é um evento mas não é por si só uma acção. Este é então o espaço preenchido por aquilo a que McCann chama *volição*, o meio pelo qual executamos intenções anteriores. Uma intenção, na medida em que é um estado, afirma McCann, não se pode executar a si própria. Outra diferença importante aqui (e também uma discussão clássica na filosofia) para McCann é a diferença entre desejos e *volição*. Um desejo é sempre algo de passivo em nós, acontece-nos, damos por nós tendo-o, enquanto uma *volição* não – nos seus termos, ela é *'actional'*, i.e. é actividade. A aposta de McCann na *volição* ganha em ser vista à luz histórica que ele próprio nos propõe na entrevista: McCann traça a história da consideração filosófica da acção desde o behaviorismo lógico, passando pela teoria causal da acção, hoje bastante difundida e largamente aceite, de forma a vir a expôr a sua própria posição, centrada no conceito de *volição*, como uma crítica a esta última.

As implicações morais e teológicas da discussão sobre acção e agência são, como disse atrás, algo que interessa McCann; no entanto ele tem o cuidado de sublinhar na entrevista o facto de a descrição acima deixar ainda em aberto a questão de saber se, ao agirmos, a vontade é causada a fazer o que faz, quer por razões quer por intenções. Por um lado temos um domínio de abordagem de questões da acção que deve vir a permitir-nos conceber a forma como exercícios da vontade poderiam ser, digamos, *'pecaminosos'* em si mesmos. Noutras palavras, e de forma um pouco crua, deve ser possível *'culpar'* o querer, deve ser possível esse querer ser um querer mau de alguém, deve haver algo como maus actos de formação de intenções, como quando alguém decide matar uma pessoa – isso será nos termos propostos por McCann uma questão da actividade volicional através da qual se executa tais (más) intenções. Mas a questão acerca de se a vontade está sujeita à causação não se confunde com esta questão. Seremos nós livres? Se a vontade estiver sujeita à causação não poderemos possivelmente ser livres? Na entrevista, McCann considera o problema do livre arbítrio que aqui se ergue como parte dos problemas da filosofia da acção, e que introduz discussões clássicas sobre compatibilismo e incompatibilismo, determinismo e indeterminismo³⁸.

³⁸ Cf. Miguens 2009, Liberdade, cognição, acção.

No sentido libertário (incompatibilista), a vontade livre consiste em formar ou executar uma intenção sendo que o nosso decidir, o nosso querer, não é produto de causação determinista. O libertário tem pois de defender que razões podem ter força explicativa sem no entanto determinarem causalmente a formação da intenção e tem ainda de explicar por que é que uma decisão não é um mero acidente pelo qual o agente não é responsável. Os deterministas compatibilistas por seu lado devem encontrar uma maneira de distinguir acções livres e não livres num cenário determinista e oferecer uma teoria da responsabilidade e da culpa que seja compatível com as nossas práticas e também com o determinismo. Nenhuma destas alternativas gera tarefas simples. O impasse, entre libertários incompatibilistas e deterministas conduz então à alternativa de McCann, uma perspectiva prática sobre a acção que convoca a fenomenologia – a nossa experiência de agirmos. Eis por que McCann considera vácuo a noção de causação do agente: “Da perspectiva prática do agente (...) é conceptualmente impossível que a decisão ou o querer contem como ‘irrupções, porque é conceptualmente impossível que estes fenómenos ocorram de forma accidental ou inadvertida. Para vermos isto, simplesmente imaginemos o exemplo que mencionei atrás: como é que reagiria se um estudante viesse ter consigo e dissesse “tenho muita pena de ter faltado ao seu curso ontem, Professor X. Decidi acidentalmente não vir.”³⁹

Finalmente, relativamente à acrasia, esta é, de acordo com McCann, perfeitamente possível e consiste em casos em que não temos motivação para executar a intenção que formámos. McCann não recua perante a ideia de que o conceito de acrasia tem conotações morais, e na sua opinião não há nada de errado com isso: a acrasia é frequentemente uma falha na rectidão ou resolução (*fortitude*) da pessoa e enquanto tal algo acerca do carácter desta.

O trabalho de McCann sobre racionalidade prática mostra claramente o quanto as questões em torno de acção que nos propusemos trabalhar têm implicações directas em assuntos morais e filosófico-teológicos, estando enquanto tal plenamente instaladas em terrenos clássicos da história da filosofia. Essa foi parte da sua importância no nosso projecto.

³⁹ Cf. Hugh McCann, entrevista, versão on-line em <http://mlag.up.pt>.

3.3 Michael Bratman

Michael Bratman é Durfee Professor no Departamento de Filosofia na School of Humanities and Sciences da Universidade de Stanford – um lugar central na história da ciência cognitiva da segunda metade do século xx – e está também ligado ao Symbolic Systems Program (Programa de Sistemas Simbólicos) da mesma universidade. Este é um programa de ciência cognitiva, que inclui várias áreas disciplinares, desde a Inteligência Artificial à filosofia e que visa o estudo de computadores e de mentes, i.e., em geral, de agentes naturais e artificiais que usem símbolos para representar informação. Os seus interesses têm como centro a filosofia da acção e a filosofia moral, indo desde questões gerais relativas à natureza da agência, da intenção e da razão prática, do livre arbítrio e da responsabilidade moral, até estudos muito específicos de fenómenos tão importantes para compreender a socialidade e a coordenação dos humanos (e possivelmente de outros agentes) como a intenção partilhada e a agência partilhada. O seu trabalho conduziu-o à proposta do modelo BDI (Belief-Desire-Intention) do raciocínio prático humano, um modelo hoje usado em muitas áreas, incluindo a Inteligência Artificial. Em 2008 recebeu o IFAAMAS (International Foundation for Autonomous Agents and Multi-Agents Systems) Award for Influential Papers, um prémio que assinala contribuições fulcrais para a investigação acerca de agentes e sistemas multi-agentes⁴⁰.

Mas o que é que um filósofo como Bratman, interessado em compreender a natureza do raciocínio prático humano pode ter a ver com Inteligência Artificial, de forma a ter mesmo visto um modelo em Inteligência Artificial (o modelo BDI, um modelo de software desenvolvido para programar agentes inteligentes⁴¹) ser inspirado nas suas

⁴⁰ Cf. o site da Fundação, em <http://www.ifaamas.org/>

⁴¹ Um agente BDI é um agente de software com racionalidade limitada, dotado de atitudes mentais, que são as ‘crenças’, ‘desejos’ e ‘intenções’ referidas no nome do modelo. A arquitectura de um sistema BDI pode ser caracterizada pelos seus componentes crenças, desejos, fins, intenções, planos e eventos. As **crenças** representam o estado informacional do agente, i.e. as suas crenças acerca do mundo, inclusive sobre si próprio e sobre outros agentes. Podem estar incluídas aqui regras de inferência que permitem chegar a novas crenças. Dizer ‘crença’ em vez de conhecimento é uma marca do reconhecimento de que as crenças dos agentes não têm que ser verdadeiras, e podem mudar. As crenças estão armazenadas numa base de dados (‘base de crenças’ ou ‘conjunto de crenças’). Os **desejos** representam os estados motivacionais dos agentes. Representam objectivos ou situações que o agente gostaria de alcançar ou fazer acontecer, por exemplo: ficar rico, ou encontrar o melhor preço. **Fins (goals)**: um fim é um desejo que está a ser adoptado para procura activa pelo agente. Uma restrição importante aqui é que o conjunto dos desejos activos deve ser consistente. O agente pode ter dois fins

investigações?

O primeiro livro de Bratman *Intention, Plans, and Practical Reason* (1987; Harvard University Press, 1987; reimpresso em 1999, CSLI Publications) com o seu foco em intenções e planos para compreender a natureza da racionalidade prática chegaria só por si para dar parte da resposta: o livro teve um enorme impacto ao chamar a atenção dos teóricos da acção para o papel constitutivo e complexo das *intenções* e do *planeamento* na agência racional. Para Bratman a intenção de agir não é apenas o resultado do raciocínio prático – é também o input deste, sendo fundamental para compreender a agência e crucial para conceber a natureza da diferença entre racionalidade prática, relativa ao que se deve fazer, e racionalidade teórica, relativa ao que se deve acreditar. Esta é de alguma forma a ‘marca de Bratman’ nos estudos da racionalidade prática: o sublinhar da importância específica de intenções e planos na racionalidade prática e o sublinhar da forma como isso distingue a racionalidade prática da racionalidade teórica. O referido modelo BDI de Bratman é uma proposta acerca da natureza da racionalidade prática e da agência que precisamente tem no centro a importância da intenção de agir, separando a actividade de selecção de um plano da execução de planos actuais e considerando que agentes são capazes de equilibrar o tempo passado a deliberar sobre planos com a execução de planos. Bratman continuou o trabalho iniciado em *Intention, Plans, and Practical Reason* com *Faces of Intention: Selected Essays on Intention and Agency* (Cambridge University Press 1999), uma colecção de ensaios. Em *Structures of Agency: Essays* (Oxford University Press, 2007) analisa a forma como a agência humana se relaciona com questões de auto-determinação, auto-governo, e autonomia. Além da importação do modelo BDI na Inteligência Artificial, o seu trabalho sobre agência partilhada (*shared agency*), intenção partilhada (*shared intention*) e valorização partilhada (*shared valuing*) tem sido influente na filosofia do direito e na teoria política, na

como desejáveis – ir à festa e ficar em casa – mas não pode tê-los como fins concorrentes simultaneamente. **Intenções** representam o estado deliberativo do agente, aquilo que ele escolheu fazer; de alguma forma intenções são desejos com os quais o agente se comprometeu, começando a executar um plano. **Planos** são sequências de acções que um agente executa para ver realizadas intenções; podem incluir sub-planos (o plano de ir à festa pode incluir o plano de encontrar o vestido certo). No modelo de Bratman, os planos são inicialmente apenas parcialmente ‘preenchidos’, o preenchimento de detalhes acontecendo enquanto progridem. **Eventos** são despoletadores de actividade reactiva do agente, no sentido em que podem actualizar crenças, iniciar planos ou modificar fins; podem ser gerados externamente ou internamente. O modelo BDI implementa assim os principais aspectos da teoria do raciocínio prático humano de Bratman e as componentes desta (crença, desejo e intenção).

medida em que tem um claro impacto na nossa compreensão das capacidades humanas básicas de socialidade e auto-governo. Estas são de resto capacidades centrais para o que faz dos humanos *humanos* (vale a pena dizer que para Bratman o fenómeno nuclear é o da *intenção partilhada*), e temas cuja compreensão recruta tópicos tratados noutras áreas da filosofia tais como o conhecimento mútuo e a natureza da convenção⁴².

Bratman vê a sua *planning theory* do raciocínio prático como uma ‘teoria modesta da vontade’ – na verdade, vê-a como uma teoria alternativa por exemplo à teoria de George Ainslie, de que falarei em seguida⁴³. Para Bratman e ao contrário do que um incompatibilista suporia, a vontade não é algo de exterior à ordem causal: criaturas acerca das quais dizemos terem uma ‘vontade’ são criaturas dotadas de determinadas estruturas psicológicas e capacidades que permitem impor uma certa estrutura no pensamento e na acção, no nosso “*downstream practical thinking and acting*” (é esta a expressão que utiliza). A questão que se coloca é a de compreender essa estrutura, e Bratman pensa que o conceito e o fenómeno do planeamento, bem como as normas associadas, são fundamentais, dado nomeadamente o seu papel nas nossas vidas. Desde logo, as nossas vidas estendem-se no tempo e são socialmente coordenadas e essas são condições com as quais temos de lidar tendo as limitações cognitivas e o acesso limitado à informação que temos (convém recordar, para o colocar num quadro mais lato de história da ciência cognitiva, que Bratman foi influenciado pela teoria da racionalidade limitada de Herbert Simon); a importância dos planos reside precisamente aí. A escolha e a decisão são as formas standard de ‘dotarmos’ as nossas mentes de estados de planeamento ou planos. Nós, humanos, somos *planning agents* (agentes planeadores) e Bratman insiste que *planning agents* como nós são um tipo muito especial dos *purposive agents* (i.e. dos agentes com propósito, categoria que abrange muitos outros animais, por exemplo ratos ou morcegos ou vacas); a *purposiveness* que caracteriza todas estas criaturas animais, como nos caracteriza a nós, não exaure o fenómeno de *planning* que o interessa particularmente. Olhemos então para o específico tipo de conexão com o mundo que a agência envolve e a importância do planeamento nesse contexto segundo Bratman.

⁴² Em projectos anteriores trabalhámos sobre estes tópicos, cf. por exemplo Miguens 2006, Conceito de crença, triangulações e atenção conjunta, em Miguens e Mauro 2006.

⁴³ Bratman, Planning and Temptation, in *Faces of Intention*.

Na *planning theory* de Bratman as intenções têm um relevo especial – ora, é fácil verificar que nas discussões filosóficas clássicas em torno da motivação dos agentes humanos para a acção é notoriamente mais frequente falar de desejos do que de intenções. No seu trabalho conosco Bratman disse-nos que via a sua obra, o enfoque desta no planeamento, como resultando de ‘uma discussão que não chegou a acontecer entre Elizabeth Anscombe e Donald Davidson sobre agência e intenção’. Antes de mais, Bratman recorda que existem muitos pontos de acordo entre Anscombe e Davidson: eles concordam acerca de individuação de acções, ambos admitem ‘descrições múltiplas’ daquilo que acontece; os pontos de discórdia clássicos entre ambos dizem respeito a saber se o apelo às razões do agente para agir constitui ou não constitui explicação causal, e à existência e natureza de um tipo especial de conhecimento não-observacional que um agente como nós teria do seu agir enquanto age (algo em que Anscombe insiste). Sobretudo devido ao trabalho de Davidson, o debate sobre razões e causas tornou-se central na filosofia da acção, enquanto a questão do auto-conhecimento ficou um pouco mais esquecida (embora, como Bratman recorda, tenha vindo a ser recentemente recuperada por exemplo por um autor como David Velleman). No entanto há, segundo Bratman, ‘um pacto tácito’ entre Anscombe e Davidson acerca de *intenção de agir, ou tencionar agir (intending to act)*: nenhum deles pensava que por aí se chegasse ao que quer que fosse de fundamental para compreender a natureza da agência. É no entanto precisamente aí que Bratman pensa que devemos focar a nossa atenção.

Uma razão para isso é a seguinte: se quer as intenções quer os desejos são ‘pró-actitudes’ (actitudes mentais respeitantes à acção), as intenções distinguem-se por serem *pró-actitudes de controle do comportamento envolvendo compromisso*. Ora, é este compromisso que possibilita a persistência temporal de planos, além de proporcionar novos planos feitos com base naqueles com os quais o agente está já comprometido. Ao contrário dos desejos, as intenções estão sujeitas a *requisitos de racionalidade*, requisitos estes que Bratman considera estarem em operação na racionalidade prática: em conjunto com as crenças, as intenções devem proporcionar qualquer coisa como um modelo consistente do futuro do agente. Como Bratman afirma em “*Intention, belief, practical, theoretical*”⁴⁴, “A ideia de que intenções – em contraste com os

⁴⁴ Bratman 2009, em Robertson, Simon, 2009, *Spheres of Reason*, Oxford, OUP.

comuns desejos – estão embebidas em regularidades características e estão sujeitas a pressões racionais específicas de consistência e coerência é central na teoria do planeamento. Há, especialmente, um requisito racional (*rational demand*) de que as nossas intenções, tomadas em conjunto com as nossas crenças, produzam um modelo consistente do nosso futuro. Há ainda um requisito racional adicional de que as nossas intenções sejam coerentes em termos de meios e fins, no sentido, grosseiramente falando, que não seja verdade que tencionemos E, acreditemos que E requer que tencionemos o meio M, e ainda assim não tencionemos M. Estas normas de consistência e coerência estão em operação no raciocínio prático de um agente que planifica (*planning agent*), e conceber a natureza dos referidos requisitos racionais é uma das preocupações centrais de Bratman. Intenções assim caracterizadas são elementos de planos dirigidos-ao-futuro (*future-directed plans*) fulcrais na racionalidade prática. Uma outra característica relevante desses planos e que muito interessou Bratman é o facto de estes tipicamente serem parciais, sendo preenchidos ‘pela realidade e pelo contexto’, de uma forma que contrasta com as características da racionalidade teórica. Exemplos são os fenómenos de ‘aceitação num contexto’ em contraste com as crenças (este é o tópico de um artigo que Bratman publicou na Revista *Mind* em 1992⁴⁵) ou o ajuste do background cognitivo na deliberação (fenómenos como *positing* e *bracketing* – postular e pôr entre parêntesis, aspectos com os quais noutros contextos – por um exemplo o contexto de uma aposta – o agente não lida da mesma forma). Bratman tem-se dedicado ainda a comparar a flexibilidade de intenções e decisões e a estudar a estabilidade da intenção.

Bratman, que insiste que a tradição filosófica ocidental se concentrou sobretudo na compreensão da natureza da mente e do conhecimento mais do que na compreensão da natureza da agência e dos agentes, representa no nosso projecto a importância que a compreensão das estruturas da agência tem para não apenas o campo filosófico clássico em que esta sempre foi estudada e que tem no seu centro a nossa natureza moral, mas para um grande número de campos exteriores à filosofia, desde a Inteligência Artificial às ciências sociais e políticas. A agência, no sentido de “comportamento potencialmente sob o controle de um mecanismo de condução, que persegue um dado fim” é um conceito comum a um grande número de disciplinas e a

⁴⁵ Bratman 1992, Practical Reasoning and Acceptance in a Context, *Mind* 101, 1-14 (reimpresso em Bratman 1999, *Faces of Intention*).

investigação sobre agência pode cobrir uma enorme diversidade de agentes. Daí que a ideia de Bratman segundo a qual o *objecto da filosofia da acção* são as *estruturas básicas da agência especificamente humana* abra para esta uma enorme potencialidade de aplicações.

3.4 George Ainslie

George Ainslie é um psiquiatra, psicólogo e economista experimental americano. É autor de dois livros particularmente importantes para os temas do projecto “*Conversations*”: *Picoeconomics – The Strategic Interaction of Successive Motivational States within the Person* (1992) e *Breakdowns of the Will* (2001). A ideia de uma ‘picoeconomia’⁴⁶ é a seguinte: assim como na economia clássica se descreve a negociação por recursos limitados entre instituições e na microeconomia se descreve a negociação por recursos limitados entre indivíduos, na picoeconomia procura-se descrever *as interacções dentro de um mesmo indivíduo pelo controlo da capacidade comportamental finita desse indivíduo*, sendo que essas interacções que se assemelham também elas a uma negociação entre partes. Em *Breakdowns of the Will* Ainslie aplica estas concepções ‘picoeconómicas’ a um modelo do eu (ou da pessoa) e da vontade, que lhe permite conceber os colapsos desta. Ainslie tem sempre em mente não apenas o objecto teórico clássico de interesse dos filósofos – a acrasia – mas também condições clínicas várias, centralmente a adição (a drogas ou álcool, por exemplo, mas também muitos outros tipos de adição). Que esta condição de adição e a ambiguidade que a acompanha – o querer e não querer fazer o que se está a fazer – é deveras comum nas sociedades afluentes é um ‘diagnóstico civilizacional’ avançado por Ainslie em *Picoeconomics*. O que é importante e modelar neste tipo de fenómenos para a compreensão do eu e da vontade é a possibilidade de agentes serem seduzidos por prazeres urgentes com objectos destrutivos e que afectam aquelas que seriam as suas próprias escolhas estáveis de longo prazo, como se houvesse um eu que está sempre alerta para a possibilidade de um gozo compulsivo imediato – *quero esta dose de droga agora* -, um gozo para o qual está fortemente motivado, e esse eu vencesse e controlasse a acção, mesmo que isso envolva falhar em dar satisfação a

⁴⁶ ‘Pico’ significa ‘micro-micro’.

longo prazo ou mesmo em corresponder ao que o eu (outro eu?) ‘quer ser’. A ambivalência que acompanha estes fenómenos torna-os particularmente importantes no estudo da acção e dos agentes; para Ainslie eles marcam antes de mais o facto de que a compreensão das preferências de agentes como nós envolve mais do que a computação de valores relativos que a teoria da escolha racional considera. Ainslie relaciona as suas reflexões sobre a acção, o eu e a vontade com o estudo do comportamento de adição, com que lidava no seu trabalho clínico no Veteran Administration Medical Center, Coatesville, Pennsylvania, onde foi Psiquiatra Chefe⁴⁷. O viciado que quer e não quer a heroína é um bom exemplo da ambiguidade referida, e uma ideia básica de Ainslie é que diferentes partes ou aspectos da personalidade estão em conflito nos indivíduos (algo que relaciona com a teoria Freudiana do Id, do Ego e do Superego) e isso é particularmente claro nos comportamentos de adição. A teoria tornou-se importante na economia comportamental (*behavioral economics*) nomeadamente sob a forma da teoria dos sis múltiplos (*multiple selves*) de Richard Thaler (University of Chicago – School of Business). Conjuntamente com as aportações de Drazen Prelec (MIT Sloan School of Management) o trabalho de Ainslie constituiu uma via pela qual os estudos sobre condicionamento operante vieram juntar-se a outras abordagens da natureza da tomada de decisão (*decision making*), criando um desafio muito poderoso à teoria da decisão racional dominante no pensamento económico.

Penso que se tornou já claro que o background académico de Ainslie é bastante distinto do background de outros autores com que trabalhamos – o facto de ele provir das ciências médicas tem reflexos óbvios. Na entrevista que lhe fizemos, e de forma reveladora da sua origem académica, Ainslie começa por avançar uma definição de acção bastante diferente (bem mais ampla) de definições de outros autores: para Ainslie, acções são *processos teleológicos repetíveis que respondem a recompensas* – entre os seus exemplos estão comportamentos de bulimia e de adição. É a ‘*reward-responsiveness*’, o responder a recompensa, das acções definidas desta forma que constitui objecto especial dos interesses de Ainslie, e é nesse contexto que ele responde às questões da entrevista acerca de escolha, deliberação, acrasia, etc. Na entrevista e ainda acerca do que se entende por acção, Ainslie afirma: “Acção é

⁴⁷ Ainslie foi ainda Clinical Professor na Temple University, em Philadelphia.

qualquer processo que seja originado no interior do indivíduo e cuja operação elimine uma alternativa possível. Quando dás um pontapé, és incapaz de não estar a dar um pontapé. Quando sucumbes ao medo és incapaz de estar calmo. Quando pensas num urso branco és incapaz de não pensar num urso branco. Eu não reconheço uma distinção entre acções e paixões, ou entre acções e pensamentos. Reconheço apenas graus de motivação [*degrees of motivatedness*] e graus de voluntaridade [*degrees of deliberateness*]

Mas se concebemos a natureza da acção da forma indicada, de que falamos quando falamos de racionalidade e irracionalidade na acção? Como diz Ainslie na entrevista “É muito difícil dividir de forma estrita os produtos da negociação intertemporal [*num indivíduo*] em racionais e irracionais. O teste de racionalidade que está mais próximo do da linguagem comum é se uma escolha serve o nosso interesse de longo prazo”. Melhor será por isso focar a análise na motivação, e assim teremos que “A motivação implica uma moeda universal a que é melhor chamar recompensa, que é necessária para decidir a competição entre todos os membros das classes de processos que podem ser substituídos uns pelos outros. Refiro-me ao lugar hipotético no qual estes processos competem por expressão como o *mercado motivacional*”⁴⁸.

Ao falar desta forma de ‘recompensa’ Ainslie mostra claramente que além de provir de um âmbito científico distinto dos nossos outros entrevistados, a forma como se move no seu campo é ela própria peculiar: de forma pouco comum para um psiquiatra, Ainslie interessou-se pela investigação experimental do condicionamento operante em animais – trabalhou sobre esse assunto com a supervisão do psicólogo Howard Rachlin. Investigou nomeadamente a escolha inter-temporal em pombos, e foi o primeiro a demonstrar experimentalmente o fenómeno da inversão de preferências a favor de benefícios mais imediatos à medida que o ponto de escolha entre duas opções, uma com benefícios produzidos antes da outra, é ‘adiado’ no tempo. Explicou este fenómeno em termos de *desconto hiperbólico de recompensas futuras*, uma ideia decorrente de conceitos que Rachlin e outros tinham desenvolvido a partir da *matching law* de Richard Herrnstein⁴⁹. Ainslie integrou então tais ideias em trabalho

⁴⁸ G. Ainslie, Entrevista, *Conversations on Practical Rationality*, em <http://mlag.up.pt>

⁴⁹ Richard J. Herrnstein (1930 –1994) foi um psicólogo americano, de linha skinneriana, investigador da aprendizagem animal e um dos fundadores da análise quantitativa do comportamento. A *matching law* é a sua descoberta principal como psicólogo experimental e dá nome à tendência dos animais para

experimental e teórico sobre escolha inter-temporal, por exemplo os estudos de Walter Mischel sobre o adiamento da gratificação em crianças. Tópicos como estes constituem o material de *Picoeconomics*, onde se encontram ainda convocadas inúmeras aportações da tradição filosófica de reflexão sobre as paixões, desde os estóicos a David Hume, até, nomeadamente (e smuito centralmente) Sigmund Freud. Um ponto central é de qualquer forma que a maneira (por exemplo utilitarista) de pensar em escolhas em termos de maximização da utilidade e tendo como pedra de toque as escolhas monetárias simplesmente não captura a maioria das escolhas humanas complexas nem a *ambiguidade* destas, i.e. o querer e não querer que é tão frequente nos comportamentos humanos de escolha, e que é óbvio por exemplo no comportamento de adição.

Voltando à picoeconomia, convém descrever de forma mais específica a principal ideia a que o termo '*picoeconomics*' é associado: ele é usado por Ainslie para descrever as implicações de uma descoberta experimental específica: a tendência que as pessoas têm para ter uma forte preferência por benefícios (*payoffs*) imediatos em relação a benefícios tardios, sendo que a tendência aumenta quanto mais próximo do presente ambos os resultados estiverem. Dadas duas recompensas semelhantes, os humanos mostram uma preferência por aquela que chega mais cedo. Diz-se que os humanos *descontam o valor da recompensa mais tardia*, por um factor que aumenta com o aumento da demora.

Um grande número de experiências confirmaram as preferências espontâneas de sujeitos humanos e não-humanos como seguindo uma curva hiperbólica em vez da curva exponencial convencional que produziria escolha consistente ao longo do tempo. Por exemplo muitas pessoas, quando lhes é oferecida a escolha entre 50 dólares agora e 100 dólares daqui a um ano, imediatamente escolhem os 50 dólares. No entanto, dada a escolha entre 50 dólares daqui a 5 anos, e 100 dólares daqui a 6 anos todas escolhem 100 dólares daqui a 6 anos, mesmo se esta é a mesma escolha, vista a uma distância 5 anos maior.

Quando em *The Breakdowns of Will* (Colapsos da Vontade) Ainslie lida mais

alocarem as suas escolhas em proporção directa das recompensas que destas onbtêm. Imagine-se por exemplo duas fontes de recompensa, uma duas vezes mais rica do que a outra. Aquilo que Herrnstein descobriu foi que animais escolhem com duas vezes mais frequência aquela alternativa que parece ter o dobro do valor. É isto que é conhecido como "matching," quer na análise quantitativa do comportamento quer na psicologia matemática.

directamente com esse fenómeno central para o nosso projecto que é a acrasia, fá-lo no âmbito de um modelo do eu e da vontade. Mas o que é o eu? O fenómeno do desconto hiperbólico que tanto interessa Ainslie sugere um modelo do eu ou self como uma população de processos em busca de recompensa – interesses de curto e longo prazo de agentes são vistos como processos de procura de recompensa que competem entre si de acordo com uma regra de estrita maximização de recompensa esperada descontada em cada momento de escolha – associado a um modelo da vontade como negociação inter-temporal (*intertemporal bargaining model of the will*). E de onde vem, e o que é, então, a ‘vontade’? Segundo Ainslie, na entrevista que nos deu, “A diferença nas pessoas [relativamente a outros agentes] é provavelmente que o nosso poder de computação vastamente aumentado conduz não apenas a mais imaginação e imaginação de eventos mais remotos, mas também à modelação da imaginação de outras pessoas (“teoria da mente”) a à inclusão recursiva das nossas próprias escolhas naquilo que imaginamos, dando origem à vontade, tal como eu a imagino”.

Se a perspectiva de Ainslie focaliza o que subjaz, tanto quanto sendo agentes somos sistemas biológicos, à acção humana *par excellence*, para usar o termo de Velleman, como é que as características desta – consciência, deliberação, liberdade –, aparecerão? A questão pode colocar-se em termos da relação entre motivação e deliberação: diz Ainslie “A maioria dos processos que parecem ter sido discutidos pelos filósofos são não apenas motivados mas deliberados. Eu não quero dar uma definição intransigente [*I shrink from a hardheaded definition*], mas a ideia geral é que as acções deliberadas foram testadas no mercado da recompensa ao ponto de, quando são escolhidas, elas serem escolhidas estando o agente consciente delas e das suas alternativas imediatas. Assim, num certo sentido uma acção é deliberada se for escolhida ‘uma vez consideradas todas as coisas’ [*all things considered*]” E acerca de vontade, da qual Ainslie fala em termos da força da vontade, diz: “Assim, força de vontade não é um conceito supérfluo (como a teoria da escolha racional mantém) mas antes tem uma função necessária – a manutenção da preferência por alternativas LL [LL são ‘larger, later rewards’, que se opõem a SS ‘smaller, sooner, rewards’] ao longo do tempo”. O que nos leva à questão de saber o que são agentes que são pessoas e que são a ‘mesma pessoa ao longo do tempo’, como diria John Locke. Como é que um tal resultado – uma pessoa única – vem à existência por sobre as populações de interesses em competição retratadas por Ainsie? Acerca de pessoas e da sua identidade diz-nos

Ainslie: “Estas propriedades da acção permitem-nos ter uma teoria de como uma pessoa que é uma população de processos de procura de recompensa pode formar processos mentais superiores—“processos de ego” e uma “identidade do ego”. Uma pessoa enquanto entidade que se estende ao longo do tempo sabe que de alguma forma não pode sempre contar consigo própria – vêm as tentações, é muito possível que sucumba a elas quando estas se lhe depararem no futuro – então uma pessoa ‘negoceia consigo própria’ tais relações inter-temporais: Ainslie evoca o caso emblemático de Jon Elster, o caso de Ulisses e das sereias⁵⁰: Ulisses não quer ser tentado pelas sereias; ora, como crê que será de facto tentado quando se aproximar delas, ata-se ao mastro, planeando a sua acção de forma a impedir-se de sucumbir à tentação imediata.

António Zilhão, na entrevista publicada mais à frente no presente volume, falando de Ainslie (e de Elster) nota de forma certa que «De acordo com o pensamento de ambos, a acrasia surge com a introdução da temporalidade. Ela ocorre quando o agente não ordena a sua sucessão de escolhas no tempo de uma forma racional. Eles introduzem então na discussão a distinção entre taxa de desconto exponencial e taxa de desconto hiperbólica. E o que aconteceria com os agentes acráticos seria que a sua taxa de desconto seria tendencialmente hiperbólica em vez de ser tendencialmente exponencial. A taxa de desconto exponencial é racional, a taxa de desconto hiperbólica não é racional. O facto de muitos agentes terem uma taxa de desconto hiperbólica explicaria então porque é que as suas preferências se inverteriam temporariamente na presença de estímulos sensoriais ou emocionais muito fortes, etc».

Ainslie, com a sua teoria do self e da vontade, inspirada na psicologia behaviorista e indo buscar temas freudianos, trouxe ao nosso Projecto uma muito necessária abertura à natureza biológica de (pelo menos alguns) dos agentes de que falamos quando falamos de agência. Um interesse adicional do seu trabalho para o nosso projecto foi mostrar o quanto um modelo do eu ou self é central e incontornável na abordagem das questões da racionalidade prática.

⁵⁰ Cf. mais à frente nesta Introdução a explicação do estatuto deste fenómeno tal como é retratado por Elster.

3. 5 Daniel Hausman

Daniel Hausman é Herbert A. Simon Professor no Departamento de Filosofia da Universidade de Wisconsin – Madison. Como recorda na biografia da sua página pessoal na Universidade de Wisconsin, teve um trajecto cheio de inflexões: cresceu nos subúrbios de Chicago e frequentou o Harvard College, onde se formou primeiro em bioquímica e depois em História e Literatura inglesas. Depois de ensinar no Bronx – *intermediate school* – e de obter um grau de Master of Arts in Teaching na NYU (New York University), passou dois anos a estudar ciência morais no Gonville and Caius College, na Universidade de Cambridge (Reino Unido). Em 1978 terminou um PhD em Filosofia na Universidade de Columbia, Nova Iorque.

A sua investigação tem incidido sobre as questões metodológicas, metafísicas e éticas que surgem nas fronteiras da economia com a filosofia. Em colaboração com Michael McPherson fundou a revista *Economics and Philosophy*, tendo sido o editor desta durante os primeiros dez anos da sua existência. Coordenou *The Philosophy of Economics: An Anthology* (3ª edição 2007). Os seus livros mais importantes são *Capital, Profits, and Prices: An Essay in the Philosophy of Economics* (1981), *The Inexact and Separate Science of Economics* (1992), *Economic Analysis and Moral Philosophy* (em co-autoria com Michael McPherson, 1996), *Causal Asymmetries* (1998), e *Economic Analysis, Moral Philosophy, and Public Policy* (em co-autoria Michael McPherson in 2006).

Hausman é também o autor do artigo “Philosophy of Economics” da *Stanford Encyclopedia of Philosophy*⁵¹, artigo no qual são avançados um mapeamento e uma análise detalhada dos problemas da filosofia da economia. De forma ilustrativa, Hausman caracteriza aí a filosofia da economia como consistindo em sobretudo três tipos de investigações: (a) investigações em torno de escolha racional, (b) investigações em torno da avaliação de resultados, instituições e processos económicos, (c) investigações em torno da ontologia dos fenómenos económicos e da possibilidade de adquirirmos conhecimento sobre estes. Segundo Hausman, embora estes ramos ou orientações se intersectem de múltiplas formas no trabalho concreto do

⁵¹ Cf. <http://plato.stanford.edu/entries/economics/>

filósofo da economia, é ainda assim útil ter em mente a estruturação da filosofia da economia nestes tópicos maiores, especialmente porque eles podem ser vistos respectivamente como (1) *um ramo da teoria da acção*, (2) *um ramo da ética* (ou filosofia social e política normativa) e (3) *um ramo da filosofia da ciência*. O facto é que as teorias económicas da racionalidade, do bem-estar e da escolha social envolvem teses filosóficas substanciais e que elas são frequentemente informadas pela literatura filosófica relevante - o que lhes confere um óbvio interesse para quem se interessa, a partir de 'fora' da economia, por teoria da acção, psicologia filosófica e filosofia social e política. Quanto ao aspecto filosofia da ciência, a economia constitui um objecto particularmente interessante pelo facto de possuir tantos traços das ciências naturais (nomeadamente o uso de métodos matemáticos), sendo o seu objecto fenómenos sociais.

Na entrevista que nos concedeu, Hausman começa por notar que a identificação dos problemas maiores da filosofia da acção (a primeira pergunta do nosso *script*) depende do interesse com o qual a abordamos. No seu caso, com a economia em pano de fundo e a teoria da acção racional como referência, esses problemas maiores são a natureza das preferências, a diferença entre preferências e juízos de valor, a relação de preferências com gostos e querereres, a relação de preferências com crenças, desejos, escolha e acção. Em suma, trata-se em geral para Hausman de procurar compreender a forma como o modelo da escolha racional assumido e manipulado pelos economistas se relaciona com descrições mais comuns - mais próximas da psicologia de senso comum - da acção humana.

Difícilmente será contestado que o modelo da escolha racional dos economistas conduz a 'ângulos de visão' específicos sobre as questões da racionalidade prática. Por exemplo perante a pergunta do nosso projecto acerca do papel da deliberação na racionalidade, Hausman sublinha imediatamente que esta é ela própria controversa: em Economia assume-se que os agentes já completaram as suas deliberações, já adjudicaram, já têm um ranking de preferências. Eles escolherão de acordo com tal ranking, e isso será agir racionalmente. Nestas condições deliberar é simplesmente calcular - e nada mais precisa de ser dito sobre deliberação. Eventualmente num momento posterior colocar-se-ão múltiplos problemas sobre formação de preferências ou alteração de preferências (ou mesmo sobre a pertinência da noção de 'preferências' - um outro dos autores do nosso projecto, John Broome, prefere simplesmente falar

de bens (*goods*) de modo a não enviesar à partida as questões). Mas essas reflexões não têm por que entrar, e frequentemente não entram, no trabalho de dia-a-dia dos economistas.

Direi aqui um pouco mais acerca de agentes e ranking de preferências, uma vez que é isso que os agentes racionais são / (ou devem ser) da perspectiva da economia: agentes são entidades dotadas de um ranking de preferências e agir racionalmente ao escolher é agir de acordo com tal ranking. No seu artigo sobre filosofia da economia da *Stanford Encyclopedia* Hausman traça a história do pensamento económico desde Aristóteles, passando pelos fisiocratas, Hume e Smith para vir a sublinhar que no século xx os economistas deixaram em geral cair a roupagem hedonista de formulações anteriores do comportamento de escolha e preferência – muito ligadas à filosofia utilitarista e à ‘felicidade’ dos agentes – concentrando-se no ranking de preferências. Agentes racionais são caracterizados pelas suas preferências; preferências são rankings de objectos de escolha. Avaliar é fazer um ranking das alternativas que se enfrenta. Ora, o que é fundamental reter acerca dos agentes racionais é que estes são capazes fazer um ranking *consistente*. Isto supõe que os rankings são *completos* — para duas alternativas x e y ou o agente prefere x a y, ou prefere y a x, ou é indiferente. Em segundo lugar as preferências do agente racional são *transitivas*. A imagem do agente racional de que o economista precisa é portanto aquela que vê este como possuindo preferência completas e transitivas e escolhendo entre alternativas – as preferências são racionais se são completas e transitivas. Este ângulo de abordagem é fundamental para distinguir a economia de outras investigações sociais – o que é distintivo do inquérito económico sobre a realidade social é precisamente esta natureza (racional) das preferências do agente económico. É certo que para além das preferências racionais, precisa-se aqui de um ‘motor’, de algo como um desejo de riqueza e consumo – este é no entanto assumido como sendo simplesmente ‘natural’ aos agentes económicos racionais. É basicamente com a riqueza e o consumo que os humanos se satisfazem; preferências como a do asceta que procura a felicidade através do ascetismo são, como nota Hausman, ‘estranhas’.

Há aqui toda uma teoria da racionalidade – analisável e objectável enquanto tal, e pensar sobre a natureza desta nas suas ramificações variadas é uma das ocupações dos filósofos da economia. Pode-se dizer por exemplo que esta teoria da racionalidade é demasiado fraca, tanto quanto nada diz acerca de crenças, ou acerca do que a

racionalidade implica quando os agentes não conhecem com certeza tudo aquilo que é relevante para as escolhas que fazem. Também se pode considerá-la demasiado forte, tanto quanto por exemplo se pode argumentar que nada há de irracional em ter preferências incompletas em situações que envolvem incerteza, situações que certamente abundam nos mundos de agentes do nosso tipo. Pode ser racional suspender o juízo ou não escolher por entre alternativas cuja natureza não se compreende completamente – em virtude de que se consideraria tais posições irracionais?

Quanto à transitividade das preferências, se ela parece plausível como requisito da racionalidade, o problema é que parece haver abundante evidência experimental de que as preferências das pessoas não são de facto transitivas. Pode-se considerar que isso não faz dos agentes eles próprios irracionais, tratando-se apenas de alguns dos seus comportamentos. De qualquer forma, a interpretação de resultados acerca de alteração de preferências e escolhas ‘irracionais’ é um núcleo importante no cruzamento, nos estudos da racionalidade, dos modelos, com a investigação empírica sobre comportamento de escolha real.

Para estender a teoria da racionalidade a circunstâncias envolvendo risco e incerteza são necessários mais princípios, e instrumentos técnicos, cuja natureza caberá também ao filósofo da economia analisar, avançando nomeadamente para o domínio epistemológico do bayesianismo e das questões relativas a probabilidades subjectivas e probabilidades objectivas. De qualquer forma importa reter que quando falamos da teoria da escolha racional aquilo de que falamos é de uma formalização das condições de racionalidade, de forma a permitir cálculos. Se temos agentes com preferências completas e transitivas satisfazendo ainda uma condição de continuidade adicional, então eles podem ser representados pela função de utilidade ordinal. É possível definir uma função que representa as preferências do agente de forma tal que $U(X) > U(Y)$ se e só se o agente prefere X a Y e $U(X) = U(Y)$ se e só se o agente é indiferente entre X e Y . A função representa apenas o ranking das preferências – não contém qualquer outra informação adicional. Qualquer transformação de “ U ” que preserve a ordem representará também as preferências do agente. Quando além disso as preferências do agente satisfazem a condição de independência e outras condições técnicas podem ser representadas pela função da utilidade esperada, que tem propriedades importantes, teorizadas e problematizadas pelos economistas e filósofos da economia.

Este é apenas um esboço elementar do que está em causa na ‘passagem’ da economia para a filosofia da economia, uma vez que no nosso projecto Daniel Hausman foi precisamente o nome exemplar da filosofia da economia ‘disciplinarmente’ concebida. Na medida em que a ligação da filosofia à economia era uma das intenções iniciais do projecto, Hausman é um guia que sugerimos para um possível aprofundamento posterior.

3. 6 Joshua Knobe

Joshua Knobe é hoje um nome emblemático da filosofia experimental⁵². É neste momento Professor Assistente do Programa de Ciência Cognitiva do Departamento de Filosofia da Universidade de Yale. Estudou na Universidade de Stanford (B. A. 1996) e na Universidade de Princeton (PhD 2006), com Gilbert Harman, entre outros. O seu trabalho em filosofia experimental tem tido bastante repercussão mediática, nos media convencionais e on-line (*The New York Times*, *Slate*, *bloggingheads.tv*, etc). O seu nome é hoje associado ao ‘*Knobe Effect*’ ou “*Side-Effect Effect*”. A ideia subjacente à descoberta deste efeito foi a de pôr de lado a consulta dos filósofos às suas próprias intuições filosóficas indo em busca da forma como as pessoas comuns pensam (‘têm intuições’) sobre acção intencional⁵³. No seu mais conhecido estudo⁵⁴, Knobe confrontou transeuntes em Manhattan com o seguinte cenário: o CEO⁵⁵ de uma empresa está sentado no seu gabinete quando o Vice-Presidente de I&D entra pela porta dentro e lhe diz: “Estamos a pensar iniciar um novo programa. O programa ajudar-nos-á a aumentar os lucros, mas também será prejudicial ao ambiente”. O CEO responde que não quer saber do ambiente e só quer aumentar os lucros tanto quanto for possível. O programa é levado a cabo, os lucros sobem e o ambiente é prejudicado. Pergunta-se às pessoas: Será que o CEO *prejudicou intencionalmente* o

⁵² Outros filósofos que contariam actualmente como filósofos experimentais seriam por exemplo Shaun Nichols, Thomas Nadelhoffer, Edouard Machery, Stephen Stich, etc.

⁵³ Cf. os comentários de António Zilhão, na entrevista publicada mais à frente neste volume, sobre a filosofia experimental como movimento.

⁵⁴ Knobe, Joshua, 2006 (“The Concept of Intentional Action: a Case Study in the Uses of Folk-Psychology”, *Philosophical Studies* 130: 203-231. Reimpresso em Knobe & Nichols 2008, *Experimental Philosophy*).

⁵⁵ Chief Executive Officer.

ambiente? Uma vasta maioria de pessoas interrogadas por Knobe (82%) responderam que sim. Este é o cenário ‘*harm condition*’. Num segundo cenário, ‘*help condition*’, as coisas são postas da seguinte maneira: “O Vice-Presidente da empresa vai ter com o CEO e diz: Estamos a pensar começar um novo programa; ele ajudar-nos-á a aumentar os lucros e também beneficiará o ambiente.” O CEO responde: “Não quero saber de ajudar o ambiente. Apenas quero ter tanto lucro quanto for possível. Vamos lá começar o novo programa.” Começaram-no e o ambiente saiu beneficiado. Pergunta-se às pessoas: Neste caso o CEO *ajudou intencionalmente* o ambiente? (recorde-se que ele não quer saber de ajudar o ambiente, quer apenas ter lucro – no entanto as suas acções resultam em ambas as coisas). Confrontadas com a questão “Será que o CEO ajudou intencionalmente o ambiente?” apenas 23% das pessoas dizem que ‘sim’. É esta assimetria entre os cenários ‘prejudicar’ e ‘ajudar’ que é conhecida como “Efeito Knobe”. Knobe argumenta que, uma vez que conceitos da *folk psychology* tais como os relativos a agir intencionalmente podem ser influenciados por considerações morais gerando efeitos como o exemplificado acima, não se deve pensar neles como servindo simplesmente para prever e explicar comportamento, como é frequentemente assumido sem mais considerações. O conceito comum de acção intencional é antes, de acordo com Knobe, algo como uma ‘*multipurpose tool*’, um instrumento de usos múltiplos.

Se o que acabei de descrever é um exemplo do que se faz em filosofia experimental, neste caso em torno das intuições das “pessoas comuns”⁵⁶ acerca do que é agir intencionalmente, ele não é ainda suficientemente esclarecedor por si nem particularmente polémico – é por isso preciso dizer um pouco mais acerca das razões por que a filosofia experimental é hoje tão polémica e exaspera tanta gente no seio da filosofia académica. Muito do que se faz em filosofia experimental é investigação que poderíamos classificar como investigação em ciência cognitiva acerca de tópicos que vão desde a intenção, à consciência, ao livre arbítrio e às emoções. No entanto a filosofia experimental começou com um propósito metodológico polémico explícito relativamente à ‘tradicional filosofia de cadeirão’ (*armchair philosophy*)⁵⁷, declarando

⁵⁶ É irresistível perguntar, como um dos comentários no blogue X-Phi pergunta: significa isto que os filósofos são ‘pessoas incomuns’? E que isso tem consequências metodológicas?

⁵⁷ Cf. Joshua Knobe e Shaun Nichols, An Experimental Philosophy Manifesto, in Knobe & Nichols 2008, *Experimental Philosophy*: 3-14. Uma tradução portuguesa pode ser encontrada no blogue Filosofia Experimental, em <http://filosofiaexperimental.wordpress.com/textos-basicos-traduzidos/um-manifesto-da-filosofia-experimental/>.

que para compreender a forma como os humanos pensam e agem se deve ‘*go out and run empirical tests*’ – fazer testes e experiências – nomeadamente para mexer com as ‘intuições’ usualmente evocadas em filosofia sem mais problemas (por exemplo no caso de Knobe recordado acima, intuições acerca do que é ou não é agir intencionalmente). É por esta razão e não tanto por lidar com consciência, intenção, variação conceptual intercultural, etc, que a filosofia experimental despoleta hoje um debate metodológico importante acerca do que se faz quando se faz filosofia. Essa é também a razão pela qual muitos dos filósofos tradicionais se impacientam com a filosofia experimental, vendo nela muita *naiveté* e uma grande agitação acerca de coisa nenhuma – um ponto geral destes opositores seria: não há nada de errado com investigações empíricas acerca dos temas acima listados, mas por que razão há-de alguém como Knobe, ou qualquer outro praticante da filosofia experimental, considerar que trabalha em filosofia, por oposição a psicologia, ou ciência cognitiva? E se o que se está a fazer é investigação em psicologia ou ciência cognitiva, será que está à altura das práticas normais dessas disciplinas, feitas em diálogo com os pares e assim cientificamente controladas, ou será que o suposto interesse filosófico das experiências serve como desculpa aos filósofos experimentais para se situarem em terreno académico de ninguém e assim escaparem a críticas?

A contenda em torno da filosofia experimental pode ser facilmente reconstituída quer em inglês quer em português – devido ao blogue X-PHI de Carlos Mauro há muitas traduções disponíveis em português⁵⁸. Uma forma de reconstituir a contenda será começar por ler o manifesto de Knobe e Nichols⁵⁹ e em seguida respostas críticas (agudas), como por exemplo as de Anti Kauppinen (*The Rise and Fall of Experimental Philosophy*⁶⁰) ou comentários de Timothy Williamson no seu livro recente *The Philosophy of Philosophy*⁶¹. Anti Kauppinen termina o seu artigo dizendo que “no máximo os resultados de testes dão-nos alimento para pensamento [*food for thought*] – mas alimentar-nos-emos melhor se em vez de desenharmos experiências e situações artificiais prestarmos atenção de perto ao que é dito em situações de uso de linguagem na vida real, tal como os filósofos conscienciosos têm vindo a fazer pelo menos desde Sócrates”. Esta é obviamente uma observação metodológica acerca de

⁵⁸ Cf. <http://filosofiaexperimental.wordpress.com/textos-basicos-traduzidos/>

⁵⁹ Knobe & Nichols 2008.

⁶⁰ Kauppinen 2007.

⁶¹ Williamson 2007.

linguagem e intuições e afinal acerca do lugar da linguagem na análise do pensamento, um tema que foi metodologicamente fundamental para a filosofia analítica no século xx, na sequência de Frege. A polémica está aí, entre nós.

É verdade que um defensor da filosofia experimental poderia argumentar que a disputa não é apenas metodológica mas também temática: na entrevista que nos deu, Joshua Knobe acentuou não apenas a importância dos métodos da filosofia experimental para a filosofia da acção, mas também o facto de a filosofia experimental significar um regresso de tópicos clássicos da filosofia, tais como os tópicos tratados por Nietzsche, que dizem respeito à forma como os ‘humanos realmente são’, e que contrastam com os ‘tópicos excessivamente técnicos’ dos ‘filósofos de cadeirão’. Há um ponto aqui a sublinhar que tem tendência a exasperar os filósofos de cadeirão que alegremente se assumiriam como tal, i.e. filósofos analíticos a que os filósofos experimentais chamariam representantes do *mainstream* técnico: a filosofia de cadeirão seria caracterizada pelo ‘método de análise conceptual’, sendo este, pelo menos no *Manifesto* de Knobe e Nichols⁶², caracterizado de forma incrivelmente simplista como consistindo na ‘análise por decomposição dos conceitos que as pessoas usam’, uma definição que esquece basicamente toda a história da filosofia analítica desde fins do século XIX, aquilo que faz da filosofia analítica *filosofia analítica* propriamente dita, e que passa não apenas pela entrada em cena da lógica na investigação da natureza do pensamento como por uma reflexão semântica e metodológica acerca do significado de tal prática e pela consciência da necessidade de responder ao cientismo e psicologismo que se erguiam então, no século XIX, e que aparentemente os filósofos experimentais no século XXI têm orgulho em continuar. Que a análise de pensamentos não é análise dos ‘conceitos nas mentes das pessoas’ é algo que na história da filosofia analítica ficou claro desde Frege, i.e. desde o início. Mas essa é uma via metodológica à qual os filósofos experimentais são aparentemente insensíveis – ‘método’ nas suas palavras significa ‘método experimental’, nada tendo a ver com lógica e linguagem; se os filósofos experimentais vêm algo na história da discussão acerca de método que gerou a própria filosofia analítica, e na qual a lógica e a linguagem são incontornáveis, é simplesmente uma regressão relativamente aos temas muito ‘mais humanos’ de alguém como Nietzsche. Como afirmavam Knobe e Nichols no *Manifesto*, «A

⁶² Knobe & Nichols 2008.

esperança é em última análise que possamos utilizar esta informação para determinar se as fontes psicológicas das crenças abalam a justificação das crenças. A abordagem básica aqui envolvida deve ser-nos familiar da história da filosofia. Olhe-se apenas para a filosofia da religião do século XIX. Nesse tempo, houve um debate aceso acerca de se as crenças religiosas das pessoas eram justificadas e alguns filósofos (Marx, Nietzsche, Feuerbach, etc.) contribuíram para o debate oferecendo hipóteses acerca das fontes psicológicas da fé religiosa. Estas hipóteses conduziram a uma explosão de discussão que mostrou ser enormemente valiosa para muitíssimos assuntos. Mas então algo de estranho aconteceu. Embora argumentos deste género básico tenham sido tradicionalmente encarados como sendo extremamente importantes, eles vieram a ocupar um papel bem menos significativo na forma específica de filosofia que se tornou proeminente no século XX. A ascensão da filosofia analítica conduziu a uma diminuição do interesse por, por exemplo, as fontes da fé religiosa e a um aumento de interesse em questões mais técnicas que podiam ser tratadas a partir do cadeirão. Esta mudança foi algo peculiar. Não é que alguém tivesse avançado argumentos contra a ideia de que valia a pena compreender as fontes das nossas crenças; o que aconteceu foi antes que esta forma de inquérito tradicional saiu de moda. Nós vemos isto como um desenvolvimento lamentável. Pensamos que questões sobre as fontes das nossas crenças religiosas, morais e metafísicas são profundamente importantes e que nunca existiram boas razões para se abandonar a sua investigação. A nossa finalidade é voltar a estas questões, desta vez armados com os métodos da ciência cognitiva contemporânea⁶³».

Mas de que se fala quando se fala de tópicos nietzscheanos que teriam sido afastados pela filosofia analítica de cadeirão? Um exemplo avançado por Knobe na conversa connosco é a questão “Porque é que as pessoas acreditam em vontade livre ou livre arbítrio?”, seguida da resposta de acordo com a qual seria o desejo de justificar actos de punição (Nietzsche fala de uma ‘metafísica do carrasco’). Algo que se poderia aqui dizer é os filósofos experimentais são eles próprios muito ‘analíticos’, agora no sentido apenas cultural e não já teórico, i.e prestam atenção apenas à filosofia feita em inglês: se não tivessem os olhos fechados para a filosofia de língua francesa, seria claro que tais tópicos e um autor como Nietzsche, nunca foram abandonados:

⁶³ Knobe & Nichols 2008: 13-14.

Nietzsche é simplesmente omnipresente na filosofia dita continental do século xx, e o uso desejado por Knobe encontra-se explorado de forma sofisticada por exemplo em Foucault ou Deleuze.

No nosso projecto, para além de ter introduzido a polémica discussão em torno da filosofia experimental, Knobe é alguém que aparece como defendendo que os modelos normativos de racionalidade propostos pelos filósofos simplesmente ‘não representam possibilidades reais para criaturas como nós’. Ele vai ainda mais longe: evocando o tópico clássico da falácia naturalista⁶⁴ Knobe alerta para o perigo do que chama a ‘falácia não-naturalista’: o erro de pensar que resultados experimentais são completamente irrelevantes para questões morais, e que os filósofos podem simplesmente ignorar tudo o que sabemos sobre natureza humana (Knobe refere-se aos dados recolhidos de forma experimental na filosofia experimental) e responder a questões filosóficas (por exemplo morais, ou acerca de acção) a partir do ‘cadeirão’.

4. Um pouco mais de espaço no domínio: J. Broome, M. Smith e J. Elster

Até aqui identifiquei os autores com os quais terminámos entrevistas; tinham no entanto sido ainda seleccionados por nós outros autores com quem o processo da entrevista não chegou a avançar: é importante para a caracterização do domínio do projecto explicar o nosso interesse e a razão porque os seleccionámos.

4.1 John Broome

John Broome é um filósofo e economista inglês, White’s Professor of Moral Philosophy na Universidade de Oxford e Fellow do Corpus Christi College. Estudou na Universidade de Cambridge, na Universidade de Londres e no MIT, onde obteve o seu PhD em Economia. É hoje um dos nomes fundamentais no campo de intersecção da ética com a economia, e alguém em cuja obra a importância da economia para a ética é claramente acentuada e trabalhada. A suposição do seu trabalho é que apenas o

⁶⁴ Cf. mais à frente neste volume S. Miguens, *Hilary Putnam acerca de ética*.

reconhecimento dessa importância nos permite pensar filosoficamente sobre problemas relativos a alocação de recursos económicos, quer se se trate por exemplo de políticas médicas para o futuro, quer de questões como o aquecimento global, quer de muitas outras. Isto é manifesto em livros seus tais como *The Microeconomics of Capitalism* (1983), *Weighing Goods – Equality, Uncertainty and Time* (1991), *Counting the Cost of Global Warming* (1992), *Ethics Out of Economics* (1999) e *Weighing Lives* (2004). No seu trabalho nesta área Broome dialoga com autores bem conhecidos na ética e na filosofia política tais como John Harsanyi⁶⁵, John Rawls⁶⁶, Derek Parfit⁶⁷ ou Thomas Nagel. Mas além do seu trabalho em torno de ética e economia, Broome é ainda conhecido pelo seu trabalho acerca da natureza e requisitos da racionalidade, de que são exemplo escritos como por exemplo *Normative requirements* (1999), *Normative practical reasoning* (2001), *Are intentions reasons? And how should we cope with incommensurable values?* (2001), *Practical reasoning* (2002), *Reasons* (2004), *Does rationality give us reasons?* (2005) *Have we reason to do as rationality requires? A comment on Raz* (2005), *Wide or narrow scope* (2007), *The unity of reasoning?* (2009), *Is rationality normative?* (2008), *Does Rationality consist in responding correctly to reasons?* (2007), *Reasoning with preferences* (2006) ou *Motivation* (2009)⁶⁸.

Como Bratman, também Broome se interessa pela natureza dos requisitos de racionalidade, procurando compreender a natureza de tais requisitos. À primeira vista, a racionalidade diz-nos que devemos pensar certas coisas e fazer certas coisas (por exemplo ter crenças consistentes, ter intenções consistentes ou tencionar fazer aquilo que se pensamos ser um meio necessário para o fim que prosseguimos). Mas como havemos de tomar esse ‘dever’ que a racionalidade dita? Noutras palavras, em que sentido é a racionalidade *normativa*? A racionalidade é certamente normativa no sentido em que é um sistema de requisitos ou regras, mas, citando *ipsis verbis* um

⁶⁵ Que é, recorde-se, um importante utilitarista. Cf. o artigo de Pedro Galvão “Teoria da decisão, racionalidade e ética – o utilitarismo de John Harsanyi” em *Trólei – revista electrónica de filosofia moral e política* nº 4, 2004 (on-line no site do MLAG em <http://mlag.up.pt>).

⁶⁶ Que argumenta, recorde-se, contra o utilitarismo.

⁶⁷ Além de uma metafísica reducionista da identidade pessoal, Parfit defende o princípio utilitarista da distribuição (Parfit 1984, *Reasons and Persons*). Para Parfit, as sub-divisões em vidas de pessoas são de certa forma como as divisões entre pessoas, e isto pesa sobre a forma como se deve pensar sobre as questões da distribuição. Uma formulação geral do princípio utilitarista da distribuição seria: “Uma alternativa é pelo menos tão boa como outra se e só se a ela corresponde um total idêntico de bem de pessoas” (cf Broome 1991: 16).

⁶⁸ Cf. referências no Apêndice do presente volume.

título de Broome, dizer isso não nos diz ainda *se temos razões para fazer aquilo que a racionalidade requer*. Será que a racionalidade nos dá razões para fazer φ , razões para seguir os requisitos da racionalidade? A racionalidade dá-nos, ou coloca perante nós, regras de correcção mas há uma questão posterior que não fica respondida apenas com essa colocação: será que devemos seguir as regras que a racionalidade nos dá? Temos razões para isso? São estas algumas das perguntas ‘disruptivas’ de Broome, que não aceita que exista simplesmente uma conexão conceptual entre racionalidade e razões e que seja, por exemplo, simplesmente irracional agir contra razões. Na verdade Broome revela-se céptico quanto à resposta que se deve finalmente dar à questão de saber se devemos seguir as regras que a racionalidade nos dá⁶⁹, acabando por concluir que não é possível mostrar que os requisitos da racionalidade são normativos, e assim rejeitando a tese da ‘Normatividade forte’⁷⁰ e mesmo da ‘Normatividade fraca’⁷¹ acerca da racionalidade.

Não avançarei mais aqui mais aprofundadamente pelo trabalho de Broome sobre requisitos de racionalidade – passo a focar o seu trabalho ‘ético-económico’, tomando como exemplo os temas com que lida no livro *Weighing Goods* (1991), ‘um livro sobre ética que usa alguns métodos da economia’⁷², e onde Broome investiga os ‘traços da estrutura do bem’ e a ‘*betterness relation*’ (‘a relação de melhor’⁷³), em diálogo com a tradição filosófica utilitarista e com as críticas a esta, e sem evitar incursões metafísicas acerca de pessoas e bens, que obviamente o obrigam a dizer o que entende quer por ‘pessoas’ quer por ‘bens’ (no primeiro caso, a sua *dis-uniting metaphysics*, metafísica da desunidade, responde a desafios e teses vários do célebre livro *Reasons and Persons* de Derek Parfit⁷⁴). O objectivo de *Weighing Goods* é uma proposta acerca de como ‘ponderar ou pesar’ [*weigh*] bens.

⁶⁹ Broome 2007, Is rationality normative?

⁷⁰ Formulada em Broome 2007 como: “Necessariamente, se a racionalidade requer de (uma pessoa) N que F, N deve [*ought to*] fazer F, porque a racionalidade requer que N (faça) F”.

⁷¹ Formulada em Broome 2007 como: “Necessariamente, se a racionalidade requer de (uma pessoa) N que F, há uma razão para N (fazer) F”.

⁷² Broome 1991: ix.

⁷³ “Some things are better than others, so that a relation of betterness holds between them” (Broome 1991: 1).

⁷⁴ Parfit 1984. O livro de Parfit assume especial importância por ter colocado directamente, na discussão ético-metafísica, a questão da *unidade moralmente significativa*. Esta é uma questão decisiva também para Broome nos seus escritos: será essa unidade a *pessoa*? ou serão *experiências*, cuja forma de agregação em ‘pessoas’ coloca ela própria vários problemas? Essa opção em última análise metafísica é fundamental para a abordagem de questões de distribuição e ‘bem’. Algumas das ideias de

Pretendendo ser ouvido quer por filósofos quer por economistas, e também obter aplicações económicas práticas das suas propostas, Broome tem que explicar aos filósofos por que razão os números e os teoremas formais comuns na prática dos economistas importam tanto para se pensar em questões éticas e que explicar aos economistas por exemplo por que prefere falar do bem [*good*] em vez de preferências, ao contrario do que é usual em economia. O seu ‘*principle of personal good*’ – princípio do bem pessoal⁷⁵ – será fundamental para articular as propostas do livro, nomeadamente quanto à relação de melhor [*betterness relation*]. De acordo com Broome o princípio do bem pessoal aplica-se a *probability relative goodness* [bem relativo a probabilidades].

É o bem (de uma pessoa) e não as preferências desta que Broome toma como conceito central na sua discussão – e esta é uma opção, relativamente à qual ele se explica, sobretudo perante os economistas, que usam de preferência as ‘preferências’ para falarem de indivíduos e suas relações. O bem ocorre em diferentes locações (*locations*). A análise em *Weighing Goods* prossegue visando a forma como os bens localizados ou alocados em cada uma de três dimensões – pessoas, tempos, estados de natureza – compõem o ‘bem’ (*goodness*) global. Evidentemente isto coloca a questão de saber como é que o bem é agregado. Divisão, continuidade, agregação são as questões teóricas que se colocam para todas as dimensões – pessoas, tempo, estados

Parfit retomadas e discutidas por Broome são, por exemplo, o Paradoxo da Mera Adição, considerando vidas que poderiam não existir (por exemplo um filho que um casal decide não ter) do ponto de vista da quantidade de ‘bem’ (cf. Broome 1991: 15-16), a descrição da justiça distributiva avançada no capítulo XV (Personal Identity and Morality) de *Reasons and Persons* (Broome 1991: 39-51), considerando além da distribuição do bem na dimensão das pessoas, a distribuição do bem ao longo da dimensão do tempo da vida de uma pessoa (obviamente colocando a questão acerca do que faz a ‘unidade’ dessa pessoa) e a ‘*priority view*’, de acordo com a qual a contribuição para o bem geral é maior quanto pior (*worse off*) a pessoa ficar (o princípio entra na discussão do igualitarismo) (Broome 1991: 198-199). Um outro autor convocado por Broome na discussão das questões metafísicas da identidade pessoal, tanto quanto estas têm um reflexo directo nas discussões ético-económicas, é David Lewis.

⁷⁵ “...the good of individuals determines general good (so there can be no increase in general good without an increase in some individual’s good)” (Broome 1991: 36). Este é, segundo Broome, um ‘princípio de separabilidade no domínio das pessoas’, que também pode ser lido como dizendo que uma alternativa não pode ser melhor do que outra a não ser que seja melhor para alguém. Broome também formula o princípio – nomeadamente para criticar o Princípio de Pareto da seguinte forma: “*The principle of personal good* (a) Two alternatives are equally good if they are equally good for each person. And (b) if one alternative is at least as good as another for everyone and definitely better for someone, then it is better”. O Princípio de Pareto diria antes “(a) Two alternatives are equally good if everyone is indifferent between them, and (b) if everyone either prefers the first of two alternatives to the second or is indifferent between the two, and someone definitely prefers the first, then the first is better than the second” (Broome 1991: 152)

de natureza. Por exemplo no caso da dimensão pessoas, em que circunstâncias serão os bens para uma pessoa ao longo do tempo os bens de *uma mesma* pessoa?

Broome argumenta que as três dimensões acima estão ligadas por aquilo a que chama o *Teorema da adição interpessoal*, paralelo ao teorema de Harsanyi⁷⁶, que sustenta o princípio utilitarista de distribuição. O Teorema da adição interpessoal formula-se da seguinte forma: “Se as relações de melhor [*betterness relations*] individual e pessoal são coerentes, e se o princípio do bem pessoal é verdadeiro, então a relação de melhor geral pode ser representada por uma função de utilidade expectacional que é a soma das funções de utilidade expectacionais que representam as relações de melhor individuais”⁷⁷. A convicção de Broome é que este teorema diz algo de muito significativo acerca da estrutura do bem: diz-nos, nomeadamente, de algo acerca da forma como o bem individual é agregado para compor o bem geral. Essa é de resto segundo Broome, a principal razão porque *Weighing Goods* é escrito em torno do bem e não das preferências – na sua opinião, o teorema paralelo acerca de preferências não chega a dizer nada de significativo pois as suas premissas são inconsistentes⁷⁸. Qual é a importância geral desta discussão? Se a estrutura do bem tem as características que Broome pretende que tem, isso é extremamente importante para a economia do bem estar (*welfare economics*), já que uma preocupação central desta é a agregação ‘do bem ‘entre’ [*across*] pessoas, quando se trata de avaliar situações sociais e escolhas entre resultados alternativos quanto a estas.

Em *Weighing Lives* (2004) Broome continua esta investigação, considerando a intuição segundo a qual adicionar pessoas é algo de eticamente neutro, que acaba por rejeitar. Estes trabalhos sistemáticos são complementados pela sua colecção de artigos *Ethics Out of Economics* (1999), onde se encontram escritos na intersecção da filosofia moral com a teoria económica sobre, por exemplo valor, igualdade, equidade (*fairness*) e utilidade.

Alguns outros artigos importantes sobre questões de igualdade, valor, economia e ética são por exemplo “What’s the good of equality?” (1989), “Equality versus

⁷⁶ “Suppose that each person has coherent preferences, and that social preferences are coherent and Paretian. Then social preferences can be represented by an expectational utility function that is the sum of expectational utility functions representing the individual’s preferences.” (Broome 1991: 162)

⁷⁷ Broome 1991: 162-163

⁷⁸ Broome 1991: 163

priority: a useful distinction”, “Respects and levelling down” “Should we value population?” (2005), “Can There Be a preference-based utilitarianism?” (2007), “Why economics needs ethical theory”, “What is your life worth?” ou “The ethics of climate change” . Broome tem dedicado muita atenção a este último tópico, como se pode ver por exemplo em “Counting the Cost of Global Warming” (White Horse Press, 1992), cujo texto se encontra integralmente on-line e onde a relação deste tópico com questão a justiça inter-geracional é central, ou “Valuing policies in response to climate change: some ethical issues”, “The ethics of climate change”, “The most important thing about climate change”.

Uma outra manifestação do interesse de Broome por aplicações práticas das suas investigações é o seu trabalho com a Organização Mundial de Saúde (World Health Organization), disso sendo exemplos artigos como “Measuring the burden of disease by aggregating wellbeing”, “Fairness, goodness and levelling down”, “All goods are relevant”, “Measuring the burden of disease” ou “Equality versus priority: a useful distinction” .

Como Amartya Sen, John Broome é um nome central a ter hoje em mente quando queremos considerar as relações entre ética e economia. Neste caso encontramos alguém que está em diálogo constante com a filosofia ético-política de tradição analítica, e em cuja obra é claríssima a importância prática – ética, económica, política – enorme das questões da racionalidade prática.

4.2 Michael Smith⁷⁹

Michael Smith é desde 2009, McCosh Professor of Philosophy em Princeton, onde é também Associated Faculty Member do Departamento de Política. Estudou (BA e MA) na Monash University (Austrália) e em seguida na Universidade de Oxford (BPhil e DPhil), onde foi orientado por Simon Blackburn. Os seus interesses básicos incluem a ética, a psicologia moral, a filosofia da mente e da acção, a filosofia política e a filosofia do direito. É autor do conhecido livro *The Moral Problem*

⁷⁹ Agradeço à Susana Cadilha as notas de leitura da obra de Smith.

(Blackwell, 1994), e também de *Ethics and the A Priori: Selected Essays on Moral Psychology and Meta-Ethics* (CUP, 2004). É co-autor de *Mind, Morality and Explanation: Selected Collaborations* (OUP, 2004), uma colecção de artigos escrita em combinações de autoria várias pelo próprio Smith, por Frank Jackson e por Philip Pettit. Coordenou *Meta-Ethics* (Dartmouth, 1995), *Reason and Value: Themes from the Moral Philosophy of Joseph Raz* (OUP, 2004) (com Jay Wallace, Philip Pettit, e Samuel Scheffler); *The Oxford Handbook of Contemporary Philosophy* (OUP, 2005) (com Frank Jackson); e *Common Minds: Themes from the Philosophy of Philip Pettit* (OUP, 2007) com Geoffrey Brennan, Robert Goodin, e Frank Jackson.

Foi a particular forma de colocar a natureza do problema moral em *The Moral Problem* que tornou este livro tão discutido e emblemático: Smith aborda directamente a tensão que considera existir nos juízos morais entre, por um lado, a sua (possível) objectividade e envolvimento com questões de racionalidade, e por outro lado o seu carácter prático, motivador da acção. Nestas circunstâncias, qual poderá ser a real natureza dos juízos morais? Se existe objectividade moral, as questões morais são questões de facto (*matters of fact*) e as respostas que lhes damos podem estar correctas (*right*) ou incorrectas, e os argumentos morais são uma espécie dos argumentos racionais, cuja finalidade é descobrir a verdade - neste sentido estaríamos num domínio de racionalidade. Mas por outro lado se os juízos morais são práticos, eles impelem ou movem a agir aqueles que os aceitam. Se nós somos agentes de um tipo tal que aquilo que nos impele à acção são os desejos, e os choques dos desejos, esta forma como os requisitos morais se nos apresentam, solicitando racionalmente coisas, tem algo de estranho. Ou eles não são crenças mas nesse caso não poderão ser objectivos, ou eles são de facto crenças mas nesse caso não poderão ser motivadores. Isto é um problema, o problema moral. Ele é formulado nos termos das duas características da moralidade, a objectividade do juízo moral e a practicalidade deste [*objectivity of moral judgement, practicality of moral judgement*]. Estas que ‘fazem pressão’ em direcções opostas, ameaçando tornar a ideia de moralidade incoerente⁸⁰. E de acordo com Smith o problema moral é o problema organizador central da meta-ética, sendo que a meta-ética deve ser o ponto de partida do estudo da ética.

⁸⁰ Smith 1991, 5-7.

As posições específicas que Smith defenderá em *The Moral Problem*, após uma discussão de autores e temas importantes para se compreender o seu posicionamento no debate ético e meta-ético contemporâneo, e que vão de Ayer (o emotivismo) a Phillipa Foot ou a seu professor Blackburn, até Nagel e McDowell, são o realismo moral e o cognitivismo. Mas defender tais posições é algo que só poderá ser feito uma vez tendo ajustado contas com a ‘concepção standard da natureza da psicologia humana’, que Smith considera ser a visão humeana acerca de crenças e desejos, no âmbito da qual se ergue uma particular e muito influente forma de conceber a motivação para a acção. Voltemos então à motivação – de que falamos quando falamos de *motivação*, e dos juízos morais como sendo *motivadores*? Smith ele próprio defende uma teoria humeana da motivação para a acção. Hume, como se sabe, defendeu que a razão sozinha é insuficiente para nos motivar a agir: a razão só pode ser escrava das paixões. Noutras palavras, são precisos *desejos* e não apenas representação de estados de coisas para os humanos serem movidos à acção. Os neo-humeanos contemporâneos não têm de se ater de forma totalmente estrita às afirmações de Hume, mas de acordo com a sua teoria humeana da motivação não é possível uma crença motivar alguém sem a presença de um desejo que essa pessoa já tenha. Ora, como se sabe também, os desejos, para Hume, não são de todo racionais. A questão que se coloca a Smith é saber se também aí deve seguir Hume.

Na análise de Smith, uma acção só pode ser explicada por um par crença-desejo, e crenças e desejos são, como Hume manteve, naturezas distintas, existências distintas. Esta posição vai ser apoiada numa concepção disposicional de desejo. Mas por outro lado Smith tem uma concepção anti-humeana de razões normativas, rejeitando o princípio da razão instrumental, formulado como a ideia de que os imperativos da razão são hipotéticos e não categóricos. O que será então eu ter uma razão normativa para φ (fazer algo)? Para Smith é ter um desejo de que eu faça φ - se eu fosse plenamente racional eu desejaria fazer φ na circunstância C. Em geral Smith pretende resolver o problema moral acima caracterizado defendendo uma teoria dos juízos morais em *termos daquilo que desejaríamos se fôssemos plenamente racionais*. Smith defende ao contrário do que Hume faria, que *as razões podem conduzir a certos desejos*, ou à eliminação de certos desejos, o que – na sua leitura de Hume, que outros humeanos, como por exemplo Simon Blackburn, criticariam - Hume não admitia.

Dadas certas crenças, é racional que eu tenha certos desejos. Os requisitos morais são requisitos racionais.

Mas será realmente possível chegar à posição de Smith mantendo Hume em pano de fundo? Uma outra forma de ver o que Smith está a fazer é vê-lo como equacionando razões normativas com aquilo que o *'self racional'* desejaria – Razões normativas são, podem ser, o objecto das nossas crenças (este é um ponto anti-humano da posição de Smith) e essas crenças podem ter um papel causal na produção da acção. Em suma, se somos plenamente racionais, numa circunstância dada desejaríamos tal-e-tal (convém notar, apesar de esta posição poder soar kantiana, que Kant não faria nunca uma tal ligação entre desejo e moralidade). Quando não chegamos a desejar aquilo que é racional que desejemos, ou quando não chegamos a elaborar a crença certa acerca do que devemos desejar em dadas circunstâncias, agimos de forma irracional. Noutras palavras, somos irracionais quando desejamos o que não é desejável. Em contraste, Smith pensa acerca da psicologia do *self* racional que este é plenamente informado e que tem um conjunto de desejos (*desire set*) maximamente unificado e coerente.

No núcleo das posições de Smith está, como já se percebeu, uma teoria das relações da (nossa) psicologia moral com a metafísica, e o resultado último das suas investigações é sustentar o realismo moral. Assim, no nosso projecto Smith representa a importância da psicologia moral para uma perspectiva metafísica sobre as questões da moralidade – a sua última palavra é que as questões normativas são questões acerca de matéria de facto, sendo que no entanto é ao mesmo tempo possível continuar a aceitar a força motivacional dos juízos morais. É uma posição de referência, mesmo que entendamos não a aceitar.

4.3 Jon Elster

Jon Elster é um filósofo norueguês radicado nos Estados Unidos que ensina na Universidade de Columbia, em Nova Iorque, e também no Collège de France, em Paris, onde é titular da Cátedra de Racionalidade e Ciências Sociais desde 2006⁸¹.

⁸¹ Cf. http://www.college-de-france.fr/default/EN/all/rat_soc/index.htm Podem ser aqui encontrados materiais vários relativos aos cursos de Elster, nomeadamente Powerpoints, podcasts e vídeos.

Estudou em Oslo e em Paris, onde defendeu a sua tese «*Production et Reproduction: Essai sur Marx*» em 1972 (Doctorat ès lettres et sciences humaines). Ensinou na Universidade de Paris VIII (Sociologia), vindo posteriormente a integrar os departamentos de Filosofia e de História de Oslo. Nos anos 80 começou a ensinar ciência política nos Estados Unidos, primeiro na Universidade de Chicago, depois como Robert K. Merton Professor of Social Science, na Columbia University em New York. Foi um dos nomes centrais do chamado ‘marxismo analítico’⁸². Da imensa obra de Elster, considere-se apenas alguns dos livros que escreveu e coordenou: *Leibniz et la Formation de l'Esprit Capitaliste* (1975); *Logic and Society* (1978); *Ulysses and the Sirens* (1979); *Explaining Technical Change* (1983); *Sour Grapes* (1983); *Making Sense of Marx* (1985) (a partir da sua tese de doutoramento em francês); *An Introduction to Karl Marx* (1986, coord. com A. Hylland); *Foundations of Social Choice Theory* (1986) (coord); *The Multiple Self* (1986, coord); *Rational Choice* (1986, coord); *Karl Marx: A Reader* (1986); *The Cement of Society* (1989); *Constitutionalism and Democracy* (1988, coord. com R. Slagstad); *Solomonic Judgments* (1989); *Nuts and Bolts for the Social Sciences* (1989); *Alternatives to Capitalism* (1989, coord. Com K.O. Moene); *Psychologie politique* (1990); *Interpersonal Comparisons of Well-Being* (1991, coord. com J. Roemer); *Local Justice* (1992); *Choice over Time* (1992, coord. com G. Loewenstein); *Ethique des choix médicaux* (1992, coord. Com N. Herpin); *Local Justice in America* (1995, coord.); *The Round Table Talks in Eastern Europe* (1996, coord); *Deliberative Democracy* (1998, coord); *Getting Hooked: Rationality and Addiction* (1999, coord. Com O.J. Skog); *Addiction: Entries and Exits* (1999, coord.); *Alchemies of the Mind: Rationality and the Emotions* (1999); *Strong Feelings: Emotion, Addiction and Human Behavior* (1999); *Ulysses Unbound* (2000); *Closing the Books: Transitional Justice in Historical Perspective* (2004); *Retribution and Restitution in the Transition to Democracy* (2006, coord); *Understanding Choice, Explaining Behavior: Essays in Honour of Ole-Jørgen Skog* (2006 coord, Elster et al.); *Explaining Social Behavior*

⁸² O marxismo analítico foi uma abordagem desenvolvida nos anos 80 por filósofos e cientistas sociais do mundo anglófono (associada ao chamado ‘Grupo de Setembro’, em referência aos seus encontros bienais). Os marxistas analíticos descreviam a sua abordagem como “*Non-Bullshit Marxism*”; noutras palavras, eles pretendiam – ou pelo menos viam-se a si próprios a essa luz – pensar de forma clara e rigorosa sobre as questões de um quadro marxista que acreditavam serem ao mesmo tempo fundamentais e frequentemente ocultadas por um véu ideológico. Alguns membros do grupo foram G. A. Cohen, John Roemer, Jon Elster, Adam Przeworski, Eric Olin Wright, Philippe van Parijs, e Robert-Jan van der Veen.

(2007); *Agir contre soi* (2007); *Le désintéressement* (2009); *The First Social Scientist – A study of Alexis de Tocqueville* (2009); *L'irrationalité* (2010).

Como nota um outro autor no nosso projecto, Vasco Correia, que trabalha actualmente com Elster, na conversa que se encontra mais à frente neste volume, o pensamento de Elster tem um duplo enraizamento académico francês e anglo-saxónico, o que lhe traz uma particular riqueza. Do lado “francês” encontramos nos inícios da sua carreira um doutoramento feito com o sociólogo, filósofo e teórico político Raymond Aron, um nome importante da geração de Sartre⁸³, uma tese sobre Marx e o interesse por moralistas dos séculos XVII e XVIII tais como Montaigne, Pascal, La Rochefoucault e La Bruyère. Do lado anglo-saxónico encontramos uma atenção, e um envolvimento, em estudos da racionalidade directamente ligado a disciplinas empíricas como a economia e a psicologia. A isto junta-se o gosto de Elster (um gosto metodologicamente reflectido e justificado) por análises históricas, sociais e políticas específicas: muito do seu trabalho tem como foco a compreensão de tempos e acontecimentos particulares, nomeadamente momentos de desenho e criação de constituições, em particular a constituição francesa e a constituição americana, e momentos transição para a democracia (por exemplo no Leste europeu); após a queda do comunismo, Elster devotou muito tempo e estudo à observação dos novos sistemas políticos que aí surgiam. Na verdade, o trabalho sobre constitucionalismo é muito importante na sua obra, e estende-se desde a democracia ateniense, passando pela Convenção Federal americana até ao presente⁸⁴. Noutras palavras, estamos perante um leque riquíssimo de materiais teóricos, empíricos, históricos, sociais e literários, todos eles usados como instrumentos de abordagem dos fenómenos que centralmente interessam Elster: os fenómenos da racionalidade e da irracionalidade individual e colectiva.

É importante compreender que este interesse de Elster pelos fenómenos da racionalidade e irracionalidade é um interesse que aparece do ponto de vista das ciências sociais. Elster é fundamentalmente um filósofo (e psicólogo...) das ciências sociais: as sua abordagens da racionalidade e irracionalidade constituem, como sublinha também Vasco Correia, uma reflexão crítica sobre os próprios fundamentos destas. Elster é nomeadamente um crítico da teoria da escolha racional, que é uma

⁸³ Sartre e Raymond Aron foram colegas na École Normale Supérieure, em Paris, antes da IIa Grande Guerra.

⁸⁴ Elster 2000, Preface x.

referência metodológica central nas ciências sociais e em particular na economia. A sua estratégia para questionar a universalidade dos axiomas da teoria da escolha racional é confrontá-la com fenómenos que ao mesmo tempo lhe escapam e que são extremamente, ou razoavelmente, comuns, nomeadamente fenómenos de irracionalidade como a acrasia e o auto-engano, mas também fenómenos como a justiça ou o desinteresse (a acção desinteressada).

A marca da escrita de Elster são as ‘análises interdisciplinares e a procura dos mecanismos⁸⁵’ – ora, esta forma de fazer investigação constitui só por si uma posição acerca do que a história e as ciências sociais podem e não podem fazer enquanto ciências. De acordo com Elster, elas não podem, nomeadamente e em contraste com as ciências naturais, encontrar leis – pensar que esse poderia ser o caso é para Elster uma posição implausível e frágil; no entanto a impossibilidade de encontrar leis não relega a prática das ciências sociais para a mera produção de narrativas. Os mecanismos da racionalidade e irracionalidade constituem, na perspectiva de Elster um foco alternativo para o olhar do cientista social. Esses mecanismos – padrões causais recorrentes e inteligíveis, despoletados em condições geralmente desconhecidas e com consequências indeterminadas – dão-nos um módico de poder explicativo. Foi esta descoberta que segundo o próprio Elster o resgatou da sua tendência para um niilismo explicativo, um niilismo que se lhe impunha ao constatar a total impossibilidade de prever fenómenos sociais, históricos e económicos, bem como o falhanço de todas as previsões perante acontecimentos históricos, como por exemplo a queda do comunismo na Europa de Leste. *Mecanismos* constituem então uma plataforma intermédia de inteligibilidade. Em grande medida estes mecanismos dizem respeito ao eu, e às relações deste consigo próprio – por exemplo na auto-imposição de restrições à liberdade de escolha. As questões em causa dizem por exemplo respeito à posição em que cada um de nós se encontra quanto a si próprio ao longo do tempo, e nomeadamente a nossa relação como o nosso próprio eu futuro. Recorde-se o famoso caso lançado à discussão pelo filósofo inglês Derek Parfit⁸⁶, o caso do nobre russo, que Elster recorda em *Ulysses Unbound*. Este, ao doar a sua fortuna à revolução socialista diz à sua mulher ‘Se alguma vez me vires a pretender recuar quanto a esta acção que agora pratico, eu já não serei eu, e por isso não deves

⁸⁵ Cf. Elster 1999, A Plea for Mechanisms, in *Alchemies of the Mind – Rationality and the Emotions*.

⁸⁶ Cf. Parfit 1984, *Reasons and Persons*.

dar-me ouvidos'. Recuar é no entanto precisamente aquilo que o nobre russo pretende fazer vinte anos passados, e pede ajuda à mulher. O que deve esta fazer quando ele lhe pede que o ajude? O nobre russo que via a sua identidade essencialmente ligada ao acto de doar toda a fortuna veio a ter um eu futuro que lamentou profundamente tal acção – mas quem ‘manda’ no eu que eu sou? A questão aplica-se a todos nós: como pode cada um de nós saber com o que pode contar de si próprio no futuro? Trazendo ao presente esta posição, concluímos no mínimo que enquanto indivíduos temos boas razões para querermos proteger-nos de nós próprios, das nossas paixões, da alteração da nossas preferências e de inconsistências temporais de diferentes tipos.

É nestas circunstância relativas ao nosso eu, e ao nosso eu futuro, e ao que podemos fazer para ter controlo, assumindo que este controlo pode ser perdido, que como diz Elster, por vezes, e ao contrário do que pode parecer, *less is more*: se é verdade que podemos pensar que o ‘progresso moral’ na história, a ter existido, consistiu no levantamento de restrições materiais e legais de vários tipos, é verdade também que pode ser vantajoso, em circunstâncias específicas, termos menos opções para a nossa conduta futura. Há constrangimentos que são benéficos – este é um tema importante em *Ulysses Unbound*. E os constrangimentos criados são uma espécie de racionalidade ao longo ou através do tempo.

Anteriormente, em *Ulysses and the Sirens*, Elster tinha tratado, de uma forma que se tornou muito conhecida, o fenómeno do *compromisso prévio (pre-commitment) ou auto-restrição (self-binding)*. É o compromisso prévio que é ilustrado pelo célebre exemplo de Ulisses e das Sereias, sendo este um mecanismo que Elster usa nomeadamente para pensar na natureza das constituições políticas – a sua tese era então que a auto-restrição era a própria natureza da elaboração de constituições. O caso de Ulisses e as sereias representa portanto o mecanismo do compromisso prévio, uma estratégia para lidar com o futuro de indivíduos e grupos, em situações em que se vislumbra possível perda de controle: Ulisses parte do princípio que será incapaz no futuro próximo de resistir ao canto das sereias, pede por isso que o amarrem ao mastro da embarcação. Quando o seu ‘eu futuro’ ouvir as sereias, estará preso e resistirá. Um mecanismo assim é muito importante para pensar na natureza das leis: uma ideia influente de Elster é que as constituições políticas que tanto o interessam historicamente podem ser vistas como dispositivos de compromisso prévio (é verdade que Elster veio posteriormente, em *Ulysses Unbound*, a defender que a ideia de auto-

restrição era excessiva: socialmente as pessoas não tentam tanto *bind themselves* (coagir-se a si próprias) como *bind others* (coagir as outras)⁸⁷). E o que nos pode levar a querer tanto a auto-restrição? Por exemplo as paixões - elas constituem razão para uma auto-restrição (*self-binding*). E é precisamente o interesse pela natureza e controlo das paixões que conduz Elster ao diálogo com toda uma tradição filosófica, de Aristóteles a Freud, sobre a natureza destas.

Um outro exemplo de mecanismo é o mecanismo das uvas amargas (*sour grapes*). Na fábula de La Fontaine há uma raposa que, morta de fome, depara com belas uvas, lá no alto. Adoraria comê-las, mas como não consegue chegar-lhes diz: “estão demasiado verdes e são para os grosseiros”. O mecanismo é portanto um mecanismo de menosprezo daquilo que não se pode alcançar, uma menorização do valor disso que não se poderá obter e um ajuste das aspirações às possibilidades. Está em causa a formação adaptativa de preferências, ou alteração adaptativa de preferências. Elster utiliza-o para lidar com problemas nos fundamentos da teoria utilitarista, e nas abordagens da ética e da questão da justiça em contexto utilitarista.

De acordo com Elster ‘o raciocínio por mecanismos’, tal como exemplificado pelo compromisso prévio e pelas ‘uvas amargas’, tem um lugar privilegiado no pensamento dos moralistas, e é por essa razão que estes tanto o interessam. Recorde-se que *mecanismos permitem explicar, mas não prever*. Mecanismos ‘psicológicos’, ou melhor mecanismos da racionalidade e irracionalidade nos indivíduos, tendo em mente compreender a motivação individual, tendo o individualismo metodológico como princípio, são então o instrumento de Elster para tentar compreender, tanto quanto isso é possível, a natureza das decisões colectivas, a natureza das maiorias e as decisões nessas condições. E esse é o muito característico olhar de Elster sobre os fenómenos da racionalidade prática, um olhar que o leva desde figuras literárias e mitológicas, psicologicamente iluminadoras, como Ulisses ou Medeia, aos regimes pós-comunistas no Leste Europeu, passando pela natureza da democracia na América.

⁸⁷ Elster 2000, Preface ix.

Alfred Mele, Hugh McCann, Michael Bratman, George Ainslie, Daniel Hausman, Joshua Knobe, John Broome, Michael Smith e Jon Elster foram os autores centrais do Projecto “*Conversations on Practical Rationality and Human Action*”; os seus trabalhos e os seus temas definiram o domínio do projecto. Por sua vez o projecto constituiu o núcleo da área Acção Agência e Racionalidade do MLAG entre 2007 e 2011. Se as entrevistas resultantes compõem o volume em inglês, que era um primeiro objectivo do projecto, o presente volume é de alguma forma posterior em espírito, e as entrevistas e conversas com (e entre) autores portugueses que o compõem, nas suas duas Partes (Parte I – Filosofia da Acção e Racionalidade Prática e Parte II – Ética, Juízos e Objectividade), de alguma forma pressupõem todo esse campo e esse trabalho. Espero que o presente volume sirva assim para deixar registadas algumas das discussões que marcaram os anos de trabalho no Projecto *Conversations*. Ao longo destes anos, lidamos com tópicos extremamente variados, desde tópicos que poderíamos encontrar na psicanálise e no marxismo, até discussões metodológicas da filosofia experimental, até repercussões teológicas das questões da acção. As incursões pelas questões da racionalidade prática e acção humana geradas pelo Projecto levaram-nos assim a cobrir, é essa a nossa convicção, um território vasto e fértil da filosofia contemporânea.

Porto-Londres, Fevereiro de 2011

Sofia Miguens Travis

Investigadora-Responsável do Projecto *Conversations on Practical Rationality and Human Action*
(2007-2010)

Referências

As referencias bibliográficas relativas à obra de Alfred Mele, Hugh McCann, Michael Bratman, George Ainslie, Daniel Hausman, Joshua Knobe, John Broome, Michael Smith e Jon Elster - os autores centrais do Projecto - encontram-se no Apêndice final do presente volume. Abaixo incluem-se apenas as restantes referências desta Introdução.

- Anscombe, G. E. M., 1957, *Intention*, Cambridge MA, Harvard University Press.
- Aristóteles, [2009], *Ética a Nicómaco*, Lisboa, Quetzal, trad. António Caeiro.
- Davidson, Donald, [1963] Actions, Reasons and Causes, in *Essays on Actions and Events*, Oxford, Oxford University Press, 1980.
- Galvão, Pedro, “Ética”, in *A Filosofia por Disciplinas*, no prelo.
- Guisán, Esperanza, 2010, *Introducción a la Ética*, Madrid, Ediciones Cátedra.
- Kauppinen, Anti, 2007, “The Rise and Fall of Experimental Philosophy” *Philosophical Explorations* 10 (2): 95-118.
- Miguens, Sofia, 2002, *Uma Teoria Fisicalista do Conteúdo e da Consciência – D. Dennett e os debates da filosofia da mente*, Porto, Campo das Letras, Coleção Nous.
- Miguens, Sofia, 2004, *Racionalidade*, Porto, Campo das Letras, Coleção Nous.
- Miguens e Mauro, 2006, *Perspectives on Rationality*, Porto, FL-UP, Coleção MLAG Discussion Papers.
- Parfit, Derek, 1984, *Reason and Persons*, Oxford, Oxford University Press.
- Putnam, Hilary, 2002, *The Collapse of the Fact-Value Dichotomy*, Cambridge MA, Harvard University Press,
- Putnam, Hilary, 2004, *Ethics Without Ontology*, Cambridge MA, Harvard University Press,
- Sen, Amartya, 1973, "On Economic Inequality", Oxford, Clarendon Press (edição expandida e revista: 1997)
- Sen, Amartya, 1979, Equality of What? (Tanner Lecture on Human Values, delivered at Stanford University, May 22nd 1979)
- Sen, Amartya, 1981 "Poverty and Famines: an essay on entitlement and deprivation", Oxford, Clarendon Press.
- Sen, Amartya, 1987 "On Ethics and Economics", Oxford, Blackwell.
- Sen, Amartya, 1999, *Development as Freedom*, Oxford, Oxford University Press.
- Williamson, Timothy, 2007, *The Philosophy of Philosophy*, Oxford, Blackwell.

Notas sobre os membros da Equipa de Investigação do Projecto “Conversations on Practical Rationality and Human Action”⁸⁸:

Sofia Miguens é Professora Associada com Agregação no Departamento de Filosofia da Faculdade de Letras da Universidade do Porto e investigadora no Instituto de Filosofia da mesma faculdade. As suas

⁸⁸ Para mais informações e para outros escritos on-line, cf. o site principal do Mind Language and Action Group, <http://mlag.up.pt>

principais áreas de ensino e investigação são a filosofia da mente e da linguagem e a epistemologia. Foi *visiting scholar* na Universidade de Nova Iorque (Fall 2000) e *visiting research fellow* no Institut Jean Nicod em Paris (2007-2008). 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), *Será que a minha mente está dentro da minha cabeça? Da ciência cognitiva à filosofia* (2008) e *Compreender a mente e o conhecimento* (2009), e coordenou *Análises* (2006) *Perspectives on Rationality* (2006), *Mente, Linguagem e Acção* (2008), *Aparência e Realidade* (2010), *Aspectos do Juízo* (2011). É ainda autora de numerosos artigos, entre eles os seguintes, mais directamente relacionados com os tópicos do projecto: *Um certo desdém de Habermas face a Derrida, Pessoas e Acções – a filosofia da mente e os fundamentos cognitivos da filosofia moral*, *Problemas de Identidade Pessoal*, *Identidade Pessoal e Posição Original Rawlsiana*, *Hume e Kant*, *Razões e Paixões* e *Liberdade, cognição e acção*.

Carlos Mauro é Professor Auxiliar Convidado da Universidade Católica – Porto (Departamento de Comportamento Organizacional) e Investigador de pós-doutoramento do Instituto de Filosofia da Universidade do Porto, onde dirige a área Acção Agência e Racionalidade do MLAG e o Porto X-Phi-Lab. É autor da dissertação de doutoramento com o título *Uma Teoria Deflacionista da Racionalidade na Acção*. Trabalha actualmente no Projecto *Judgements of Moral Wrongdoings and the Emotions*, financiado pela Fundação Bial.

Susana Cadilha é Investigadora no Instituto de Filosofia da Universidade do Porto e autora de vários artigos sobre teoria da acção, ética e meta-ética. Está prestes a completar uma dissertação de doutoramento intitulada *O que é agir moralmente?*

Parte I

Filosofia da acção e racionalidade prática

Filosofia da Acção – uma introdução

Susana Cadilha e Sofia Miguens

Nesta introdução à filosofia da acção procuraremos dar conta, de forma não exaustiva mas, esperamos, suficientemente aprofundada e estimulante, das questões mais relevantes que constituem esta área da filosofia. Nesse sentido, dividimos o texto em duas secções organizadas em torno da apresentação dos dois principais problemas da filosofia da acção – saber em que consiste uma acção e como se explica uma acção –, seguidas de uma secção mais breve, que remete para problemas considerados de charneira, em estreita relação com outros domínios do conhecimento ou outros domínios da filosofia, como por exemplo a filosofia da mente ou a metafísica.

A filosofia da acção começa por ser, no seu ponto mais básico, a tentativa de problematização de uma constatação de senso-comum: a nossa condição de agentes. Constantemente vemo-nos a fazer coisas (como andar na rua, conduzir um carro, enviar um email ou votar), reportamo-nos a acções (nossas e dos outros), assumimos responsabilidade por algumas delas (somos os culpados num acidente, ou rejeitamos tal culpa), atribuímos valor a outras (por exemplo o acto heróico de salvar uma pessoa com o risco da própria vida num cenário de guerra). Descrevermo-nos como agentes é a base de qualquer tipo de interacção social: vemo-nos como actores, intervenientes, encontramos ou projectamos essa capacidade nos outros e somos por isso capazes de detectar, explicar e até de algum modo prever as acções dos que nos rodeiam – assim nos orientamos, com mais ou menos sucesso, no mundo social. Certo é que existe um ponto comum a essas observações: a ideia de que agimos, e de que há algo a que podemos chamar acção. A primeira questão que se impõe, portanto, é a de saber com que direito podemos referir-nos a algo que acontece no mundo – o mundo que os físicos, por exemplo, tratam em termos de partículas e energia, e que a ciência natural nos diz ser regido por causalidade fechada e leis independentes de nós – como tratando-se de uma ‘acção’. Como é que algo pode ser feito acontecer por agentes nesse mundo? O que é isso a que chamamos *acção*? Existirá de facto ou, nas palavras

de Thomas Nagel, quanto mais soubermos de um ponto de vista externo, científico-natural, acerca das leis que regem o mundo, maior será ‘a erosão daquilo que nós fazemos pela subtracção daquilo que acontece’?

Por outro lado, e mesmo sem fazer apelo ao conhecimento de terceira pessoa, pensando no conhecimento de primeira pessoa que temos das acções, como sabemos que agimos e que somos nós os actores daquilo que o nosso corpo faz? Será que pensar que ajo é agir? Assim como nada garante que o que eu vejo exista realmente, nada garante que o que eu faço seja eu a fazê-lo realmente – podemos imaginar um cenário análogo ao do Génio Maligno cartesiano em que há um Neurocientista Maligno que me deixa pensar que sou eu que ajo quando realmente todas as minhas acções e decisões são controladas por ele. Acontece também que eu creio que faço coisas mas aquilo que eu faço fazendo-as de certa forma escapa-me: ofereço morangos ao meu amigo, procurando ser gentil e dar-lhe prazer, e ele vai parar ao hospital com uma grave intoxicação. O que é fazer alguma coisa acontecer? O que é ser eu o actor daquilo que é feito?

1. A natureza da acção

1.1. Como se distinguem acções de não-acções?

De entre todas as coisas que vemos acontecerem há umas a que chamamos propriamente acções, distinguindo-as assim de meros acontecimentos. Se o dia clareia e amanhece, diríamos que é algo que simplesmente acontece; do mesmo modo, se uma pedra rola de uma montanha, constatamos o acontecimento; mas se alguém a chutou, tendemos a considerar que nesse caso estamos diante de uma acção. Esta é pelo menos a forma como comumente diferenciamos as situações. Mas há alguma coisa que, de facto, as distinga ou devemos antes considerar que as diferenciações que comumente estabelecemos são talvez úteis mas não têm qualquer sustentação?

Não pode ser o resultado, isto é, o evento físico que eventualmente ocorra, a distinguir a acção. Considerando o exemplo a que recorremos acima: em qualquer das situações há uma pedra que rola. Mas num caso teríamos uma acção, no outro não. Podemos ainda pensar num outro exemplo, mais complexo, em que a presença de alguém seja necessária em ambos os casos, para tentar compreender em que ponto se encontra a linha que separa acção de não-acção. Imaginemos que uma pessoa A morre atingida por uma bala. E imaginemos duas situações que possam ter provocado essa

morte: num caso alguém está a brincar com uma arma, não sabe que ela está carregada, e a arma dispara. Num outro caso, alguém é contratado para matar essa pessoa, procura-a e dispara sobre ela. Mais uma vez, o resultado é o mesmo, mas as situações parecem-nos diferentes – e parece que a diferença terá que se situar ao nível do que se passa no ‘interior’ do agente. É neste ponto que convocamos coisas como intenções ou razões.

Para tentar explicar essa diferença, vale a pena socorreremo-nos da hipótese de resposta de Donald Davidson (1917-2003), filósofo americano que é uma referência incontornável na filosofia da acção. Na terminologia associada a Davidson (e em geral ao modelo que ficou conhecido como modelo crença-desejo), mostrar que uma acção é uma acção, que em dadas circunstâncias existem acontecimentos – por exemplo, movimentos de um corpo humano – que são também acções, chama-se *racionalizar uma acção*. Racionalizar uma acção é dar a razão do agente para ter levado a cabo aquela acção, tornando assim inteligível enquanto acção aquilo que acontece. Para tal, é sempre necessário atribuir a um agente uma intenção, e para falarmos de intenção precisamos de (pelo menos) um desejo e (pelo menos) uma crença relevante. Acções são então acontecimentos intencionais – o que distingue uma acção de um mero evento é o facto de a acção poder ser descrita de um ponto de vista mentalista. Diremos, portanto, que existe uma acção se eventos no mundo forem racionalizáveis a partir de crenças e desejos de um agente: certas coisas que *acontecem* no mundo tornam-se inteligíveis como um *fazer* desse agente. A intencionalidade é assim a marca da acção – e vale a pena lembrar que nisto Davidson segue a filósofa britânica Elizabeth Anscombe (1919-2001), cuja obra *Intention* (1957) é por ele considerada seminal na filosofia da acção, ‘a mais importante desde Aristóteles’, este último consensualmente visto como o primeiro a abordar o campo de forma marcante. Para considerarmos algo que acontece como tratando-se de uma acção temos de ser capazes de atribuir intencionalidade ao indivíduo que a realizou e para isso tem que existir pelo menos uma descrição da acção sob a qual detectemos crenças e desejos relevantes que possam ter constituído a *razão primária da acção*. Facilmente se constata que um particular evento pode ter várias descrições; ora é apenas sob certas descrições desse evento que podemos pensar estar diante de uma acção intencional. No exemplo acima referenciado: se o indivíduo X pega numa arma para brincar, não sabe que ela está carregada e prime o gatilho, provocando a morte de Y, podemos considerar que sob a descrição “matar o indivíduo Y que se encontrava a

2 metros de distância”, não se tratou de uma acção intencional. Por contraste, no segundo caso (em que alguém contratado para matar Y, procura essa pessoa e dispara sobre ela), a acção de matar o indivíduo Y que se encontrava a 2 metros de distância foi intencional. No entanto, no primeiro caso estamos também diante de uma acção, porque existe pelo menos uma descrição do que aconteceu – “ X pega na arma”, “X brinca com a arma”, “X prime o gatilho” – para a qual encontramos crenças e desejos que racionalizam a acção.

Portanto, uma das hipóteses de resposta para a questão de saber o que distingue uma acção de uma não-acção é a chamada hipótese causalista – o traço distintivo é o facto de a acção ser causada de uma determinada forma, por certos estados mentais (se mantivermos o modelo crença-desejo, esses estados mentais serão crenças e desejos, precisamente). Davidson não é o único filósofo que advoga esta posição causalista; no contexto da filosofia da acção contemporânea, Alfred Mele ou Michael Bratman, por exemplo, partilham da mesma ideia. Como veremos mais adiante, o causalismo é uma hipótese que visa não apenas a questão da natureza da acção, mas também o problema da explicação da acção. Isto é facilmente compreensível: se no que respeita à distinção entre acção e não-acção se considerar que a diferença reside no facto de um dado acontecimento ser causado por estados mentais, é natural que no toca a saber como se explica uma acção se asseverar que esses estados mentais constituam a causa da acção.

Mas nem o causalismo nem o modelo crença-desejo são consensuais. É para alguns autores no mínimo estranho que aquilo que faz da acção propriamente uma acção seja um elemento exterior à própria acção, como a sua história causal. O filósofo Hugh McCann é exemplo disso mesmo. Parece que faria mais sentido se a diferença residisse na própria acção, em alguma característica distintiva que os movimentos do corpo que designamos como acções manifestassem, e que estaria ausente dos eventos meramente físicos. Seria de esperar, particularmente, que a questão de saber como são causadas as acções seria independente, e posterior, relativamente à questão de saber o que é uma acção. Até porque podemos duvidar que razões possam ser causas (argumento que discutiremos adiante), mas normalmente não temos dúvidas quanto a saber distinguir entre algo que simplesmente acontece e uma acção. Uma forma de marcar essa diferença relativamente ao causalismo é mostrar que há algo absolutamente essencial para que possamos definir um acontecimento como tratando-se de uma acção, e é nesse contexto que alguns autores identificarão acção com

volição. Isto não equivale a assumir que a intencionalidade não é a marca da acção, porque as acções continuam a ser eventos intencionais. Equivale, sim, a defender que uma acção é intencional não por ser causada pela crença e desejo apropriados, mas por tratar-se de um acto volitivo. E a volição é intrinsecamente intencional. De acordo com a tese volicionista, define-se o que é essencial para que a acção tenha lugar e fica em aberto a questão de saber se o acto volitivo é ou não causado por razões. Continua a pensar-se que a diferença reside no que se passa ao nível da vida mental do agente, mas há divergência quanto ao que se passa (opondo-se ao modelo crença-desejo) e quanto ao que é essencial para que a acção ocorra (opondo-se à tese causalista).

A proposta de H. McCann é a seguinte: o que está presente em toda e qualquer acção, e que por isso a distingue dos meros movimentos físicos, é o acto volitivo – a actividade mental através da qual a “agência” é realizada. Agir é sobretudo querer agir, isto é, querer que algo aconteça: a acção não está nos acontecimentos mas no facto de os produzir. As crenças e desejos podem ser elementos necessários para que a acção tenha lugar, mas apenas o acto volitivo é elemento suficiente, no sentido em que eu posso desejar sair para apanhar sol, considerar que a melhor forma de satisfazer esse desejo é pegar na bicicleta e dirigir-me até ao parque, mas há um espaço em branco daí até que eu efectivamente leve a cabo a acção. Esse espaço em branco é preenchido, consideram os defensores de uma teoria volitiva da acção, pelo acto volitivo. Só quando eu quero é que alguma coisa acontece – e isto mesmo que o movimento físico de pegar na bicicleta e dirigir-me ao parque acabe por não ter lugar. A acção reside no querer agir; se depois daí se segue o movimento físico ou não é algo que já não depende de nós, mas do mundo. Neste caso, dependerá de o meu sistema motor estar apto a funcionar, e também das condições climáticas se manterem, etc. Mas o movimento físico será sempre o efeito de uma acção – o querer, o acto volitivo em causa – e nunca a própria acção. Já tínhamos visto que o que distingue uma acção de uma não-acção não poderia ser o evento físico, aquilo que realmente acontece (o dirigir-me ao parque), na medida em que podemos chegar ao mesmo resultado sem que qualquer acção seja realizada (imaginem que sou sonâmbula). Mas isto poderia funcionar ao contrário, e constituir razão para considerar que estamos na verdade à procura de algo que não existe. Porque onde está realmente a acção, aquilo que *nós* fazemos? Ou, como diria Wittgenstein, “o que é o resto que fica se eu subtraír ao facto de que eu ergo o meu braço, o facto de que o meu braço se ergue?”. Para além do acontecimento ou do movimento físico, para além

daquilo que o “mundo” se encarrega de fazer, onde está o que eu faço, e o que distingue uma coisa da outra? A resposta destes autores centra-se nesse acto volitivo, “fundador”, que constitui uma acção pura, ou básica. Isto significa que, para retomar o exemplo de Wittgenstein, a acção de levantar o braço não se identifica com o movimento físico que ocorre – o braço levantar – mas tão-somente com o querer levantar o braço: e isto é precisamente o que está a mais. Não é o meu querer levantar o braço que causa a acção de levantar o braço; toda a nossa acção, tudo o que fazemos, é querer levantar o braço – o resto compete ao mundo, no caso, ao sistema neurofisiológico que faz com que o braço se levante.

Reduzir a acção a um evento mental é, claro, problemático – se tudo quanto fazemos é querer que o meu braço se levante, e não propriamente o levantamos (esse é o efeito físico que decorre da nossa acção de querer, mas que não depende de nós) então parece que somos de algum modo espectadores dos nossos próprios movimentos; são coisas que vemos acontecerem, mas que não causamos, o que é manifestamente estranho. Ao mesmo tempo, é difícil explicar como pode um puro evento mental – o acto volitivo – originar um efeito físico – o movimento (e esse é o velho problema de perceber como mente e corpo interagem, que, neste caso, coloca-se de pleno direito).

Referimo-nos, até agora, a formas de elucidar a questão de saber se há alguma coisa que separe e diferencie as acções dos meros acontecimentos. Focamo-nos em duas tentativas de resposta: 1) uma acção é um evento acerca do qual podemos apresentar uma descrição e uma explicação mentalista; 2) uma acção é um evento mental particular: um acto volitivo. Num caso, uma acção é intencional porque é possível pôr a claro qual a razão que levou o agente a agir, no outro uma acção é intencional precisamente porque é um acto volitivo do agente.

Se olharmos para este debate de um ponto de vista histórico, daremos conta que, em última análise, as duas hipóteses se tocam num ponto que Wittgenstein já tinha atacado. Tanto os causalistas como os volicionistas, para explicar o que pode contar como uma acção, separam de alguma forma os eventos mentais dos eventos físicos. Ora, toda a abordagem wittgensteiniana será anti-causalista não apenas no sentido em que o volicionismo se opõe ao causalismo mas por dissolver o problema com que estes lidam. Na verdade, Wittgenstein formulou o ‘problema do resto’ (“o que é o resto que fica se eu subtrair ao facto de que eu ergo o meu braço, o facto de que o meu braço se ergue?”) com a intenção de sublinhar que não se pode separar numa acção um elemento mental e um elemento físico e não para lançar a investigação numa

procura dos dois lados; essa busca só faria sentido no quadro de um dualismo alma/corpo, ou mente/corpo, que Wittgenstein e os wittgensteinianos simplesmente rejeitam (em última análise, para os wittgensteinianos, falar de ‘coisas separadas’ como crenças, desejos e intenções é sempre marca de dualismo, e é disso que quer causalistas quer volicionistas precisam).

1.2. *Como se distinguem as acções umas das outras? O problema da individuação.* Independentemente da questão de saber como se diferenciam acções de meros acontecimentos, um problema posterior se levanta a respeito da natureza das acções. Trata-se de um problema de cariz ontológico e coloca-se da seguinte forma: que tipo de entidades são as acções? Como devemos categorizá-las? Serão enquadráveis numa ontologia em que se fale por exemplo de substâncias, acidentes, propriedades? E como devemos individuá-las? Recuperando o exemplo anterior, imaginemos que somos capazes de detectar como tratando-se de um evento intencional a situação em que o indivíduo X é contratado para assassinar o indivíduo Y, procura-o e dispara sobre ele, premindo o gatilho. Podemos ainda querer saber o seguinte: com quantas acções nos deparamos aqui? Ou, no famoso exemplo de Anscombe em *Intention*: temos um homem que bombeia a água para encher o depósito de água de uma casa. Alguém encontrou uma forma de contaminar sistematicamente a fonte de água com um veneno cumulativo cujos efeitos são imperceptíveis até não poderem já ser combatidos. Os habitantes dessa casa estão portanto a ser envenenados. Anscombe imagina em seguida que os habitantes da casa eram os governantes de um estado e suas famílias, estando os primeiros envolvidos em planos de guerra e extermínio. O que é que o homem está a fazer? Quantas acções e intenções existem aqui? Mover o braço, accionar a bomba, encher a cisterna, envenenar os habitantes da casa, libertar o país, seriam cinco acções distintas feitas existir pelos mesmos acontecimentos e movimentos, ou uma só acção? As hipóteses de resposta a essa questão dependerão do modo como entendermos, sob o ponto de vista ontológico, o que é uma acção, e se acções são indivíduos relativamente aos quais essas questões de contagem devam colocar-se.

O que nos fará decidir, em situações como as descritas, se estamos diante apenas de uma acção, de várias acções, ou de uma mesma acção composta por vários elementos? No primeiro caso, se X mata Y disparando sobre ele, a acção de matar é distinta da de disparar? E a acção de premir o gatilho, distinta das duas anteriores? A.

Goldman avançou com uma ideia: duas situações são idênticas entre si – são a mesma acção – apenas nos casos em que o mesmo agente, o mesmo tempo e a mesma propriedade estiverem envolvidos. No caso em apreço, encontrar-nos-íamos diante, portanto, de acções diferentes, pois ainda que o mesmo agente e o mesmo tempo estivessem envolvidos, a propriedade de matar Y é distinta da de disparar a arma e distinta, por sua vez, da propriedade de premir o gatilho. Trata-se de três acções diferentes pois são três as propriedades a ser instanciadas. Obviamente, isto depende de uma ontologia de propriedades e instanciação ou exemplificação dessas propriedades por actos de agentes num dado instante de tempo. Ora, podemos perfeitamente considerar que uma tal ontologia apela a abstracções não desejáveis (propriedades existindo, e que depois são instanciadas). Donald Davidson, filósofo a que já nos referimos, apresentará uma solução distinta. Em parte devida à sua análise da forma lógica das expressões linguísticas de acções (por exemplo, 'Brutus matou César'), Davidson é conduzido a propor uma ontologia de *eventos*. Nesta, ao contrário do que assumia Goldman, uma acção específica não representa a instanciação de uma determinada propriedade; é, isso sim, um evento, um particular espaciotemporalmente localizado e irrepitível. De acordo com Davidson, apenas uma ontologia de eventos nos permitirá fazer sentido, do ponto de vista da forma lógica, de exemplos como o exemplo acima. Fica-se com o seguinte: aos eventos podem ser atribuídas diferentes descrições. Voltando ao exemplo acima referido, encontrar-nos-íamos, portanto, de acordo com Davidson, perante uma mesma acção, que pode assumir várias descrições (X mata Y, X dispara, X prime o gatilho). Goldman, porque acredita na existência de entidades como propriedades, considera que se trata de acções diferentes; Davidson não reconhece no mundo propriedades, daí que o evento seja um só. De acordo com Davidson, a diferença situa-se não ao nível ontológico mas ao nível das descrições possíveis, isto é, ao nível do discurso.

A questão da individuação da acção é tendencialmente neutra relativamente aos outros problemas que a filosofia da acção coloca, e que têm vindo a merecer mais atenção por parte dos seus teóricos, nomeadamente, o problema de distinguir acções de acontecimentos e o de perceber como se explicam as acções. Por exemplo, quanto a este último ponto, tanto A. Goldman como D. Davidson são adeptos de uma teoria causal da acção. Também quanto à natureza da acção, facilmente se compreende que as questões são independentes, uma vez que assumir uma posição no que respeita à ontologia – por exemplo, defender, como Davidson, que tanto acções como

acontecimentos são eventos particulares – deixa ainda em aberto o problema de saber como os distinguir.

2. A explicação da acção

Consideremos então, agora, outro dos problemas fundamentais da filosofia da acção. Regressando novamente ao vocabulário do senso-comum, todos reconhecemos as situações em que procuramos explicar uma acção recorrendo a estados psicológicos como crenças e desejos, razões ou intenções, como no exemplo seguinte: João lê o livro *Essays on Actions and Events* porque quer escrever um artigo sobre filosofia da acção para uma revista. Parece claro que há algo que permite perceber a acção do João; particularmente, parece óbvio que apenas certos estados mentais são passíveis de tornar compreensível essa particular acção. Em termos simples: João podia não estar interessado em escrever artigo algum ou então podia querer escrever o artigo mas considerar que o livro em causa não seria indispensável para realizar essa tarefa – o ponto é que, em qualquer dos casos, João não estaria a ler o livro em questão, o que mostra que essa acção particular só se entende à luz das crenças e desejos apropriados. Parece que há um modo de compreender a acção específica do João, que passa por apresentar a razão ou o par crença-desejo que a explica, de tal forma que nos é, de facto, permitido afirmar, sem medo de errar, que o João lê o livro do Davidson *porque* quer escrever um artigo sobre filosofia da acção para uma revista. Mas as dificuldades começam aqui. Em particular, o que pode querer dizer esse “porque”?

2. 1. Podem razões ser causas?

Aqui recuperamos a proposta causalista a que já tínhamos feito menção na primeira secção: no chamado modelo crença-desejo, a ideia de que as acções são aqueles eventos passíveis de se tornarem compreensíveis através da racionalização é acompanhada pela ideia de que o par formado pelo desejo e pela crença relevantes para a racionalização da acção também constitui a *causa* dessa acção. Isto significa que as razões (recordemos: razões são conjuntos crenças-desejos de agentes) não apenas tornam as acções compreensíveis mas explicam causalmente as acções. O argumento principal é o seguinte: há imensas razões para levarmos a cabo uma acção; imensas razões capazes de tornar inteligível uma acção (por exemplo, o João pode estar a ler o livro do Davidson porque quer impressionar alguém), mas só conseguimos explicar propriamente uma acção quando encontramos a razão que é a

sua causa, a real razão pela qual o agente fez o que fez. Apontar razões possíveis não é o bastante para explicar uma acção, e isso só se compreende porque há uma razão que tem, de facto, relevância causal (se, pelo contrário, as razões não causassem, qualquer uma serviria).

Esta proposta está de acordo, como vimos, com as explicações de senso-comum, que quotidianamente avançamos para dar conta do que fazemos – e está também de acordo, diga-se, com uma longa tradição que vem desde Aristóteles. No entanto, durante muito tempo ela foi posta em causa: Wittgenstein e os seus seguidores – com E. Anscombe à cabeça – negavam o carácter causal das explicações psicológicas. A linha de abordagem dos wittgensteinianos é a seguinte: questões ‘Porquê...?’ quanto a um dado movimento corporal tanto podem ser respondidas apontando razões como apontando causas; ora, é vital para a teoria da acção evitar ‘confusões gramaticais’, que nos fariam falar de razões como causas. Os wittgensteinianos consideram uma falácia naturalista falar de razões como causas de acções. Porquê? Porque razões e causas seriam conceitos de ordem diferente e dizer que uma razão causa uma acção seria um erro conceptual, uma transgressão dessas ordens. Por exemplo, uma causa é sempre susceptível de investigação empírica, susceptível de ser descoberta e portanto de ser ignorada por mim enquanto ‘agente’; ora, para os wittgensteinianos, ao contrário do que acontece com causas, não tem sentido dizer que eu descobro as razões, que anteriormente ignorava, da minha acção. Outra pessoa pode vir a dizer-me quais foram as causas dos movimentos do meu corpo (levantei o meu braço) mas não quais foram as razões para o mover. Pode-se chamar ao argumento dos wittgensteinianos o argumento da conexão lógica. Esse argumento é o seguinte: se as razões fossem causas de acções, então a causa suposta (a razão) deveria ser separável do seu efeito (a acção ou comportamento) que seria extrínseco e contingente. Ora, não é possível descrever a razão de uma acção independentemente da própria acção e do seu resultado (a intenção de fazer A menciona A). Logo existe uma relação intrínseca e necessária entre a razão e a acção, que não tem as características de uma relação causal.

Neste ponto, devemos novamente chamar a atenção para a importância do artigo “Actions, Reasons and Causes” (1963), de D. Davidson, onde ele ataca esse argumento. A importância histórica desse artigo foi precisamente ter restituído legitimidade ao causalismo ao mesmo tempo que delineava uma posição suficientemente sofisticada (nomeadamente pelo papel que atribui à racionalidade na

teoria do lugar da mente no mundo) para responder às críticas de origem wittgensteiniana. A resposta de Davidson é peremptória: a relação causal é uma relação que se estabelece entre eventos – independentemente do modo como os descrevemos. Onde há causalidade tem de haver lei: os acontecimentos que têm uma relação de causa e efeito entre si são subsumidos por leis deterministas estritas (Davidson chama a isto o princípio do carácter nomológico da causalidade). Na explicação psicológica (“João lê o livro *Essays on Actions and Events* porque quer escrever um artigo sobre filosofia da acção”), aquilo que se passa é que descrevemos sucintamente a situação em termos mentalistas, associando uma razão a uma acção. Mas do facto de podermos fazer isso, e de existir uma relação de implicação lógica entre razões e acções, não se segue que não exista também uma relação causal entre os eventos físicos particulares que ocorreram.

Daqui se depreende que Davidson, ao defender uma teoria causal da acção, não pode deixar de fazer afirmações importantes acerca do lugar da mente no mundo. Particularmente, o que ele acaba a defender é um certo tipo de teoria ontológica do lugar da mente no mundo, que ficou conhecida como monismo anómalo. A teoria é monista porque se considera que o mundo é de uma só natureza (fundamentalmente física) e enquanto tal regido por leis estritas, mas esse monismo é anómalo porque algumas coisas (as coisas mentais, como acções ou percepções) não caem sob (não são capturadas ou capturáveis) as leis estritas que regem o mundo físico. Isto não significa que estejam fora do mundo ou sejam imateriais mas que as descrições que as capturam são de uma outra ordem que não nomológica. Todas as ‘entidades’ mentais supõem a racionalidade no sentido em que apenas descrições mentalistas as capturam, mas estados mentais como crenças, desejos ou razões são acontecimentos físicos particulares. Se preferirmos, Davidson está a dizer que acontecimentos físicos particulares podem ser descritos de um ponto de vista mental. É isso que fazemos quando invocamos explicações psicológicas do tipo atrás referido, e é disso que precisamos quando em teoria da acção queremos saber como explicar uma acção. Isto significa ao mesmo tempo, evidentemente, que ‘explicar uma acção’ não é o mesmo que subsumir a uma lei, sendo essa a forma de explicação característica das descrições do que acontece enquanto físico. Explicar uma acção é precisamente dar razões. Mas o facto de invocarmos razões e de estabelecermos relações lógicas necessárias entre razões e acções (o facto de sermos capazes de descrever os acontecimentos dessa forma) não significa que as razões não possam simultaneamente

ser as causas dessas acções: elas são acontecimentos físicos, e é nessa qualidade que causam as acções. Portanto, a proposta que Davidson avança é a seguinte: a ligação entre razão e acção é uma conexão que se obtém entre dois eventos (por um lado, o agente acreditar e desejar e, por outro, o agente agir) e que pode ser descrita de maneiras diferentes. Essa conexão é (i) *racional*, tanto quanto o par crença-desejo especifica a razão para a acção, (ii) *causal*, tanto quanto um evento causa o outro.

Claro que a questão não fica resolvida por aqui. É lícito questionarmos se faz de facto sentido falarmos de explicação quando apresentamos as razões de uma acção. Qual é de facto o poder explicativo das racionalizações? De acordo com a acepção própria, central, de explicação, esta tem apenas lugar na circunstância de podermos reconhecer a instanciação de uma lei, assim se distinguindo explicação de justificação. Apesar de reconhecer que de facto não existem leis psicofísicas (isto é, leis que relacionem estados mentais como crenças e desejos a estados físicos como movimentos de um corpo), Davidson não considera isto um problema pelas razões que já apontámos. Mesmo que as únicas leis estritas (isto é, sem excepção) sejam as leis físicas, os eventos descritos como mentais são também eventos físicos, e é nessa qualidade que se encontram submetidos a leis causais. Só que isto levanta um novo problema: será que desta forma fica ainda assegurada a causalidade mental, ou de alguma forma a intervenção da mente no mundo? Se é enquanto eventos físicos que as razões causam as acções, que espaço resta ainda propriamente para a explicação psicológica? Será que de facto o mental *causa*?

O impasse é o seguinte. Parece que para falarmos de acções precisamos de falar de razões. É essa, como vimos, a sua proposta no que toca à questão de saber o que distingue uma acção de um qualquer evento que ocorre no mundo (problema da natureza da acção). Se descrevêssemos o acto de beber um copo de água, por exemplo, em termos puramente físicos, estaríamos a falar de movimento físicos e eventos cerebrais, mas nunca de acções. Só podemos considerar estar diante de acções quando somos capazes de reconhecer as razões que as explicam. No entanto, ao lidar com o problema da explicação da acção, damos agora conta que essa explicação, que implica uma relação causal, estabelece-se não entre eventos mentais e eventos físicos mas apenas entre eventos físicos. Portanto, definiu-se acção pela sua propriedade de ser causada/explicada por estados mentais, mas agora vemos que não é por serem estados mentais que as razões causam/explicam as acções. Daí as acusações de epifenomenismo dirigidas a Davidson – as propriedades mentais seriam afinal

causalmente inoperantes e por isso irrelevantes. Outro problema ainda: será que podemos continuar a falar, com correcção, de ‘agentes’?

2.2. *O caso das cadeias causais desviantes*

A questão adensa-se ainda sob um outro ponto de vista. Está por esta altura assente que de acordo com o modelo crença-desejo que Davidson, por exemplo, preconiza, a acção intencional define-se pelo facto de ser causada por um par crença-desejo que a torna compreensível. Mas é o próprio Davidson que lembra situações em que essa relação entre os estados mentais e a acção se complexifica. Concretizando, Davidson admite a possibilidade de alguém ter uma razão para realizar uma acção, levar a cabo a acção, e essa não ter sido a razão pela qual ele fez o que fez. E apresenta um exemplo já célebre: imaginemos o caso de um montanhista que segue à frente de um outro, sentindo-se incomodado pelo peso, pelo cansaço que essa situação lhe provoca e pelo perigo que ela representa. Ele quer livrar-se do peso e sabe que largando a corda conseguiria isso. Pensar em tal coisa pode enervá-lo a tal ponto que acaba por soltar a corda e o segundo montanhista cai. O que ele pretende mostrar é podem existir diferentes crenças e desejos (ou diferentes razões) passíveis de racionalizar ou tornar inteligível uma dada acção, mas encontramos a razão que explica a acção apenas na circunstância de ela ter causado a acção de forma apropriada, ou seja, no caso de ter sido essa realmente a razão pela qual a acção é realizada. Neste exemplo, esse não foi o caso. Vemos algo a acontecer (um montanhista que solta uma corda, provocando a queda de um outro), existe uma razão (pelo menos um par crença-desejo) que racionalizaria isso que vimos acontecer, e que acabou por causar o dito comportamento, mas não podemos considerar que tal comportamento constitua uma acção intencional, pois parece claramente que algo escapou ao seu controle, que não foi algo que ele quis fazer.

Portanto, se todos os casos de acção intencional se definem do seguinte modo: X agiu intencionalmente se e só se agiu por uma razão que é a causa dessa acção, o contrário não se verifica – isto é, nem todos os casos em que temos uma razão a causar um comportamento, temos acção intencional. O que nos mostram casos como este – a que se convencionou chamar situações de cadeias causais desviantes – é que nem sempre as razões causam da forma apropriada.

Exemplos como este colocam portanto um problema às teorias causais, pois parece que há situações em que o agente tem o desejo de fazer X e a crença correspondente

de acordo com a qual pode obter X se fizer Y; esse par crença-desejo acaba por causar Y, e ainda assim não estamos perante uma acção intencional. Ou seja, aquilo que de acordo com a teoria causal (como vimos na primeira secção) permite distinguir um mero acontecimento de uma acção intencional – a propriedade de ser causada por estados mentais como crenças e desejos – não é afinal condição suficiente para fazer essa distinção.

Algo que à partida parecia simples e funcional – fazer da atribuição de estados mentais como crenças e desejos a marca distintiva da acção – parece afinal não ser a solução, pois podemos ter os estados mentais causais e não ter a acção correspondente. O exemplo das cadeias causais desviantes coloca, assim, em cheque, o causalismo: a solução que permitia distinguir uma acção de uma não-acção e, conseqüentemente, a ideia de que as razões podem explicar as acções (na verdade, algo mais importante será colocado em causa: a própria possibilidade de explicar acções, como veremos adiante).

A resposta por parte dos causalistas tem passado por apontar as condições específicas em que uma razão é, de facto, a causa da acção: uma razão explica a acção apenas na circunstância de ela ter causado a acção de forma apropriada – e a expressão a reter aqui é “de forma apropriada”. Não basta que esteja presente uma razão que poderia racionalizar a acção; não basta que exista uma coincidência, mas é preciso que exista uma correspondência entre a acção e a razão de tal modo que seja esta última que, de alguma forma, guie a acção. Agir por uma razão é agir em resposta a essa razão específica – tem que existir uma conexão directa entre ambas e a percepção, da nossa parte, que estamos a agir por essa razão. Outro famoso exemplo de Davidson usado para ilustrar essa distinção é o seguinte: suponhamos que alguém quer matar outra pessoa atirando sobre ela, mas falha. No entanto, o disparo assusta um grupo de porcos selvagens que se encontrava por perto e estes, em fúria, acabam por esmagar a pessoa que tinha escapado ao disparo. O que este tipo de exemplos mostra é que para que possamos com propriedade designar algo que acontece como uma acção, é exigível que ela represente a realização de algo como uma representação mental do agente, caso contrário não estaremos diante de uma acção intencional propriamente dita, mas apenas de um efeito não-intencional.

Mas estas adendas à solução inicial soam por vezes pouco esclarecedoras, algo

tautológicas: parece somente a reafirmação da ideia de que a razão é a causa de uma acção intencional apenas nos casos em que...a razão é propriamente a causa de uma acção intencional. Por esse motivo, alguns autores preferem fazer entrar em cena uma outra noção – a noção de intenção – pondo em causa o modelo crença-desejo, mas salvaguardando a teoria causal da explicação da acção.

2.3. *Uma outra equação: crença + desejo + intenção = acção*

Fazer entrar a noção de intenção implica considerar que se trata de um elemento novo, i.e, que uma intenção não se reduz a uma combinação de uma crença e um desejo relevantes. Desde “Actions, Reasons and Causes” que Davidson se refere a intenções: “conhecer a razão primária pela qual alguém agiu da forma que agiu é conhecer a intenção com a qual a acção foi realizada”. Mas, nesse artigo, para atribuir a um agente uma intenção precisávamos de um desejo e de uma crença relevantes. Ou seja, nesse artigo, Davidson comprometia-se com o modelo crença-desejo e com a ideia de que a intenção se identifica com o par crença-desejo. Neste ponto, avança-se, porém, com uma solução diferente, que passa por considerar que a acção intencional envolve um outro estado mental distinto, que acrescenta algo às crenças e desejos: a intenção. A nossa vida mental não é apenas constituída por estados mentais “actuais” e básicos como crenças e desejos. Somos capazes de vários tipos de actividade mental, como a de planear acções para o futuro, próximo ou distante, que exigem um tipo de estado mental mais complexo, e que exige um maior comprometimento da nossa parte. Neste sentido: crenças e desejos são coisas pelas quais somos, não raras vezes, assaltados; são coisas sobre as quais, também não raras vezes, não chegamos a ter controlo. Podemos até ter desejos contraditórios. Formar uma intenção implica dar o passo seguinte, predispor-nos a agir. Não é apenas algo que temos, mas é algo que de algum modo construímos. E não conseguimos formar intenções contraditórias – posso ter o desejo de ir à praia esta tarde e o desejo concomitante de acabar o artigo esta tarde, mas não formo a intenção de ir à praia juntamente com a intenção de acabar o artigo. Claro que ter a “intenção de” não garante que de facto a acção se realize, mas é um elemento intermediário essencial para que a acção tenha lugar. Regressando ao exemplo do montanhista: não basta, como vimos, que eu tenha o desejo de livrar-me do peso e a crença de que posso fazê-lo se largar a corda. Para que se trate de uma acção intencional, de acordo com esta solução, eu teria além disso de ter formado a

intenção ou delineado o plano de largar a corda, por forma a livrar-me do peso. E é isso que marca a diferença (entre comportamento e acção intencional).

Trata-se, portanto, de uma proposta alternativa de explicação da acção, que faz intervir a intenção e concede-lhe o papel de guiar e coordenar as nossas acções. Mas é ainda uma explicação causal, que trata a intenção como a causa próxima da acção. É claro que continuamos a formar as intenções a partir de crenças e desejos, mas a intenção invoca um compromisso de ordem psicológica que vai mais longe do que a simples presença de crenças e desejos, e que é portanto essencial para que uma acção intencional tenha lugar.

2.4. *Podemos de facto explicar as acções? A alternativa não-causal*

Até aqui temos apresentado diferentes tentativas de explicar a acção. Mas há ainda uma outra hipótese a considerar: a ideia de que não é, de todo, possível explicar a acção, porquanto o único sentido próprio de explicação – a subsunção a leis, o estabelecimento de relações causais estritas – não é aplicável quando falamos de acções e da sua relação a itens mentais como crenças, desejos ou intenções. A dúvida aqui não é saber qual a cadeia causal apropriada, mas simplesmente a assunção de que a relação de causalidade, uma relação contingente que se descobre entre factos empíricos, própria do domínio natural, não se estabelece entre razões e acções. Recupera-se, claro, o argumento de Wittgenstein que focámos atrás: “A proposição de que a tua acção tem tal ou tal causa é uma hipótese. A hipótese é bem fundada caso tenhamos tido um certo número de experiências que, grosso modo, concordam em mostrar que a tua acção é o que normalmente ocorre no seguimento de certas condições, às quais passamos então a chamar causas da acção. [Mas] Para sabermos a razão que tiveste para fazer uma certa afirmação, para agir de uma certa maneira, etc., não é necessário um certo número de experiências concordantes, [por isso] a afirmação da tua razão não é uma hipótese.” Para os não-causalistas é simplesmente absurdo que se possa dizer que a minha acção é o que “normalmente ocorre” no seguimento de certas condições, simplesmente porque não podemos afirmar que alguma coisa se siga de certos estados mentais. Neste ponto podemos também recuperar Anscombe, discípula de Wittgenstein, que tinha como um dos seus objectivos, ao escrever *Intention*, afastar precisamente aquilo que considerava serem

as más interpretações estabelecidas da caracterização do raciocínio prático feita por Aristóteles – ou seja, a ideia de que o conhecimento prático deve ser visto sob o prisma do conhecimento teórico, o que conduziria nomeadamente a considerar o silogismo prático como prova de uma conclusão. Pense-se no exemplo: Alimentos secos são apropriados para qualquer humano / O alimento x é um alimento seco / Eu sou humano / Este é um pedaço do alimento x / Este alimento é apropriado para mim. Temos vontade de responder: sim, e depois? Não somos necessariamente movidos a acção alguma.

O ponto é que mesmo que sejamos capazes de considerar, numa dada ocasião, que João fez X porque queria Y, o que não podemos é afirmar que sempre que João quer Y, ele faz invariavelmente X. Os não-causalistas não duvidam que o que se passa possa ser descrito em termos de crenças, desejos e intenções, duvidam sim da natureza mental separada de entidades tais e consequentemente do papel causal que elas possam desempenhar enquanto mentais, bem como da possibilidade de generalizar sobre tais bases. Isto é, duvidam que a relação que se estabelece entre intenções, razões e acções seja do tipo explicativo; intenções e razões servem para tornar as acções compreensíveis e não para as explicar. Estas abordagens apelam normalmente para o carácter normativo das relações entre estados mentais e acções, por oposição ao carácter explicativo. E, portanto, para a ideia de que a acção intencional plena só é possível e só se compreende à luz de uma rede de crenças e desejos coerente e de um conjunto de normas que regulam a nossa vida mental. É por relação a esse fundo que as acções intencionais se entendem e, em certo sentido, se justificam. A acção intencional é assim uma actividade plenamente consciente e reflexiva – agir não é qualquer coisa que me aconteça, como o resultado de uma medição e comparação de crenças e desejos (como no modelo causal), mas é algo que eu faço, num quadro de actividade reflexiva. Portanto, não é possível estabelecer entre os estados mentais e a acção uma relação de causalidade simples do tipo “se A deseja X mais do que Y, sabe como obter X, e é livre de obter X, então ele intencionalmente fará X”, porque muitas coisas se intrometem – particularmente, a nossa capacidade de reflectir sobre os nossos desejos. Mesmo que eu deseje X e me considere livre para o obter, posso ainda não o fazer por considerar, por exemplo, que não devo fazê-lo. O que é importante reter é que nesse quadro, onde intervêm noções normativas, não é possível falar de leis nem de explicação, num sentido estrito, de acções. Mas é por

isso que podemos, com propriedade, falar de agentes livres – é nesse hiato reflexivo que se estabelece entre os estados psicológicos e a acção que se encontra o agente. Porque a acção intencional tem por base uma capacidade de reflexão e de comprometimento que não é previsível nem quantificável e porque o agente pode intervir sobre os seus desejos e crenças, modificando-os, não é possível prever uma acção tendo por base as crenças e os desejos do agente, nem tão pouco o inverso – explicar a acção por relação a eles. As razões incitam-nos a agir enquanto razões, não enquanto causas.

É lícito perguntarmo-nos, por esta altura, se a este problema não foi já dada resposta cabal pela solução engenhosa de Davidson. Afinal, Davidson reconheceu a ‘anomalia’ do domínio do mental, i.e., reconheceu que toda a esfera do mental é anómala relativamente ao mundo regido por leis que é o mundo físico. Deixou claro, de igual modo, que ‘explicar’ no âmbito das ciências físicas é diferente do sentido de ‘explicar’ a acção humana – a explicação de uma acção envolve a racionalidade, o que queremos não é descrever eventos físicos sob leis mas sim dar razões, dar as razões que conduziram à acção.

Essa não é, porém, considerada uma solução satisfatória por aqueles que estão preocupados em defender a liberdade do agente, na medida em que a relação normativa e a relação causal estão apartadas: a relação normativa estabelece-se entre conteúdos proposicionais ao passo que a relação causal existe somente entre eventos. Isto significa que não é em virtude das suas razões que o agente intervém no mundo (no fundo, a acusação de epifenomenismo a que já fizemos menção). Em Davidson, ao contrário da perspectiva que acabamos de apresentar, as razões incitam a agir enquanto causas e são causas enquanto eventos físicos.

Terminamos assim o mapeamento de posições perante os dois problemas centrais da filosofia da acção – a natureza da acção e a explicação da acção. No que se segue procuramos exemplificar algumas relações da filosofia da acção com outras disciplinas da filosofia, como a filosofia da mente, a metafísica ou a filosofia moral.

2. 3. Filosofia da acção e possíveis ligações externas: da filosofia da mente à metafísica

3.1. Acção e pessoalidade

A acção parece-nos frequentemente central, e por boas razões, para pensarmos no tipo de seres que somos, seres capazes de ‘vontade’ e ‘liberdade’, capazes de ‘fazer ser’ e não apenas de observar coisas a acontecer, mesmo que pelos movimentos dos nossos corpos. Alguém que procurou explicitar a relação entre a acção e o tipo de ser que somos, focado naquilo a que chamou ‘pessoalidade’ (*personhood*), que nos caracteriza, foi o filósofo americano da mente Daniel Dennett. Basicamente, ele defende que a eficácia das razões no controlo da acção de um sistema cognitivo físico é fundamental para a existência de pessoalidade no mundo.

Segundo Dennett, ‘pessoas’ e ‘acções’ são conceitos mentalistas, razão pela qual não podemos ter concepções absolutas de pessoa ou de acção. Ele prefere ver 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). Devemos considerar que uma entidade no mundo se estabelece enquanto pessoa através de uma relação constitutiva com acontecimentos que têm determinadas características e que são as suas acções. Este ponto pode ser formulado evocando a noção lockeana de pessoa. 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»). O que a teoria dennettiana das condições de pessoalidade (*conditions of personhood*) faz é relacionar intimamente os dois componentes da noção lockeana, aquele que evoca a auto-consciência e aquele que evoca a acção própria e a nossa relação com esta.

Se pessoa não é um conceito absoluto, aplicável a qualquer indivíduo da espécie humana (mesmo que legalmente venhamos a estipular que as coisas devem ser como se isso fosse assim), o que acontece é que para uma entidade ser pessoa é necessário que determinadas condições sejam preenchidas por um particular sistema físico: esse sistema deve ser 1. racional (no sentido instrumental), 2. susceptível de descrições intencionais, 3. capaz ele próprio de considerar intencionalmente outros sistemas, 4.

capaz de comunicação verbal (com o aprofundamento mental inerente, revelado pelas análises griceanas – ter crenças acerca de crenças acerca de crenças). Isto é parte daquilo de que a auto-consciência essencial ao conceito de pessoa depende. Mas até aqui nada explica porque é que podemos considerar por vezes pessoas individuais como responsáveis pelas suas acções. A chave para a ‘pessoalização da pessoa’ (o que é preciso para uma pessoa ser especificamente esta ou aquela pessoa) é outra e faz entrar em cena a acção. É necessária a auto-avaliação, e, por exemplo, um comunicador griceano (que tem crenças acerca de crenças acerca de crenças) pode não ser capaz desta auto-avaliação. Para compreender a última condição de pessoalidade, Dennett apela às ideias de H. Frankfurt e Ch. Taylor, respectivamente ‘volições de segunda ordem’ e ‘avaliação forte’ (avaliação de si e dos desejos e crenças próprias e não apenas de acções).

Harry Frankfurt 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 implica ser-se livre para se querer o que se quer querer. Assim, a liberdade da vontade só pode existir em seres capazes de formar volições de segunda ordem, seres capazes de quererem que um determinado desejo seu seja (ou não seja) a sua vontade (no exemplo do viciado não existiria essa liberdade de querer ou não querer). Um comportamento voluntário de tendência para o objecto do desejo apenas constitui uma acção livre se envolver volições de segunda ordem. 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 última condição envolve avaliação forte. O conceito é de Ch. Taylor, que o reporta à responsabilidade pela identidade própria. 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. De qualquer forma, e é isso que nos interessa aqui, em *Conditions of Personhood* Dennett defende que sem acções não pode haver pessoalidade no mundo e apenas cumpridas estas condições a ‘ordem que

está lá’, a aparência de propósito, os *free floating rationales* observáveis no comportamento de um sistema, se tornam razões propriamente ditas.

O que ficou dito acima é importante por exemplo para os debates sobre a consciência na filosofia da mente: significa, por exemplo, que para Dennett a auto-consciência envolve mais do que auto-apercebimento e auto-representação: envolve acção e uma certa relação com a acção própria; a consciência nunca poderia, nestas circunstâncias, ser, por hipótese, identificada com qualia.

Esta forma ‘não absolutista’ de pensar na pessoalidade, ligando pessoa e acção, tem óbvias implicações éticas e políticas.

3.2. Racionalidade

Um tema quase sempre ligado à acção é o tema da racionalidade. Convém ter em mente, no entanto, quais os sentidos de racionalidade disponíveis. Antes sequer de nos reportarmos a agentes humanos, que mais ou menos conscientemente escolhem cursos de acção, com base em crenças e desejos ou num determinado plano, antes sequer de podermos com propriedade falar de acções e agentes intencionais, há um outro sentido, mais básico, de racionalidade que é detectável no mundo, no comportamento de sistemas mais ou menos complexos que na sua relação com o ambiente pode dizer-se que manipulam informação de forma mais ou menos adequada, assim ajustando ‘meios’ e ‘fins’ (ou podendo ser descritos como tal). A um tal ajustamento de meios e fins chama-se ‘racionalidade instrumental’ e os sistemas em causa são objecto de estudo em áreas tão diversas como a biologia evolucionista ou a economia.

Quando estamos já perante agentes humanos tendemos a qualificar como ‘racional’ o que se passa em circunstâncias deste género: temos um agente que crê determinadas coisas, tem determinados desejos, e que em função dessas crenças e desejos age de forma a obter aquilo que pretende, de forma a desencadear o estado do mundo que corresponderá à satisfação dos seus desejos (por exemplo, alguém que espera uma mensagem importante, deseja avidamente lê-la, crê que essa mensagem acabou de chegar, e dirige-se imediatamente à caixa do correio para a ler). Aquilo de que estamos a falar quando qualificamos como racional o comportamento desse agente é de uma acção apropriada a uma dada finalidade, da selecção e mobilização de meios com vista a um determinado fim (os fins do agente são relativos àquilo que ele deseja, e os agentes chegam supostamente à situação de decisão já munidos de desejos). Não

chamaríamos racional ao comportamento do agente se este, esperando uma mensagem importante, desejando avidamente lê-la, acreditando que a mensagem acabou de chegar, em vez de se dirigir imediatamente à caixa do correio fugisse desta a sete pés (convenhamos que isto pode perfeitamente acontecer – o funcionamento das crenças e desejos de agentes humanos não é nada simples).

Reencontramos assim a racionalidade instrumental – a chamada definição instrumental da racionalidade diz-nos o que é racionalidade na acção, e é na verdade a mais consensual e comum. A definição instrumental pode ainda fazer referência a processos mentais (crenças, desejos) envolvidos num processo de controle da realidade por parte de um ser inteligente. Se há alguma coisa quanto à qual as pessoas que falam de racionalidade estão de acordo é a definição instrumental de racionalidade. Mas isso é muito pouco. Antes de mais, várias coisas têm que estar já no lugar se quisermos aceitar uma tal definição de racionalidade – temos nomeadamente que saber que coisas são essas a que chamamos ‘acções’ e ‘agentes’, por contraste com coisas que simplesmente acontecem no mundo e corpos que se movem – mas temos quase um mínimo denominador comum. É verdade que não sabemos bem que estatuto tem esta definição, nomeadamente como se relaciona com aquilo a que podemos chamar cânones de racionalidade (lógica, teoria das probabilidades, teoria da decisão). A definição instrumental de racionalidade parece ser uma definição puramente exterior, comportamental – como se relacionará ela com as questões de normatividade e princípios evocadas nessas teorias? É verdade também que considerar que ‘ser racional é agir de forma apropriada a uma dada finalidade, mobilizar meios com vista a um determinado fim, o qual satisfaz os desejos do agente’ não nos diz ainda por exemplo exactamente que princípios de decisão vão reger a passagem à acção, que princípios vai o agente seguir tomando como ponto de partida as crenças e desejos que tem: agentes como nós têm muitas crenças e muitos desejos, acreditam e querem simultaneamente muitas coisas, e assim a aparente simplicidade da ideia de mobilizar meios com vista a um ‘fim que satisfaça o desejo do agente’ rapidamente se esvai. No âmbito da teoria da decisão, por exemplo, são discutidos princípios de decisão alternativos, tais como a maximização da utilidade esperada ou o princípio maximin. Dizer que os princípios de decisão são alternativos quer dizer que a partir da mesma informação e dos mesmos desejos do agente poderão surgir diferentes decisões de acção.

Qual é o problema com isto? É que por muito úteis e mesmo inabdicáveis que sejam

estas caracterizações da racionalidade elas são formais, são modelos. Como se relacionam com agentes que são reais (por exemplo, corpos físicos humanos, controlados por cérebros que não são ‘movidos’ a proposições mas de forma bioquímica) abre um conjunto muito grande de problemas, desde a crítica dos fundamentos filosóficos da teoria da decisão racional, até investigações psicológicas acerca de racionalidade e irracionalidade em que filósofos e psicólogos têm estado envolvidos nas últimas décadas. Em suma, é muito mais fácil evocar a racionalidade do que saber exactamente do que falamos quando falamos de racionalidade. Ora, tanto quanto evocações de racionalidade no comportamento e na acção concebidas de forma razoavelmente simples são comuns nos mais variados contextos teóricos (desde a biologia evolucionista à filosofia política – pense-se em Rawls falando de escolha racional dos princípios da justiça na posição original), se há problemas acerca da natureza da racionalidade, estes serão importados/herdados.

Por esta altura será claro que este sentido de racionalidade que apelidámos como básico, transversal ou mínimo excede em muito a própria definição da acção racional, a que os filósofos da acção dedicam mais atenção, assim como a própria noção de raciocínio prático. Por exemplo, para Davidson, a racionalidade é uma questão de coerência entre os estados mentais e a acção e uma característica que por defeito atribuímos aos agentes, como condição de possibilidade para interpretar o seu pensamento ou a sua acção. Isto é, parte-se do princípio de que os agentes humanos são em larga medida racionais no sentido em que as suas crenças, desejos e intenções formam um todo coerente e no sentido em que a sua acção se compreende à luz dessas crenças e desejos. No contexto da filosofia da acção, portanto, para avaliar da racionalidade da acção, o cenário tem já que incluir coisas como agentes humanos que deliberam e escolhem caminhos alternativos de acção com base nas suas crenças e desejos, e essa avaliação tem por base um entendimento holista do que é a vida mental, muito além de uma mera mobilização de meios para atingir um fim.

Neste contexto, um exemplo típico de acção irracional é a acção acrática: a ocasião em que o agente vai contra o seu melhor juízo, i.e., age contrariamente às razões que ele próprio considera serem as mais relevantes na situação em causa. Por aqui se vê como o conceito de acção racional é mais estreito do que o de acção intencional que analisámos na primeira parte – para qualificar uma acção como racional é necessário não apenas reconduzir a acção às razões do agente, mas verificar se entre as razões e a acção se estabelece uma ligação adequada. No caso da acção acrática, a acção decorre

das razões do agente, e por isso trata-se de uma acção intencional. No entanto, o agente reconhece que, considerando todos os aspectos em questão, há um melhor curso de acção disponível, e portanto age contra o seu próprio juízo – daí que se avalie tal acção como irracional, devido à incoerência por relação com as crenças e desejos relevantes do próprio agente (trata-se de um caso de irracionalidade subjectiva).

A questão da racionalidade pode adquirir vários sentidos e conotações, e interessou-nos por isso chamar a atenção para o problema de saber de que falamos quando aplicamos esse termo. Desde esse sentido mínimo que podemos detectar em múltiplos comportamentos de agentes humanos e não-humanos, até um sentido mais “robusto” de acção racional que encontramos nas teorizações próximas da filosofia moral, estão em causa diferentes considerações do que é ‘ser racional’. Por exemplo, a definição puramente instrumental da racionalidade acima avançada é considerada insuficiente por muitos filósofos morais. Para os filósofos que querem fazer da moral uma expressão da racionalidade, uma acção moral será racional não porque representa a selecção de meios adequados para alcançar um determinado fim, mas porque, podendo deliberar não apenas a respeito dos meios mas também dos fins, somos capazes de seleccionar cursos de acção válidos de um ponto de vista imparcial. Deste ponto de vista, a acção racional caracteriza-se tanto pela deliberação acerca dos meios para atingir um dado fim, como pela deliberação acerca do que queremos (o que está de algum modo próximo da ideia de Frankfurt acerca da liberdade da vontade) e do que devemos querer. O que isto mostra é que em caso de conflito moral, por exemplo, nunca poderíamos calcular qual o melhor curso de acção a partir de um princípio da teoria da decisão como o princípio da maximização da utilidade esperada – ou seja, avaliar a racionalidade da acção pode ir muito além do mero cálculo instrumental. Há casos em que, ao deliberar, não nos limitamos a calcular desejos mas intervimos sobre eles e avaliamo-los de acordo com padrões que não podem ser medidos em termos de desejos. Este é todo um outro sentido de racional, ligado a uma capacidade reflexiva e consciente partilhada por agentes humanos de pensarem no que deve ser o caso, e não já à mera capacidade de adequar meios a fins. Para alguns filósofos, a moral só é possível a partir daí, se estivermos já perante um ser consciente e reflexivo, capaz de linguagem.

Este é outro dos tópicos, portanto, pelos quais a questão da acção se liga à filosofia moral e política. De salientar, no entanto, a tentativa que a filosofia da acção tem

vindo a empreender no sentido de se livrar de qualquer conotação moral na análise dos seus problemas e portanto no sentido de se afastar de possíveis implicações morais dessas análises. Se o tem conseguido ou não, é ainda um problema adicional.

Pessoalidade e racionalidade são, então, apenas dois dos tópicos nos quais as questões relativas à acção confluem, existem vários outros que não teremos oportunidade de explorar aqui.

3.3. Livre-arbítrio e responsabilidade moral

Uma outra ligação possível é a que se estabelece entre a filosofia da acção e as questões do livre-arbítrio ou da possibilidade da responsabilidade moral. O problema do livre-arbítrio, *grosso modo*, consiste em saber se agentes humanos podem ser considerados como detentores de uma vontade livre. Há mais do que uma forma de pensar em que pode consistir essa liberdade, mas o modo mais comum de colocar o problema define o livre-arbítrio como a capacidade de poder escolher a cada momento, de entre um leque de alternativas possíveis, que curso de acção tomar. O problema é conciliar isso com a ideia determinista central de acordo com a qual tudo o que acontece tem uma causa. Se isto vale também para a acção humana, então a escolha que fazemos é afinal determinada por um estado anterior do mundo (e pelas leis da natureza), e assim sucessivamente, até chegarmos a um estado de coisas sobre o qual não temos qualquer controle ou responsabilidade. Portanto, dado o determinismo, não é verdade que eu, a cada momento, possa agir ou escolher de modo diferente daquele que acabo por fazer – que tenha controlo sobre o meu próprio querer. Em risco está, então, também a ideia de que somos responsáveis pelas nossas acções.

O problema do livre-arbítrio tem alguma importância no contexto das discussões em filosofia da acção uma vez que é em atenção a esse problema que alguns teóricos optam por rejeitar a alternativa causal da explicação da acção. Precisamente porque a existir essa relação causal, nómica, entre razões e acções, então não poderíamos ter acções livres. Uma teoria causal da explicação da acção faria depender a minha decisão de estados mentais cuja proveniência, em última instância, não depende de mim. A acção seria como que o resultado de um embate entre as razões que eu por

acaso tenho, e não propriamente algo que eu faça ou origine, como a “fenomenologia da acção” parece indicar. A existir essa relação causal, então, dadas as nossas crenças e desejos, parece que nada nos resta senão a prossecução de um certo caminho. Em causa está portanto a ideia de que somos autores das nossas acções, de que somos agentes no pleno sentido do termo.

Porque lhes interessa manter o carácter não determinado das nossas acções, alguns teóricos da acção defenderão que ao fornecer as razões da acção, a explicação que se avança será unicamente de carácter teleológico e não nomológico ou causal. Se eu decido fazer X é à luz de certas razões, mas a relação entre as razões e a minha decisão será essencialmente intrínseca e não extrínseca; ou seja, a minha decisão leva em consideração as minhas razões mas não procede/emerge causalmente delas.

Autores causalistas como A. Mele, por seu lado, sustentarão que o chamado problema do livre-arbítrio só constitui um problema para as teorias causalistas se a explicação causal implicar o determinismo, se o conceito de causação for essencialmente determinista. E nem todos os autores causalistas se comprometerão com uma visão determinista.

Bibliografia

Anscombe, E. (1957), *Intention*, Oxford: Blackwell.

Audi, R. (1986), “Acting for reasons”, *Philosophical Review*, 95.

Bratman, M. (1987), *Intention, Plans and Practical Reason*, Cambridge: Harvard University Press (reeditado em 1999).

Cohen, J.L. (1981), “Can human irrationality be experimentally demonstrated?”, *Behavioral and Brain Sciences*, 4: 317-370.

Cosmides, L. & Tooby, J. (1992), “Cognitive adaptations for social exchange”, in Barkow, J., Cosmides, L. & Tooby, J. (1992), *The Adapted Mind: evolutionary psychology and the generation of culture*, Oxford: Oxford University Press, pp. 163-228.

Darwall, S. (1990), “Autonomist Internalism and the Justification of Morals”, *Noûs*, 24: 257-268.

Davidson, D. (1980), *Essays on Actions and Events*, Oxford: Oxford University Press.

_____ (2001), *Subjective, Intersubjective and Objective*, Oxford: Oxford University Press.

Frankfurt, H. (1982), "Freedom of the Will and the Concept of a Person," in Watson, Gary, ed. (1982), *Free Will*, Oxford: Oxford University Press, pp. 81-95.

Gigerenzer, G. (1991), “How to make cognitive illusions disappear: beyond heuristics and biases”,

- European Review of Social Psychology*, 2: 83-115.
- Ginet, C. (1990), *On Action*, Cambridge: Cambridge University Press.
- _____ (2002), "Reasons Explanations of Action: Causalist versus Noncausalist Accounts," in Kane, Robert, ed. (2002), *Oxford Handbook on Free Will*, New York: Oxford University Press, pp. 386-405.
- Goldman, A. (1970), *A Theory of Human Action*, Princeton: Princeton University Press.
- Harman, G. (1976), "Practical Reasoning", *Review of Metaphysics*, 79: 431-63.
- Holton, R. (2006), "The Act of Choice", *The Philosophers Imprint*, 6 (3): 1-15.
- Kahneman, D., Slovic, P. & Tversky, A., eds., (1982), *Judgment under Uncertainty: heuristics and biases*, Cambridge: Cambridge University Press.
- Kim, J. (1998), *Mind in a Physical World – an essay on the mind-body problem and mental causation*, Cambridge: The MIT Press.
- Korsgaard, C. (1986), "Skepticism about Practical Reason", *Journal of Philosophy*, 83(1): 5-25.
- Locke, J. (1999), *Ensaio sobre o entendimento humano*, Lisboa: Fundação Calouste Gulbenkian.
- McCann, H. (1974), "Volition and basic action", *Philosophical Review*, 83: 451-73.
- _____ (1998), *The Works of Agency: On Human Action, Will and Freedom*, Ithaca: Cornell University Press.
- McDowell, J. (1996), *Mind and World*, Harvard University Press.
- Mele, A. (1992), *Springs of Action*, New York: Oxford University Press.
- _____, ed. (1997), *The Philosophy of Action*, New York: Oxford University Press.
- _____ (2003), *Motivation and Agency*, New York: Oxford University Press.
- Miguens, S. (2002), *Uma Teoria Fisicalista do Conteúdo e da Consciência – D. Dennett e os debates da filosofia da mente*, Porto: Campo das Letras.
- _____ (2004), *Racionalidade*, Porto: Campo das Letras.
- Moya, C. (1990), *The Philosophy of Action*, Polity Press.
- Nagel, T. (1970), *The Possibility of Altruism*, Princeton: Princeton University Press.
- _____ (1979), "Moral Luck", in Nagel, T. (1979), *Mortal Questions*, Cambridge: Cambridge University Press.
- O'Shaughnessy, B. (1973), "Trying (as the mental 'pineal gland')", *Journal of Philosophy*, 70: 365-86.
- Taylor, Ch. (1976), "Responsibility for self", in Rorty, A., ed. (1976), *The Identities of Persons*, Los Angeles: University of California Press.
- Wittgenstein, L. (2002), *Investigações Filosóficas*, Lisboa: Fundação Calouste Gulbenkian.

_____ (1958), *The Blue and Brown Books*, Oxford: Blackwell.

Ação, decisão e explicação da ação – Conversa com António Zilhão

Carlos Mauro, Susana Cadilha e António Zilhão

António Zilhão é Professor no Departamento de Filosofia da Faculdade de Letras da Universidade de Lisboa desde 1990. Tem ensinado em áreas que vão desde a lógica à filosofia da mente, filosofia da acção e teoria da decisão. É autor de *Linguagem da filosofia e filosofia da linguagem – estudos sobre Wittgenstein* (1993), *Animal racional ou bípede implume? – Um ensaio sobre acção, explicação e racionalidade* (2010) e coordenador de *Evolution, Rationality and Cognition – a cognitive science for the 21st century* (London, Routledge, 2010), entre muitas outras publicações. É ainda tradutor de Wittgenstein e de Frege (*Os Fundamentos da Aritmética*, Lisboa, INCM, 1992) para o português. Estudou em Lisboa, Graz e Londres (onde se doutorou-se, no King's College). Foi investigador visitante no Max Planck Institut für Bildungsforschung de Berlin e no Center for Philosophy of Science da Universidade de Pittsburgh (USA).

Carlos Mauro (CM) e Susana Cadilha (SC): O que é a Filosofia da Acção?

António Zilhão (AZ): A Filosofia da Acção é aquela parte da filosofia que tem como tópico o estudo da acção humana. Há, evidentemente, um conjunto de problemas que a constituem. Estes definem, por sua vez, sub-áreas de estudo dentro da Filosofia da Acção. A título de exemplo, posso indicar os seguintes problemas e sub-áreas. Em primeiro lugar, o problema da identificação das acções. Refiro-me aqui ao problema de saber como se caracteriza uma acção – o que é uma acção? Em segundo lugar, o problema da individuação das acções. Refiro-me por meio desta expressão ao problema de determinar como é que se distinguem acções umas das outras. Em terceiro lugar, o problema da dimensão temporal da acção, isto é, o problema de determinar em que período do tempo é que é apropriado considerar-se que uma acção decorreu. Estes três problemas definem uma sub-área da Filosofia da Acção que poderíamos, talvez, designar como Ontologia da Acção. Outra sub-área desta disciplina é a da Metafísica da Acção. Aqui interessa, sobretudo, saber se as acções podem ser livres e, caso possam ser livres, em que é que consiste essa liberdade. Outra sub-área ainda é a da Epistemologia da Acção, a qual se constitui a partir da tentativa de perceber em que é que consiste explicar uma acção. Finalmente, há também problemas associados à compreensão da acção colectiva, ou seja, àquele conjunto de acções cuja execução pressupõe não apenas que o agente aja de uma determinada maneira, mas também que os outros agentes ajam igualmente de uma

determinada maneira. Neste sentido, é necessário compreender como é que emerge a coordenação necessária a que a acção do agente possa ter o conteúdo que nós lhe atribuímos, quais são os princípios subjacentes a essa coordenação, etc. Creio que estes são os problemas mais importantes com que a filosofia da acção lida, e que tem que discutir, e que, por conseguinte, aquelas que eu referi são também as sub-áreas nucleares da Filosofia da Acção.

CM e SC: Alguns autores incluem na filosofia da acção algumas questões de filosofia moral. Qual a sua opinião sobre isso? Considera que a filosofia da acção é uma área completamente separada da filosofia moral ou não?

AZ: Eu creio que a filosofia da acção pode ser prosseguida de uma forma independente da filosofia moral.

CM e SC: E quais são as relações objectivas que vê entre filosofia da acção e economia?

AZ: Há uma relação muito óbvia, que é a relação que se estabelece em virtude de uma das categorias essenciais subjacentes ao pensamento económico ser a categoria da racionalidade. Uma das coisas que os economistas debatem é precisamente a de saber o que é, em geral, ser racional, e como é que o agente pode agir racionalmente em certas situações específicas. Ora, a caracterização do conceito de racionalidade é também uma questão que é tratada na Filosofia da Acção. Na apresentação que há pouco fiz dos principais problemas que esta deve tratar, não me referi especificamente a este problema porque ele aparece integrado no âmbito daqueles tópicos que eu designei como sendo os tópicos da identificação das acções e da individuação das acções. O primeiro, como disse, investiga o que é que faz com que um certo evento seja considerado uma acção (ou seja, como é que nós distinguimos acções de não-acções) e o segundo investiga como é que distinguimos umas acções das outras. Ora, se prosseguirmos a discussão em torno destes tópicos, veremos que uma das categorias fundamentais que é chamada à colação para especificar as distinções necessárias ao seu tratamento é a categoria da racionalidade. Neste sentido, a

Filosofia da Acção lida com um aspecto fundador da teoria económica, que é precisamente o aspecto de determinar o que é ser racional no âmbito da acção.

CM e SC: Comecemos então pela questão da identificação das acções. O que pode contar como uma acção?

AZ: A este respeito, posso começar por dizer que acções são coisas que acontecem no tempo em associação com agentes. Mas é claro que podem acontecer coisas no tempo em associação com agentes que não são acções. Então o que é que faz com que possamos distinguir os comportamentos dos agentes que são acções dos que o não são? De uma forma geral, a resposta pode ser dada em termos da recondução dos primeiros às razões do agente. E aqui entra em cena o conceito de intencionalidade. Portanto, diz-se das acções que elas são intencionais, no sentido em que elas decorrem das razões do agente, ao contrário de outros comportamentos (instintivos, reflexos, etc.) que, sendo protagonizados também por agentes, não são, porém, considerados acções, precisamente porque não decorrem de razões do agente para se ter comportado da forma como se comportou. O que aparece como distintivo das acções é, então, a um primeiro nível, muito elementar, a sua associação com agentes (ao contrário de outras coisas que acontecem no mundo e que não estão ligadas a agentes), e, a um segundo nível, – aquele no contexto do qual temos que fazer a distinção entre comportamentos de agentes que são acções e comportamentos de agentes que não são acções – o facto de haver intencionalidade associada ao comportamento agencial. Depois podem ainda introduzir-se algumas especificações mais subtis. Por exemplo, num caso em que o agente faz alguma coisa intencionalmente, no sentido em que está a implementar um comportamento que decorre das suas razões, ele pode também, ao mesmo tempo, e de forma não intencional, gerar outras consequências. Considera-se que essas consequências são resultado da acção do agente, e portanto são atribuíveis ao agente como sua acção, apesar de não terem sido intencionais. Como é que então podemos incorporar estas duas dimensões do problema? Dizendo, por exemplo, que um determinado comportamento dirigido do agente conta como uma acção se houver uma descrição desse comportamento de acordo com a qual esse comportamento foi intencional. Mesmo que tenha gerado outras consequências que não foram intencionais. Mas

como há pelo menos uma descrição desse comportamento debaixo da qual esse comportamento conta como uma acção, então essas consequências contam também como consequências da acção, o que permite que o agente seja responsabilizado, etc.

CM e SC: Um movimento instintivo do corpo, por exemplo: afastar um mosquito enquanto estudamos. Em que sentido esse movimento não pode ser considerado uma acção, ou porque é que não pode ser considerado intencional?

R: Todas as definições podem, evidentemente, ser postas em causa. Mas a tradição da Filosofia da Acção é considerar como acções aqueles comportamentos que decorrem de razões. Se remontarmos, por exemplo, a Aristóteles, o que nele encontramos é o estabelecimento de uma analogia com a razão teórica. Aristóteles argumenta reflexivamente que, no âmbito da razão teórica, a admissão de certas proposições parece compelir-nos à concordância com outras proposições. É claro que nem sempre isso acontece – por vezes eu tomo certas proposições como verdadeiras, mas, no entanto, não sinto em mim qualquer necessidade que me leva a pensar que, se eu aceitei aquelas proposições como verdadeiras, tenho que aceitar esta ou estas outras como verdadeiras. Os casos em que eu detecto essa necessidade são precisamente aqueles em que as proposições que eu aceitei como verdadeiras podem ser tomadas como premissas e solicitam que eu efectue uma inferência de carácter peculiar para uma outra proposição que será a conclusão. Estamos aqui perante a descoberta dos princípios da inferência lógica – se eu concordo com a proposição A e com a proposição B, sinto-me obrigado a concordar também com a proposição C, sendo que C decorre, como conclusão lógica, de A e B, tomadas como premissas. Isto é uma coisa que se nos impõe quando reflectimos sobre a nossa própria prática de argumentar e de pensar reflexivamente. Como disse, em Aristóteles, as considerações acerca da razão prática surgem por analogia com as considerações acerca da razão teórica. Há casos nos quais o pensamento parece implicar uma acção e há casos em que isso não acontece. Quando Aristóteles reflecte sobre qual é a diferença específica que acompanha os casos em que há esta conexão da acção com o pensamento que parece ser semelhante à conexão que conclusões têm com premissas, o que ele constata é que essa diferença específica é que os pensamentos em causa constituem *razões* para a acção. Isto é, segundo ele, a partir do momento em que eu tenho aqueles pensamentos, sinto-me compelido a agir de certa maneira, de uma forma não

exactamente igual a, mas com parecências muito relevantes com, aquela por meio da qual me sinto compelido a extrair uma conclusão a partir do momento em que aceito as premissas das quais ela se segue logicamente. A Filosofia da Acção nasce, portanto, a partir desta tradição, na qual se considera que o objecto a ser tratado por esta disciplina são aqueles comportamentos que resultam, de uma forma análoga ao modo como as conclusões resultam das premissas, de pensamentos anteriores que estão com eles numa determinada relação de consequência. Neste sentido, e dado este enquadramento, comportamentos instintivos ou reflexos, que são desempenhados de forma independente de qualquer pensamento, não contam, por definição, como acções.

CM e SC: Para que um movimento de um corpo seja considerado uma acção é necessário que seja consciente, ou não?

AZ: Aqui a pergunta não tem tanto que ver com a acção mas com o pensamento. Dada a definição anterior de o que é uma acção, acções têm que estar conectadas de determinado modo com razões, e as razões têm que ser conteúdos de pensamentos. Agora, a questão da presença à consciência ou não presença à consciência já depende do modo como se define pensamento. Se tivermos uma definição cartesiana de pensamento, não pode haver acções que não estejam associadas a razões conscientes porque, deste ponto de vista, qualquer razão tem necessariamente que ser consciente. Mas, evidentemente, podemos ter uma concepção não cartesiana de pensamento. Ora, em contextos teóricos não cartesianos, o facto de um conjunto de itens de vida mental não ser consciente não quer automaticamente dizer que ele não possa constituir pensamento e, constituindo pensamento, que não possa estar naquela relação particular de consequência com o comportamento desempenhado pelo agente que nos permita classificá-lo como um conjunto de razões. Portanto, dentro de algumas destas concepções não cartesianas, pode ter-se pensamentos não conscientes que são razões. Nestas condições, o comportamento deles decorrente é uma acção.

CM e SC: E a sua opinião qual é?

AZ: Eu não sou um cartesiano.

CM e SC: O movimento de um sonâmbulo conta como uma acção?

AZ: Porquê? Do facto de se dizer que pode haver pensamento não consciente não se segue que toda a actividade não consciente do cérebro seja pensamento. Em princípio, creio que ninguém vai considerar como acções comportamentos desempenhados quando o agente está a dormir. No sentido em que se usa o termo razões, mesmo quando se aceita que elas sejam não conscientes, não creio que esse sentido seja aplicado aos casos em que o agente está a dormir. Repare que o conceito de consciência também é um conceito ambíguo. Creio que subjacente à sua pergunta se encontra a mistura de usos diferentes deste conceito. De um ponto de vista não cartesiano, eu posso dizer de alguém que agiu, no sentido técnico do termo, apesar de essa pessoa não ter considerado de forma consciente na sua mente reflexiva as razões para a sua acção; mas isso não me obriga a dizer que essa pessoa agiu inconscientemente. Quando digo que existe pensamento inconsciente, refiro-me a pensamento que decorre sem que os seus conteúdos sejam definíveis como conteúdos presentes à consciência no momento anterior à acção. Isto é uma coisa. Mas, frequentemente, quando se diz de alguém que agiu inconscientemente pretende-se com isso dizer que essa pessoa agiu, por exemplo, de forma automática ou leviana ou irresponsável. Mas não é de todo claro que os segundos casos sejam assimiláveis aos primeiros. Precisamente, um dos fundamentos das abordagens não cartesianas do pensamento é que não o são. O par de termos consciente/inconsciente é usado de modo diferente num e noutra conjunto de casos.

CM e SC: No caso do sonâmbulo, não há razões, por isso não há acção. Mas não lhe parece estranho que aquilo possa ser considerado um mero movimento do corpo?

AZ: Não, não me parece estranho. Há com certeza uma teoria científica acerca do sonambulismo que explica esse comportamento, e essa teoria tem que passar pela compreensão dos processos que decorrem na mente de um agente quando ele está a

dormir. Portanto, há uma definição dos casos normais, e depois temos os casos de sonambulismo que são não-normais. Mas, seja como for, penso que os casos de sonambulismo têm que ser compreendidos dentro de uma teoria do sono. E eu creio que as categorias por meio das quais nós podemos formular uma teoria do sono não são categorias como acções, razões, etc. Em princípio as pessoas quando estão a dormir não agem.

CM e SC: Mas continuam a ser agentes?

AZ: Continuam, porque a vida de um agente continua enquanto ele está a dormir.

CM e SC: E o que acontece no caso do auto-engano? Quando o agente age de forma auto-enganada, qual a possibilidade que ele tem para não agir daquela maneira? Pode afirmar-se que o indivíduo auto-enganado sofre de algum tipo de bloqueio de acesso às reais razões?

AZ: Para se poder falar em auto-engano, tem que ser o caso que o agente conheça as verdadeiras razões, senão seria um caso de ignorância. Quando ignora, o indivíduo crê que B numa circunstância em que não-B; quando engana, o indivíduo sabe que não-B e, no entanto, afirma B. O indivíduo que se auto-engana mente a si próprio. Por isso é que há um problema. Isso é considerado algo parecido com um paradoxo. Como pode alguém enganar-se a si próprio? A condição necessária para eu estar a enganar é que eu saiba que aquilo que afirmo como verdade é, na realidade, uma falsidade. Se o destinatário do meu acto de fala é outra pessoa, não há problema; mas se sou eu o destinatário do engano, eu não devo poder enganar-me a mim mesmo, na medida em que, se eu sei que algo é falso, o facto de o afirmar a mim mesmo como verdadeiro não deveria ter qualquer efeito sobre mim porque, precisamente, eu sei que é falso. Mas, no entanto, existe de facto um fenómeno psicológico que é referido como auto-engano e, portanto, há necessidade de compreender como é que isso é possível, porque, *prima facie*, não parece ser possível.

CM e SC: **Mas aquilo que o indivíduo que se auto-engana faz, trata-se de uma acção?**

AZ: É uma acção porque o indivíduo age com base em razões. As razões na base das quais ele age é que podem não ser as razões na base das quais ele diz que age, mas há sempre a recondução da acção a razões. Mas, atenção, porque este contraste entre razões só por si ainda não envolve auto-engano; pode envolver apenas dissimulação para outros. Só se fala em auto-engano quando o indivíduo age com base em razões que não aquelas que ele diz que são as razões com base nas quais ele age, sabe que o faz, e, além disso, se convence a si próprio disso que diz.

CM e SC: **Mas quando ele age de maneira auto-enganada já está convencido, caso contrário ele tem consciência de que se está a enganar e já não é auto-engano.**

AZ: De acordo com a minha experiência, pelo menos, creio que os casos de auto-engano são, com muita frequência, fictícios. De uma forma geral, creio que o que prevalece é que as pessoas sabem que não estão a agir com base nas razões com base nas quais dizem que agem. O auto-engano tem esta dimensão de ilusão, envolve mentira. Isto é uma coisa. Há outra coisa diferente, que é haver engano, no sentido de erro, do agente acerca das razões com base nas quais ele agiu. Mas isso não é auto-engano. Eu posso estar convicto que agi com base num certo número de razões quando, na realidade, não agi com base nessas razões e agi com base noutras. Mas aqui não estou a enganar-me a mim próprio; estou enganado acerca de mim próprio, o que é diferente. Claro que, neste caso, põe-se a questão de saber como é que se pode descobrir as verdadeiras razões; mas, supondo que isso é possível, pode dar-se o caso de eu estar convicto que agi com base nas razões x e y quando, na realidade, agi com base nas razões w e z. Como disse, um caso de engano acerca das razões subjacentes à minha acção não é auto-engano porque eu não estou a enganar-me a mim próprio, estou apenas enganado a meu respeito, como quando penso que a dor que tenho nas costas é provocada por um problema na coluna vertebral quando ela nada tem a ver com a coluna vertebral mas com outra condição subjacente (uma pedra num rim, por exemplo).

CM e SC: **Então afirma que o auto-engano acontece menos do que parece.**

AZ: Eu acho que é uma coisa que roça a patologia... Mas não sei; se calhar acontece com frequência nos consultórios de psiquiatria...

CM e SC: **E nesse sentido pode ser um problema filosófico?**

AZ: Se ele é realmente possível, é um problema filosófico porque, dada a maneira como os termos são usados, parece haver aqui um paradoxo.

CM e SC: **Continuando a explorar a questão da identificação e da individuação da acção, qual pode ser então a relação entre a intencionalidade e a racionalidade?**

AZ: Suponho que a vossa questão visa saber se os conceitos de intencionalidade e racionalidade não serão congruentes, isto é, se não referirão, no fundo, a mesma coisa. Creio que eles têm conteúdos diferentes. O conceito de ser intencional é aquele conceito que se encontra subjacente à classificação de um comportamento como uma acção; de uma forma geral, ele é para ser entendido como classificando aqueles comportamentos dirigidos que foram desempenhados pelo agente a partir de um conjunto de razões que ele tinha para proceder daquela forma. Portanto, o conceito de intencionalidade é um conceito que reconduz a acção às suas razões. O conceito de ser racional é, creio, mais específico. Quando se classifica uma acção como racional, não se trata apenas de reconduzi-la às razões do agente para agir daquela forma; trata-se também, e sobretudo, de analisar se a conexão entre a acção e as razões das quais ela deriva é a conexão apropriada. Portanto, nós podemos considerar que uma determinada acção é intencional, no sentido em que decorre de razões que o agente teria tido para agir daquela forma, mas, no entanto, podemos, acerca dessa mesma acção, dizer que ela não é racional, no sentido em que o modo como ela decorre das razões do agente das quais efectivamente decorre não é o modo mais apropriado. É

neste sentido que o conceito de racionalidade é mais estreito do que o conceito de intencionalidade.

CM e SC: O termo “apropriado” remete para as teorias normativas acerca da racionalidade, formuladas por filósofos, economistas, etc.? E, se sim, como escolher entre as diferentes alternativas? Como argumentar a favor de uma em detrimento de outra?

AZ: A distinção que eu estabeleci entre intencionalidade e racionalidade parece supor que a caracterização de uma acção como racional subentende não apenas a recondução da acção às razões do agente para agir do modo como agiu mas também a avaliação da conexão entre razões e acção segundo um determinado padrão de medida; se a conexão entre a acção e as razões do agente estiver de acordo com esse padrão será racional, se não estiver não será racional. Essa suposição está correcta: o conceito de racionalidade supõe de facto a existência de um padrão de medida. Mas uma coisa é reconhecer-se que o conceito de racionalidade supõe, de facto, a comparação com um padrão e a efectuação de uma avaliação em função dessa comparação; outra coisa é identificar-se esse padrão com rigor; e, outra coisa ainda, é determinar-se qual é o estatuto desse padrão de medida – se é um estatuto puramente normativo, como é argumentável que tenham as regras morais, por exemplo; ou se é um padrão que tem alguma validade empírica que não se esgota apenas numa caracterização normativa; ou, ainda, uma terceira possibilidade que certos autores defendem, se esse padrão tem um estatuto prescritivo. (Prescritivo não tem que ser a mesma coisa que normativo. Nós podemos prescrever a alguém que aja de determinada maneira por ser o melhor modo de atingir os seus objectivos, mas essa prescrição não tem necessariamente que ter o estatuto de uma norma universal, etc.). E aqui pode haver, e há, respostas substancialmente diferentes.

CM e SC: E qual é a origem desses padrões?

AZ: A resposta pode ser dada ao longo de linhas diferentes. Na tradição filosófica europeia, pode reconduzir-se ao pensamento de Aristóteles a ideia de que, por um

lado, esse padrão existe, e que as acções devem ser avaliadas em função dele e, por outro lado, que a justificação para a existência desse padrão é que ele institui um paradigma de optimidade. Ou seja, concebido desta maneira, o conceito de racionalidade é, em primeiro lugar, um conceito relacional, o qual avalia a relação na qual a acção se encontra com as razões do agente, e é, em segundo lugar, um conceito de optimidade, no sentido em que considera que aquela, de entre o leque de diferentes acções possíveis em cada situação, que merece o qualificativo de racional (que está de acordo com o padrão) é a que melhor permite satisfazer os objectivos do agente em função das crenças que ele tem acerca do mundo. Esta é uma ideia muito antiga, que se encontra já nos gregos e, se formos estudar a evolução desta ideia ao longo da história da Filosofia ocidental, aquilo que verificamos é que a evolução que a ideia sofre resulta, sobretudo, de um esforço no sentido de torná-la mais precisa, mais rigorosa. Mas, em traços largos, é uma ideia que encontramos já em Aristóteles e mesmo antes dele.

CM e SC: Como escolher entre esses diferentes critérios ou padrões? Qual é a melhor forma de caracterizar o conceito de racionalidade?

AZ: Também é necessário fazer algumas distinções para poder responder a essa pergunta. Em primeiro lugar, temos a associação do conceito de racionalidade com o conceito de optimidade. Isto é uma coisa que pode discutir-se nos seus próprios termos. Pode, por exemplo, pôr-se em dúvida se, de facto, o primeiro conceito deve ser analisado à custa do segundo. Em segundo lugar, e supondo que se aceita a associação entre racionalidade e optimidade, fica ainda de pé a questão de saber como é que o próprio conceito de optimidade deve ser elucidado. Há maneiras diferentes de fazê-lo. E nem sempre é claro qual delas é a mais adequada.

CM e SC: Por exemplo, a maximização da utilidade esperada seria uma forma possível?

AZ: Sim, é aquela forma que adquiriu maior proeminência nos últimos duzentos anos.

CM e SC: E a ideia da satisfazibilidade de Herbert Simon e Gigerenzer seria outra forma possível? Quando caracteriza as diferentes possibilidades de elucidar o conceito de optimidade é a essas alternativas que se refere? A disputa encontra-se a que nível? Ao nível da utilização do próprio conceito de optimidade, ou das diferentes formas de caracterizar esse conceito?

AZ: Há disputa em todos os níveis, na realidade. Se considerarmos a discussão ao nível de se tentar saber qual é a melhor forma de caracterizar o conceito de optimidade, temos que começar por verificar que tem havido aqui uma evolução. A ideia da maximização da utilidade esperada não é a mesma que a do princípio da expectativa matemática, por exemplo. E depois de a primeira se ter imposto à segunda, novas distinções foram estabelecidas: o princípio da maximização da utilidade esperada *tout court*, o princípio da maximização da utilidade subjectiva esperada, etc. Esta é, como disse, uma discussão que decorre em torno da questão de determinar qual é a melhor forma de caracterizar o conceito de racionalidade concebido como optimidade. Agora, pode-se não aceitar que a melhor forma de caracterizar o conceito de racionalidade seja por meio do conceito de optimidade. Na segunda metade do século XX apareceu um conjunto de autores, como Herbert Simon, por exemplo, que defendeu a ideia da racionalidade como satisfazibilidade. A partir de investigações empíricas, seja em inteligência artificial, seja na psicologia cognitiva, passaram a introduzir-se na consideração do problema dados que tradicionalmente não eram nele introduzidos, nomeadamente, dados que têm que ver com as limitações cognitivas de um agente que tem que agir em tempo real. E, então, estes autores chamaram a atenção para o facto de que muitos dos modelos clássicos correm sérios riscos de não serem empiricamente realizáveis, no sentido em que supõem que os agentes têm uma capacidade computacional enorme e que têm muito tempo à sua disposição para tomar as suas decisões, o que, na vida real, normalmente não é o caso. Alguns destes autores, como Gigerenzer, por exemplo, fizeram então a proposta de substituir o conceito tradicional por um conceito ecologicamente situado de racionalidade. Esta é uma proposta que afirma que a ideia de racionalidade não pode ser entendida separadamente das condições cognitivas do agente, das circunstâncias nas quais ele age e das variáveis ambientais que são relevantes para o sucesso ou insucesso da acção. Uma avaliação de racionalidade suporia então a

consideração de todas essas variáveis e, à luz destas, estes autores argumentam que pode ser mais racional para o agente procurar um certo patamar de satisfação em vez de tentar otimizar de acordo com a ideia subjacente aos modelos clássicos.

CM e SC: A este respeito não pode ser introduzido algo como a distinção entre competência e performance? Ou seja, o critério de caracterização da racionalidade como optimidade mantém-se, e as nossas acções devem ser avaliadas em função dele; o que se passa é que o agente real nem sempre consegue agir de acordo com ele porque muitos outros factores entram em jogo.

AZ: Para avaliar esta sua sugestão temos que ver o que está em jogo com a introdução da distinção entre competência e performance na polémica em torno da cognição. Esta distinção foi introduzida por Chomsky no âmbito da Linguística. Os princípios teóricos subjacentes a essa distinção são basicamente os seguintes. Aquilo que o linguista chomskiano tem que fazer é caracterizar a competência linguística do agente. Ora, a competência linguística do agente é, digamos assim, o software da sua faculdade da linguagem. Acontece que o software da faculdade da linguagem não pode ser completamente dado pela observação do comportamento linguístico do agente porque falar é uma acção física que envolve a mobilização de um conjunto de mecanismos de carácter físico de que o corpo humano dispõe, os quais têm limitações, também elas físicas, que canalizam de uma certa forma o desempenho linguístico dos agentes. Neste sentido, a argumentação de Chomsky é que a compreensão da competência exige não apenas a observação do comportamento real, mas também a compreensão de quais são os princípios computacionais de carácter geral subjacentes por meio dos quais esse comportamento real pode ser implementado. E esses princípios podem ir, e é natural que vão, muito para lá do que o comportamento real exhibe. Daqui segue-se que deverão existir estruturas linguísticas que nós podemos modelar por meio de construções formais, as quais, dado o seu carácter puramente formal, admitem a realização de operações tremendamente complexas que nós não vamos, todavia, encontrar no comportamento linguístico efectivo (iterações sucessivas, por exemplo). No âmbito deste último, nós podemos encontrar algumas exemplificações de baixa ou muito baixa complexidade desse género de construções, mas nada que se compare com aquilo que a posse de um

mecanismo computacional de geração dessas operações nos permitiria, em princípio, fazer. Isto é, a competência instalada é algo que é representável por uma função, no sentido de uma função matemática, e a compreensão teórica da competência exige a compreensão abstracta dessa função, mesmo que, na prática, dadas as limitações físicas do aparelho fonador humano, o comportamento real apenas exiba uma muito pequena parte do que seria possível de realizar dada a presença no aparelho cognitivo humano de um mecanismo que implementa fisicamente a função. Esta é a ideia subjacente à distinção entre competência e desempenho introduzida por Chomsky a partir das suas investigações em Linguística. Trata-se, basicamente, de uma distinção entre competência computacional, a qual tem um carácter formal, abstracto, matemático, se quiser, e desempenho prático, o qual é condicionado por outros factores, de carácter mecânico, que são independentes da competência formal. A questão então é: não se poderá aplicar o mesmo método de análise à acção humana? Quando autores como Simon falam das limitações cognitivas dos seres humanos, não poderemos defender que o que se passa é que existe uma competência instalada que vai muito para além do desempenho efectivo, mas esse desempenho efectivo tem que ser entendido em função da competência computacional instalada? A este respeito há um conjunto de coisas que podem ser ditas: primeiro, mesmo falando apenas em termos de competência instalada, há modelos de tomada de decisão que envolvem uma complexidade computacional de tal modo vasta que se justifica a dúvida sobre a possibilidade de que a competência subjacente possa encontrar-se nalgum órgão biológico, mesmo esquecendo por momentos questões associadas ao desempenho. Mas, ainda que não fosse esse o caso, esta abordagem corre, em segundo lugar, o risco de perder de vista aquilo que é essencial na acção efectiva. Porque não é a mesma coisa ter uma competência complexa capaz de tratar de modos computacionalmente muito sofisticados uma multiplicidade gigantesca de variáveis, a qual tem depois que ser canalizada através de um mecanismo que limita fortemente as suas potencialidades, e ter que agir de forma útil numa situação, na qual, seja qual for a complexidade computacional da competência subjacente, a informação relevante disponível é necessariamente muito limitada. Na realidade, no âmbito da decisão e da acção, a questão põe-se, com frequência, nestes termos. Outra variável que também adquire, em circunstâncias práticas, uma importância fundamental é a variável temporal. E esta tão-pouco pode ser reduzida apenas ao facto de ter que se “encaixar” em menos tempo aquilo que foi previsto para mais tempo.

CM e SC: Quais são, então, as possíveis formas de considerar uma acção irracional?

AZ: Para responder a essa pergunta o primeiro ponto a mencionar é saber porque é que isso é um problema. É um problema porque, se um determinado comportamento dirigido só é uma acção se, por oposição a comportamentos instintivos ou reflexos, ele decorrer das razões de um agente, então pode levantar-se a este respeito a seguinte objecção: mas então não é possível haver acções irracionais porque, por definição, toda a acção é racional, no sentido em que, para um comportamento poder ser classificado como uma acção ele tem que fazer apelo a razões subjacentes, as quais se supõe que são aquelas que o tornam racional. Portanto, se se faz uma distinção entre acção racional e irracional, isso significa que podem existir acções que decorrem das razões do agente mas que não são racionais em qualquer outro sentido do termo. Quais são então as circunstâncias nas quais uma acção pode decorrer das razões do agente mas não ser racional? É claro que, para responder a esta questão, o sentido útil do termo ‘ser racional’ é aquele de que eu já falei, quando contrastei racionalidade com intencionalidade. À luz deste uso do termo, há, então, várias possibilidades de irracionalidade. Se supusermos que o princípio de racionalidade subjacente é o requisito da maximização da utilidade esperada, uma acção pode não o satisfazer e, logo ser irracional, se, por exemplo, o agente for um mau calculador; se ele calcular mal as suas utilidades esperadas. Podemos imaginar que, embora “programado” para funcionar dessa forma, o mecanismo agencial, como qualquer outro mecanismo, estaria sujeito a disfunções que o levariam a cometer erros na aplicação desse princípio e, portanto, nesses casos, a acção por ele seleccionada não seria aquela acção que ele deveria ter seleccionada para desempenhar à luz do princípio instalado. Essa acção teria assim, apesar de ter decorrido das razões do agente, sido irracional. Mas há outros casos de irracionalidade da acção possíveis, como é, por exemplo, o caso da *acrasia*. Neste caso, o agente age em função de uma razão que, sendo sua, não é, no entanto, aquela em função da qual lhe teria sido possível agir da melhor forma e o agente sabe que esse é o caso. Este caso distingue-se do anterior, portanto, porque aqui o agente sabe qual seria a acção a desempenhar de acordo com o princípio de racionalidade que ele está disposto (ou predisposto) a seguir (não houve

erro de cálculo), mas não desempenha essa acção. Desde que a acção que ele efectivamente desempenhou decorra também de razões que são suas nós podemos considerá-la como uma acção efectiva (é um comportamento intencional). Mas, sendo intencional, não é racional, porque o agente não agiu de acordo com a sua crença acerca de qual seria a melhor acção a desempenhar naquelas circunstâncias. Aqui, não se trata de a crença efectiva do agente a esse respeito estar errada; trata-se, antes, do facto de, estando certa ou errada, tanto faz, o agente não ter agido de acordo com ela. No primeiro caso, teríamos uma irracionalidade meramente objectiva, digamos assim; no segundo caso, teríamos uma irracionalidade subjectiva e, portanto, num certo sentido, mais forte.

CM e SC: Mas como se sabe que o agente sabe qual é a melhor acção possível, dado um certo princípio que ele próprio adopta? Essa afirmação que o agente sabe que deveria agir de uma certa maneira dadas as suas crenças e desejos, e dado um determinado princípio – o que a justifica?

AZ: Mesmo que não saiba, tem a convicção. O segundo tipo de irracionalidade é compatível com a existência de erro na mesma (a irracionalidade pode ser simultaneamente subjectiva e objectiva). Ele pode ter chegado à conclusão que uma determinada acção é a melhor com base num erro de cálculo também. Mas, no entanto, o que é determinante na formulação do problema é que o agente não age de acordo com a sua crença acerca de qual é a melhor acção a desempenhar, independentemente de saber se a sua crença acerca de qual é a melhor acção a desempenhar está ou não de acordo com algum dos princípios de racionalidade existentes no mercado.

CM e SC: Por que não se pode afirmar que as razões que o agente considera serem as melhores são de facto aquelas que o conduziram à acção, e não as que ele considerava em princípio serem as melhores?

AZ: É possível ter casos de *acrasia* nos quais a acção que o agente implementa é aquela que, numa análise objectiva, é a melhor para o agente e, no entanto, essa acção

é acrática porque o agente está convicto de que a acção cujo desempenho seria melhor para ele seria outra, e no entanto ele desempenhou aquela. O que interessa para a consideração de uma acção como acrática é a convicção que o agente tem acerca de qual é a melhor acção a desempenhar. Outra questão, que é diferente, é a questão de saber se essa convicção está correcta ou não está correcta.

CM e SC: E se essa convicção for um engano?

AZ: Mas o conceito de *acrasia* é um conceito subjectivo – caracteriza um fenómeno de carácter subjectivo. O que interessa é a crença que o agente tem acerca de qual é a melhor acção a desempenhar, e o facto de ele não agir de acordo com essa crença, apesar de a ter. Mesmo que essa crença que ele tem, se analisada objectivamente, for uma crença errada, e mesmo que se dê o caso que, se ele tivesse raciocinado adequadamente, teria chegado à conclusão que, de facto, a melhor acção a desempenhar seria aquela que ele realmente desempenhou, ainda assim estaremos perante um caso de *acrasia*. Tal como uma dupla negação resulta numa afirmação, a conjugação de uma irracionalidade objectiva (disfunção do mecanismo, por exemplo) com uma irracionalidade subjectiva (*acrasia*) pode, por acaso, resultar numa escolha que é, numa avaliação externa puramente objectiva, a melhor possível.

CM e SC: Está considerando que há um erro mas, e se essa crença for fruto do auto-engano?

AZ: Mas isso significa que, na realidade, o agente acha que a melhor acção para ele desempenhar é aquela que ele efectivamente desempenha, só que mente ou se ilude a esse respeito. Se ele está a auto-enganar-se quando diz ou pensa que a melhor acção a desempenhar é outra que não aquela que ele desempenhou, sendo esta última aquela que ele, de facto, acha que é a melhor, então, na realidade, não há *acrasia*, só há simulação de *acrasia*.

CM: O que é que me impediria de afirmar que todo o caso de *acrasia* é um caso de auto-engano?

AZ: Pela própria definição do conceito não pode ser. O que pode dizer, se quiser, é que essa definição do conceito não faz sentido porque, quando vamos ver casos concretos, esses casos concretos são, na realidade, casos de auto-engano e não de *acrasia* genuína. O que pode afirmar é que a *acrasia* não existe, que o conceito é vazio de instâncias. Isso pode ser dito. Há autores que afirmam que não há *acrasia*, que ela é um falso problema. Gary Watson, por exemplo, diz que a *acrasia*, tal como é tradicionalmente definida, não existe. Ele opera então uma modificação no conceito. Evidentemente, se modificarmos o sentido dos termos, podemos sempre dizer o que quisermos. Mas, partindo do princípio que a definição cobre um fenómeno efectivamente existente, então, pela sua própria definição, a *acrasia* não pode ser auto-engano. No caso do auto-engano, aquilo que acontece é que o agente diz para si e para os outros que a sua crença acerca de qual é a melhor acção a desempenhar é que essa acção é a acção y, mas, no entanto, a crença que ele próprio, de facto, tem acerca de qual é a melhor acção a desempenhar é a de que essa acção não é y, é x, que é a acção que ele de facto desempenha, e ele sabe que esse é o conteúdo da sua crença.

CM e SC: Mas o auto-engano não pressupõe convicção?

AZ: Mas aí voltamos à questão que já discutimos – a de saber se existe auto-engano genuíno...

CM e SC: Vê alguma possibilidade de melhoria da discussão sobre a *acrasia* ou auto-engano partindo de experiências na psicologia cognitiva ou filosofia experimental?

AZ: Creio que sim. Como sabem, escrevi um conjunto de artigos que têm, precisamente, como objectivo mostrar como se pode dar conta do fenómeno da *acrasia* no âmbito de uma certa visão da mente humana que recorre ao apelo a

heurísticas rápidas e frugais como antecedentes explicativos de um conjunto relevante de acções. Uso aqui a expressão ‘dar conta’ no sentido davidsoniano, isto é, no sentido de mostrar como é que a acrasia é possível definindo as suas condições de possibilidade.

CM e SC: Concorda com o uso indistinto dos termos “fraqueza da vontade” e “*acrasia*”?

AZ: O termo ‘fraqueza da vontade’ induz as pessoas em erro. Faz as pessoas pensar que se está a falar de uma coisa que não é aquela da qual se está a falar – isto é, que estamos a falar da vontade, concebida como uma das faculdades da mente, quando estamos a falar do género de relação que existe entre uma acção e as suas razões, e isso não tem necessariamente que envolver uma faculdade específica chamada ‘vontade’. Também podemos usar a palavra ‘incontinência’, que, por contraste com ‘acrasia’, tem a vantagem de ser uma palavra portuguesa e, por contraste com ‘fraqueza da vontade’, não induz as pessoas em erro, ou, pelo menos, não as induz nesse erro. Mas pode induzi-las noutros ...

CM e SC: E entre “racionalidade prática” e “raciocínio prático”? Vê alguma distinção entre esses dois termos?

AZ: O termo racionalidade pode ser usado em dois sentidos diferentes. O primeiro está associado ao adjectivo “racional” e avalia a relação entre a acção e as razões subjacentes à acção em função da sua comparação com um padrão de medida, como já expliquei antes. O segundo está associado ao substantivo “razão”, o qual refere uma faculdade da mente humana. Neste segundo sentido do termo, fala-se de razão prática e raciocínio prático para referir aquele uso da razão que lida, por meio de pensamento sistemático, com questões práticas, por oposição àquele uso da razão que lida com questões teóricas. Neste segundo sentido, fala-se de raciocínio prático de uma forma mais ampla do que quando se fala em raciocínio prático no âmbito de uma teoria da racionalidade (no primeiro sentido). O raciocínio prático, no segundo sentido, é todo o raciocínio cujo objecto sejam questões práticas; nele inclui-se, por exemplo, o

raciocínio em torno de questões de carácter ético, moral, etc., enquanto que o raciocínio prático de que se fala no âmbito de uma teoria da racionalidade, no primeiro sentido, é aquele raciocínio com base no qual se tenta compreender a acção humana enquanto uma das instanciações do conceito de racionalidade agencial, que já elucidei antes.

CM e SC: Qual é o melhor modelo de explicação da acção, na sua opinião?

AZ: Creio que é difícil dar uma resposta unificada a esta pergunta. Recentemente, tenho andado a interessar-me bastante por heurísticas rápidas e frugais, como já vos disse, e pela possibilidade de explicar certos conjuntos de comportamentos e acções humanos a partir destes mecanismos extremamente simples. Mas não creio que esta seja uma forma universal de explicar a acção humana. Muito provavelmente, vamos ter que encontrar um modelo, digamos, bipartido. Neste modelo caberão acções que decorrem de razões que, com frequência, estão presentes à consciência, e cuja formulação mobiliza a faculdade de linguagem, etc., e, portanto, a compreensão destas acções desenrolar-se-á usando, de alguma forma, os mecanismos explicativos clássicos. Mas, ao mesmo tempo, creio que temos que arranjar nesse modelo um espaço onde possam caber acções que, sendo também determinadas por razões, são-no, porém, por razões geradas em partes da mente que nem têm que ser transparentes à consciência, nem têm que transitar pela faculdade da linguagem. Creio que estas acções têm características distintivas que, por um lado, impedem que as confundamos com comportamentos instintivos, reflexos ou outras formas de comportamento não agencial e que, por outro lado, as tornam atreitas a ser explicadas pelo funcionamento de mecanismos que implementem aquele género de procedimentos que são estudados em programas de investigação como o das heurísticas rápidas e frugais. Portanto, acho que precisamos de uma teoria global que junte partes que pertencem a modelos diferentes. Creio que, de uma forma geral, os modelos que existem falham enquanto modelos gerais porque estão centrados na compreensão de certos tipos de acções para os quais de facto funcionam e generalizam abusivamente a partir da consideração desses casos. Mas a mente humana é suficientemente complexa para nós necessitarmos de uma teoria que junte tanto heurísticas rápidas e frugais como

algumas formas de implementação de princípios de optimidade formal, como o da maximização da utilidade esperada, por exemplo.

CM e SC: Acha que existem pessoas que agem mais de acordo com o princípio da maximização do que outras?

AZ: Sinto que há aqui uma precisão a fazer. A distinção que eu estabeleci entre estruturas cognitivas do género das heurísticas rápidas e frugais e outras estruturas cognitivas mais ligadas à faculdade da linguagem e, provavelmente, também à consciência, não implica que as acções que decorrem destas últimas sejam necessariamente acções cuja compreensão deve ser efectuada recorrendo ao princípio da maximização da utilidade esperada. O princípio da satisfazibilidade de Herbert Simon, por exemplo, é para ser aplicado também a acções decorrentes destas estruturas.

CM e SC: O que estaria fora do conceito de satisfazibilidade? Em que momento este não se aplicaria?

AZ: No mundo económico, por exemplo, parece sem dúvida haver agentes económicos que pautam o seu comportamento por princípios da maximização da utilidade esperada.

CM e SC: Esses modelos pretendem ser só descritivos ou também normativos?

AZ: No caso da economia, o modelo é normativo no sentido em que, se o agente quiser ter sucesso no jogo económico, terá que agir de acordo com esse princípio. Portanto, será descritivo daqueles agentes que têm sucesso no jogo económico.

CM: A partir do desenvolvimento da teoria económica, e tendo como modelo o modelo da racionalidade ilimitada, tenho dúvidas que, em algum momento,

algum autor do século XX tenha aceite essa ideia de racionalidade ilimitada como descritiva. Eles teriam consciência da limitação dessas teorias.

AZ: Elas não pretendem ser descritivas da psicologia dos agentes mas sim da realidade económica.

CM e SC: Mas a teoria da acção económica não explica a acção de um indivíduo, não era essa a sua intenção.

AZ: Ela abstrai das características psicológicas dos indivíduos, mas supõe que, na medida em que esses indivíduos sejam responsáveis por comportamentos económicos, então esses comportamentos obedecerão a esse princípio de racionalidade objectivamente, seja qual for a realidade psicológica subjacente. E são prescritivos nesse sentido.

CM e SC: No seu último livro, Gigerenzer assume a condição prescritiva da teoria dele – ele defende que aquela é, de facto, a melhor maneira de agir.

AZ: Sim, uma das coisas que distingue o programa dele do programa de investigação de Kahneman e Tversky, por exemplo, é que o programa dele tem, de facto, uma dimensão prescritiva. O programa de investigação de Kahneman e Tversky tinha como objectivo exibir a irracionalidade subjacente ao comportamento humano, mostrando como ele é irracional em muitos casos, e tentando compreender as raízes psicológicas dessas irracionalidades. Aos psicólogos caberia então intervir junto das pessoas para lhes mostrar que, em inúmeras circunstâncias, elas seriam espontaneamente irracionais e que isso deveria ser corrigido para o seu próprio bem. Enquanto que Gigerenzer defende que muitos desses comportamentos *prima facie* irracionais são, na realidade, boas soluções para certos problemas com que os membros da espécie se teriam defrontado amiúde na sua história evolucionária. Nalguns desses casos, ele tenta modelar esses comportamentos à custa de heurísticas rápidas e frugais. Nesse sentido, as soluções corporizadas nesses comportamentos seriam prescritivas (para esses problemas, os quais podem, no entanto, ser apenas

vagamente semelhantes àqueles que suscitaram o desencadeamento do comportamento em causa). Nalgumas das suas investigações, ele demonstra, inclusivamente, que há casos nos quais as soluções decorrentes dessas heurísticas funcionam melhor do que soluções rivais obtidas a partir de tentativas de optimização.

CM e SC: Mas até que ponto ele considera que aquelas escolhas que Kahneman e Tversky mostram que são irracionais são, de facto, irracionais? Ou ele considera antes que são acções melhor adaptadas a determinadas circunstâncias, e portanto não irracionais?

AZ: O que ele faz, de facto, é pôr em causa a validade de algumas das conclusões de Tversky e Kahneman. Há um conjunto de experiências que ficaram famosas. Por exemplo, numa delas, mostra-se que há circunstâncias, não muito difíceis de obter, nas quais os agentes parecem violar um princípio elementar do cálculo de probabilidades – trata-se de casos nos quais os agentes fazem escolhas nas quais atribuem maior probabilidade a uma conjunção do que a um dos seus termos. Ora, isto vai contra um princípio muito básico desse cálculo. Mas Gigerenzer põe em causa a validade das experiências por meio das quais isso é demonstrado. Ele argumenta que existem outros factores, que não apenas o raciocínio probabilístico, que entram no desenho dessas experiências, nomeadamente, factores semânticos. Basicamente, o que ele diz é que o problema revelado nessas experiências não se encontra nos agentes testados mas sim nos psicólogos que os testaram, que conceberam testes inadequados. É nesta base que ele argumenta contra algumas das conclusões de Tversky e Kahneman. Ou seja, na base de que as experiências com base nas quais eles extraíram as suas conclusões não legitimam essas conclusões porque existe a possibilidade de interpretá-las de modos diferentes, igualmente plausíveis, e que não envolvem a detecção de irracionalidades. Quer dizer, não é de todo o caso que Gigerenzer defenda que existe uma heurística rápida e frugal por referência à qual se poderia propor que seria mais racional atribuir à probabilidade de uma conjunção um valor superior ao valor que se atribuiu à probabilidade de um dos conjuntos.

PARTE II

CM e SC: Voltando à questão da ligação entre teoria da acção e economia, queríamos perguntar-lhe, mais especificamente, se é possível dizer que uma preferência é análoga a um par crença-desejo.

AZ: Creio que é possível dizê-lo. Assim como também é possível dizer que o termo 'preferência' é um modo de formalizar apenas o conceito de desejo. É, por exemplo, habitual utilizar a expressão 'escala de preferências' de um agente para referir a mesma ideia a que Jeffrey se refere como 'escala de desiderabilidades' de um agente. Se tomarmos o termo 'preferência' nesta segunda acepção, ela não contém ainda em si a ideia de que a intensidade dos desejos deve ser ponderada por meio da consideração dos graus de crença que o agente tem na actualização dos diferentes estados possíveis do mundo que poderão ser relevantes para o desfecho da acção. Se tomarmos o termo 'preferência' na acepção a que se reporta na sua pergunta, então ela contém já em si a ideia dessa ponderação. Neste caso então, o termo 'preferência' refere o resultado do funcionamento não apenas de uma das partes do mecanismo decisional mas da totalidade do mesmo. É possível usar o termo em ambas as acepções. Não creio que tenha havido uma uniformização da terminologia a este respeito.

CM e SC: Não identifica especificamente a preferência com o desejo ou com algum resultado da dupla crença-desejo?

AZ: Acho que o termo pode ser usado das duas maneiras. Convém é saber qual é o sentido no qual o termo está a ser usado em cada contexto.

CM: E quanto ao conjunto de oportunidades, que é o conceito normalmente utilizado na economia para designar aqueles elementos que estão dentro do campo de possibilidades de acção do agente, existe algum termo/ideia análogo na filosofia da acção?

AZ: Pretende saber se esse conceito, independentemente do rótulo que usamos para o descrever, pertence também à teoria da acção, ou, caso exista um rótulo diferente na teoria da acção para o mesmo conceito, qual é esse rótulo? Quer discutir o rótulo ou o conceito?

CM: A ideia.

AZ: A ideia com certeza que é também utilizada na teoria da acção. Basicamente, trata-se da ideia de que o agente escolhe dentro de um domínio de possibilidades que é limitado e não absoluto.

CM e SC: O conceito de conjunto de oportunidades estaria ligado à formação das crenças – são as crenças do agente que o informam daquilo que é possível, ou não?

AZ: Também, mas não só. Existem igualmente os constrangimentos reais dentro dos quais o agente se move. Aliás, o próprio agente pode também agir no sentido de auto-limitar o conjunto das suas oportunidades futuras. Esta é uma das estratégias a que Elster se refere na sua teoria da racionalidade imperfeita. Segundo ele, esta estratégia constitui um dos modos de que os agentes humanos se socorrem para vencer o problema da acrasia. Trata-se de uma estratégia de acordo com a qual o agente faz algo agora cujo objectivo consiste em modificar as circunstâncias externas de um modo tal que o campo de oportunidades que estarão à sua disposição num momento de escolha futura significativa diminua de modo relevante. Portanto, para a delimitação do conjunto de oportunidades é importante considerar não apenas as crenças momentâneas do agente, como também os constrangimentos reais no seio dos quais ele tem que mover-se, como, ainda, a própria postura estratégica prévia do agente.

CM e SC: Um agente que não ignora os custos afundados ao decidir/agir será considerado irracional sob o ponto de vista da economia. Mas como defender este mesmo ponto de vista na teoria da acção? Como fazer essa passagem? (O

seu artigo “Efeito dos custos afundados, incontinência e racionalidade” é sobre este ponto, mas gostaríamos que explicitasse aqui essa ideia.)

AZ: Considerando tanto a acção que o agente desenvolve, como o contexto próximo no qual ela se desenvolve, será sempre irracional, tal como dizem os economistas, honrar custos afundados. De facto, se o princípio de racionalidade operante é o de que o agente deve agir apenas em função das consequências da sua acção, então, claramente, ao honrar custos afundados, ele está, para utilizar a gíria dos economistas, “a atirar dinheiro bom para cima de dinheiro mau”. Nada disto é posto em causa no artigo que escrevi. Aquilo que eu faço nesse artigo não é dizer que afinal é bom “atirar dinheiro bom para cima de dinheiro mau”, mas antes tentar encontrar um *rationale* para alguns comportamentos – não todos – de honrar custos afundados. O *rationale* que eu proponho consiste na ideia de que é possível o agente agir no sentido de se ensinar a si próprio uma lição. Ou seja, na ideia de que podem existir circunstâncias nas quais o agente persevera num dado curso de acção que vem na sequência de um investimento que ele fez, apesar de este ter sido um mau investimento e de o agente se ter apercebido disso, mas fá-lo com o intuito de mostrar a si próprio quais são as consequências da decisão que tomou. Se o problema em causa for um problema recorrente e se, após ter passado por esse processo, nas vezes futuras em que o agente se deparar com esse problema, ele for menos descuidado na avaliação do mesmo, então podemos dizer que há uma racionalidade específica envolvida no comportamento inicial de honrar os custos afundados. Com efeito, em tais circunstâncias, um comportamento como esse terá, na realidade, sido determinado por uma preocupação com a correcção do comportamento futuro do agente e não por uma preocupação com o seu investimento passado, o qual já está irremediavelmente perdido e que não há como recuperar. Esta é uma sugestão que eu faço no sentido de mostrar como é possível que tais comportamentos – que, quando concebidos de uma maneira puramente atómica, são rotulados de irracionais, e são-no correctamente – possam, se enquadrados num contexto mais alargado, ser compreendidos como componentes de uma estratégia racional do agente. Creio, inclusivamente, que faz sentido argumentar que uma tal estratégia pode ter uma justificação de carácter evolucionário. Repare-se que um processo de aprendizagem individual tem, normalmente, dois componentes. Temos, por um lado, o componente em que o aprendiz é sujeito a um processo de adestramento protagonizado por agentes mais

velhos, mais experientes ou com mais autoridade. No contexto deste processo interpessoal de aprendizagem, é claro que mecanismos deste género existem. Mas o processo de aprendizagem agencial tem também, parece-me, um segundo componente em que o agente age sobre si próprio no sentido de se auto-instruir. Creio que, no contexto deste segundo componente, comportamentos de honrar custos afundados podem, pelo menos em alguns casos, desempenhar um papel não negligenciável na aquisição pelo agente, e na estabilização futura nele, de padrões de comportamento não desperdiçadores. É óbvio que esta ideia não surgiu do nada. Existe investigação em torno desta questão efectuada no âmbito da Psicologia Cognitiva; e existe evidência empírica, obtida nestas investigações, que parece confirmar ou, pelo menos, dar algum suporte evidencial, a esta ideia.

CM e SC: Portanto, a ideia seria a de que, em alguns casos, seria possível detectar alguma racionalidade no comportamento de honrar custos afundados, a qual decorreria do conhecimento que o agente poderia ter acerca das vantagens desse comportamento. Mas a maioria das pessoas nem imagina que esse fenómeno exista...

AZ: Mais uma vez, para nos entendermos, creio que há aqui que distinguir três objectos de análise distintos: o fenómeno empírico, o conceito que o captura e o rótulo que refere este último. É um facto que a maioria das pessoas não detém o segundo e não conhece o terceiro. Mas isso não significa que, na sua vida real, essas mesmas pessoas não se deparem, com frequência, com problemas que são correctamente caracterizados de acordo com esse conceito que elas não detêm. Repare-se que as situações que dão origem ao fenómeno empírico e que são normalmente ‘encenadas’ nas investigações em Psicologia Cognitiva em torno deste problema são situações triviais: um indivíduo que comprou, por antecipação, um conjunto de lições de ténis e que, depois, descobre que, afinal, até nem gosta muito de jogar ténis, mas que, uma vez que já pagou, resolve levá-las até ao fim, muitas vezes com sofrimento; ou um indivíduo que compra, no âmbito de um pacote apenas parcialmente reembolsável, um fim-de-semana turístico para um determinado destino que não se encontra de todo no topo da sua lista de preferências, e que, depois de já ter feito a compra, mas antes de iniciar a viagem, encontra um outro pacote para o destino da sua preferência, a um preço bastante mais barato do que o anterior, mas

que acaba por escolher viajar para aquele para onde não prefere ir, não só porque já o pagou, mas também por ter sido mais caro, etc. Ora, a generalidade das pessoas já se confrontaram com situações deste género na sua vida de todos os dias. E, independentemente de não deterem o conceito, o que muitas vezes se detecta é que as pessoas se manifestam perplexas com o seu próprio comportamento, sem saberem muito bem explicar porquê. Creio que isso indica alguma coisa.

CM e SC: Imagine uma pessoa que produz um juízo desse tipo – já que comprei as lições de ténis, apesar de não gostar muito, vou aprender. Do ponto de vista davidsoniano, este juízo parece correcto. O que é que é preciso acrescentar para poder considerá-lo irracional?

AZ: A questão que se deve começar por formular para entender esta questão é a seguinte: há algum princípio de racionalidade que é violado pelo comportamento de honrar custos afundados? Ora, a resposta a esta questão é afirmativa. Há, de facto, um princípio de racionalidade subjacente que é violado em tais comportamentos. Trata-se do princípio de que as minhas decisões acerca daquilo que vou fazer devem ser tomadas em função das consequências futuras da minha acção e não em função do meu uso irremediavelmente passado dos recursos.

CM e SC: Mas, em Davidson, parece que o princípio de continência é o mais importante – é ele que define se uma acção é ou não racional.

AZ: Sim, o princípio da continência é um princípio muito básico; mas não é, obviamente, o único. Davidson aceita como válidos os princípios da teoria da decisão racional; ora, isso significa que, para além do princípio da continência, ele aceita também princípios como o da maximização da utilidade esperada, etc., como princípios de racionalidade agencial.

CM: Portanto, uma acção para ser racional tem que estar de acordo com vários princípios. No caso de Davidson, sempre pensei no princípio de continência como

o princípio crucial, e portanto qualquer juízo avaliativo que cumprisse esse princípio seria considerado racional.

AZ: O princípio da continência é um princípio muito básico. Não respeitar o princípio da continência é *prima facie* incorrer numa contradição: o agente acha que o que é melhor para ele é fazer X e, no entanto, faz Y, sendo que Y é, obviamente, uma instância de não-X. Violar este princípio muito básico é, desde logo, de acordo com a abordagem *standard*, incorrer num caso de irracionalidade. Mas isto não significa que o campo da irracionalidade se esgote com a violação do princípio da continência. Pode haver agentes que são continentos mas que são também irracionais noutro sentido, talvez não tão básico, mas também importante, porque o princípio da continência apenas diz que o agente deve agir em função da sua própria crença acerca de qual é a melhor forma de agir, mas não estabelece nenhum constrangimento acerca de saber o que é aquilo em que consiste uma dada forma de agir ser a melhor forma de agir para um agente. Portanto, podemos também considerar irracional um agente que é continente, mas que escolhe sistematicamente como a melhor forma de agir um comportamento que é obviamente abstruso para a obtenção das suas próprias finalidades. Deste modo, a questão da racionalidade ou irracionalidade da escolha de uma acção particular por um agente está também dependente da determinação, com um mínimo de precisão, de o que se entende por ser ‘a melhor forma de agir’. Ora, o grande *apport* teórico-cognitivo da teoria da decisão (e dos seus antecessores, como a teoria da expectativa matemática, etc.) consiste precisamente em apresentar uma forma rigorosa de caracterizar essa ideia vaga de o que é que significa realmente para um agente uma dada forma de agir ser ‘a melhor forma de agir’ nas circunstâncias. Estas teorizações introduzem uma clarificação desta ideia vaga por meio da ideia de concordância com um princípio de *optimidade*, o qual passa a ser explicitamente formalizado. Neste sentido, elas tornam possível que o grau de concordância das acções efectivas com esse princípio seja mensurável. E tornam também possível que se façam previsões e que o resultado dessas previsões seja posteriormente comparado com a realidade empírica. Por sua vez, diferentes pontos de vista teóricos propõem diferentes especificações particulares do princípio de optimidade em causa. Mas, seja como for, a partir do momento em que se passa a considerar que a forma de agir que é a melhor é aquela que é óptima num sentido específico e mensurável do termo, então, se um agente fizer uma escolha de uma acção como sendo a melhor nas circunstâncias

em que ele se encontra, e agir de acordo com ela, num caso em que a escolha que ele fez viola o princípio de optimidade que se considera ser o apropriado, então essa acção é também irracional, apesar de ser continente. Há, portanto, irracionalidade a vários níveis. Regressando ao caso do efeito dos custos afundados, parece-me claro que, se um agente tiver a crença de que a forma de agir que é a melhor para ele nas circunstâncias é aquela que honra custos afundados, e se ele agir de acordo com essa crença, então ele não terá sido irracional no sentido de ter sido incontinente, mas terá sido irracional no sentido de ter seleccionado como sendo a melhor uma forma de agir que contradiz o princípio de racionalidade de acordo com o qual devemos sempre decidir em função das consequências que antevemos para a nossa acção e não em função de o que quer que seja que a tenha precedido.

CM e SC: Estão aí em causa dois sentidos de racionalidade – um subjectivo, outro objectivo? Uma acção pode ser racional do ponto de vista subjectivo e irracional do ponto de vista objectivo, e vice-versa?

AZ: Exactamente. Um indivíduo pode ter uma crença acerca de qual é a melhor forma de agir, e depois não agir de acordo com essa crença. Ao não agir de acordo com a própria crença acerca de qual é a melhor forma de agir, ele está a ser subjectivamente irracional. Repare-se que, num caso como este, um observador externo até pode chegar à conclusão de que a crença com base na qual o agente agiu, e que não era, de acordo com a perspectiva do próprio agente, aquela na base da qual ele achava que teria sido melhor agir, representa, de facto, a melhor opção para o agente. Portanto, é possível ocorrerem situações nas quais o agente é subjectivamente irracional, mas nas quais ele acaba por ser objectivamente racional, se se considerar a situação interna de um ponto de vista externo, por assim dizer. Se nos abstrairmos do facto de que em irracionalidades subjectivas deste género não é suposto que exista qualquer intenção de praticar um logro, podemos tentar imaginar os casos em que elas ocorrem por analogia com aquelas situações nas quais um indivíduo mente dizendo uma frase verdadeira. Por exemplo, se eu tiver a crença errada de que a capital do Canadá é Toronto, eu posso mentir a alguém dizendo que a capital do Canadá é Otava. Mas o facto de a capital do Canadá ser Toronto ou Otava é independente daquilo que eu penso a esse respeito, e, na realidade, a capital do Canadá é Otava. Portanto, o valor

de verdade da minha afirmação é o valor de verdade verdadeiro, apesar de eu ter mentido. De um ponto de vista subjectivo, eu estou a violar um dos princípios que permitem a manutenção da racionalidade comunicacional, a saber, o princípio da sinceridade, mas, de um ponto de vista objectivo, estou a proferir uma frase verdadeira, isto é, estou a respeitar outro desses princípios, a saber, o princípio da veracidade, mesmo que sem o saber.

CM e SC: Pode em algum sentido comparar-se esses dois casos de irracionalidade? É pior ser irracional subjectiva ou objectivamente?

AZ: Num certo sentido, a irracionalidade da incontinência é mais profunda, uma vez que podemos associá-la à auto-contradição. Já da irracionalidade objectiva, presente naquelas ocasiões em que um indivíduo é continente, mas em que a escolha que ele faz como sendo a melhor não é de facto a melhor, de acordo com a informação que ele tem à sua disposição e de acordo com os seus desejos, pode dizer-se que ela é menos profunda, uma vez que, precisamente, não envolve essa dimensão de auto-contradição. Existem depois diagnósticos diferentes acerca da razão pela qual as pessoas agem irracionalmente neste segundo sentido, objectivo, do termo. Por exemplo, do ponto de vista utilitarista, ou hedonista, a justificação standard é a de que isso acontece quando as pessoas calcularam mal. Ora, assim como, nas ciências dedutivas, um erro numa multiplicação é considerado como sendo uma anomalia menos grave do que uma contradição lógica explícita, também no ‘ranking’ das irracionalidades práticas incorrer num erro de cálculo no processo maximizador não é considerado tão grave quanto incorrer numa auto-contradição.

CM e SC: O entendimento das diferentes teorias acerca do modo como se processa o raciocínio prático tem implicações na questão da determinação da racionalidade ou irracionalidade das acções, nomeadamente, na própria definição da acção acrática? Quais as hipóteses de resposta disponíveis para a questão de saber qual é o resultado do raciocínio prático?

AZ: Podemos considerar duas linhas possíveis de resposta a essa pergunta – a linha a que se convencionou chamar ‘cognitivist’ e a linha a que se convencionou chamar

‘hedonista’. De acordo com o ponto de vista cognitivista, o resultado do raciocínio prático é ainda uma crença do agente. Nomeadamente, a sua crença acerca de qual é a melhor forma de agir nas circunstâncias. Isto é o que encontramos em Davidson, por exemplo. O ponto de vista hedonista está mais próximo da formulação original de Aristóteles. Para os hedonistas, o resultado do raciocínio prático é a acção ela própria e não a formação de qualquer crença na mente do agente acerca de como agir. Neste sentido, se, para os hedonistas, o resultado do raciocínio prático é a acção ela própria, então a definição de acção acrática de que falávamos há pouco, ou seja, de uma acção que, sendo intencional e livre, não está de acordo com a crença do agente acerca de qual é a melhor forma de agir, teria que ser modificada, porque, do ponto de vista hedonista, a consideração de uma tal crença não tem, pura e simplesmente, lugar na definição da arquitectónica do mecanismo decisional do agente. Quer dizer, para o hedonismo, a ponderação das crenças acerca do mundo e dos desejos do agente de acordo com o algoritmo decisional desemboca de imediato na acção; logo, não há lugar para a ocorrência de um passo intermédio entre o raciocínio prático e a acção, como aquele cuja existência é postulada pelo cognitivismo.

CM e SC: Mas mesmo assim a acrasia seria possível.

Az: Vamos ver o que se entende aqui por dizer que “é possível”. Há uma intuição muito forte de que existem comportamentos que podem ser designados dessa forma, mas resta saber o que é que, na realidade, se entende pelo conceito, quando ele é usado deste ponto de vista. Na realidade, o que por ele se entende pode não ser exactamente a mesma coisa que o que por ele se entende no âmbito do pensamento cognitivista. Por exemplo, referi há pouco o nome de Elster. Trata-se de um autor importante neste contexto. Outro autor que partilha muitos dos seus pontos de vista a este respeito é Ainslie. De acordo com o pensamento de ambos, a acrasia surge com a introdução da temporalidade. Ela ocorre quando o agente não ordena a sua sucessão de escolhas no tempo de uma forma racional. Eles introduzem então na discussão a distinção entre taxa de desconto exponencial e taxa de desconto hiperbólica. E o que aconteceria com os agentes acráticos seria que a sua taxa de desconto seria tendencialmente hiperbólica em vez de ser tendencialmente exponencial. A taxa de desconto exponencial é racional, a taxa de desconto hiperbólica não é racional. O

facto de muitos agentes terem uma taxa de desconto hiperbólica explicaria então porque é que as suas preferências se inverteriam temporariamente na presença de estímulos sensoriais ou emocionais muito fortes, etc. Mas repare que, muitas vezes, não é sequer claro que o género de comportamentos que estaria em causa quando a referida inversão se produziria – aqueles comportamentos a que é frequente chamar de ‘sucumbir à tentação’ – contem como acções acráticas à luz da definição cognitivista de acrasia. Porque, à luz desta definição, para um comportamento ser uma acção acrática, tem que ter sido intencional e livre. Ora, é, pelo menos, argumentável que, em situações deste género, o agente não é livre. Quando o agente age sob uma descarga emocional extremamente forte ou quando sucumbe à tentação, por exemplo no caso de adições, pode perfeitamente argumentar-se que ele está sob o efeito de uma compulsão. Mas a compulsão define-se precisamente pelo facto de retirar a liberdade ao agente.

CM e SC: Podendo ainda ser uma acção intencional?

Az: Depende também. Para ser intencional, a acção tem que ter sido determinada por razões do agente. Ora, a ocorrência de uma descarga emocional, por exemplo, não perfaz tipicamente as condições necessárias para ser considerada uma razão para agir. Portanto, nesse caso, teríamos que tais comportamentos não seriam intencionais e, não sendo intencionais, não poderiam sequer ser considerados como acções *strictu sensu* do agente. Logo, do ponto de vista cognitivista, o facto de tais comportamentos serem desempenhados contra a crença do agente acerca de qual seria a melhor maneira de agir, só por si, não seria obviamente suficiente para que eles pudessem contar como acções acráticas. O que se retira daqui é então o seguinte. A existência de pontos de vista de fundo diferentes em torno da definição da arquitectónica do mecanismo decisional – nomeadamente, um ponto de vista hedonista e um ponto de vista cognitivista – tem, na realidade, importantes consequências teóricas. Em particular, a definição do conteúdo conceptual que corresponde a alguns dos termos estruturantes da própria teoria da acção não é independente da perspectiva filosófica de fundo que se adopta.

CM e SC: Seria possível, de algum modo, afirmar que a definição de acrasia dada por uma teoria hedonista comporta uma relação com elementos morais? Por exemplo, quando Aristóteles fala da acrasia, compara o acrata com o devasso. Já nas abordagens cognitivistas – Davidson e outros – o conteúdo moral da acção não é pertinente para a determinação da racionalidade ou irracionalidade.

AZ: Não tenho a certeza que esse seja o caso – só fazendo exegese do texto de Aristóteles é que se poderia resolver essa questão – mas eu creio que Aristóteles era mais sofisticado do que isso. Mas, mesmo supondo que não o fosse, daí segue-se apenas que, na formulação do problema por Aristóteles, havia uma mistura entre os elementos específicos que são próprios do conceito e outros elementos que, por vezes, se encontram associados com ele, mas que não lhe pertencem, e esses outros elementos devem ser separados do conceito por meio de um trabalho de análise. Seja como for, e independentemente de Aristóteles, não me parece de todo que se possa dizer que o ponto de vista hedonista actual acerca da acrasia esteja preso a quaisquer considerações morais. Por exemplo, do ponto de vista de Elster e Ainslie, como referi acima, o que está em causa na acrasia é o facto de, com a simples passagem do tempo, o agente agir contra a sua escala de preferências sem que tenha havido qualquer reversão explícita das suas preferências. Como também já o disse, este fenómeno é reconduzido, em particular por Ainslie, a um facto que ele considera ser de natureza puramente empírica, nomeadamente, o facto de, frequentemente, os seres humanos descontarem irracionalmente o futuro de uma maneira hiperbólica. Aqui perguntar-se-á: e porque é que descontar o futuro de uma maneira hiperbólica é irracional? O critério de racionalidade subjacente a este diagnóstico é o critério de racionalidade subjacente aos diferentes axiomas de racionalidade da teoria da decisão: um agente que desconte o futuro com base numa taxa hiperbólica é susceptível de ser explorado até à exaustão por um agente que desconte o futuro usando uma taxa exponencial. Para dar um exemplo clássico: o agente que desconta o futuro de uma forma hiperbólica, quando o Inverno chega ao fim, está na disposição de vender o seu casaco de Inverno por um preço baixíssimo, uma vez que já não precisa dele; depois, quando o Inverno seguinte já está à porta, o agente que tem uma taxa de desconto hiperbólica está na disposição de pagar uma quantia avultadíssima pelo mesmo casaco que vendeu na Primavera por uma quantia ridícula. Evidentemente, um agente que

funcione assim não tem grandes hipóteses de sobreviver em competição com agentes que descontem o futuro exponencialmente. Por isso é que a taxa de desconto exponencial é racional enquanto que a hiperbólica é irracional. Ora, Ainslie, um autor indisputavelmente hedonista, defende que a taxa de desconto natural nos seres humanos é hiperbólica e que adoptar uma taxa de desconto exponencial exige esforço cognitivo. Daí que, com frequência, as pessoas – mesmo pessoas relativamente sofisticadas do ponto de vista cognitivo – revertam inconscientemente para a taxa de desconto hiperbólica, que, de alguma forma, teria sido deixada em nós pela evolução natural. A questão que então se coloca é a de saber como poderia explicar-se que a evolução nos tivesse deixado entregues a uma taxa de desconto hiperbólica. Ora, é de facto possível fazer algum sentido evolucionário da presença em nós de uma taxa de desconto deste género, porque é claro que, para que o próximo Inverno possa vir a ser um problema para uma criatura pertencente à biosfera terrestre, ela tem que sobreviver até lá. Seja como for, no desenho de toda esta história, não há qualquer menção a considerações de carácter moral.

CM: Partindo do princípio de que é possível fazer alguma analogia entre o que as pessoas comumente designam como fraqueza de vontade e o problema da acrasia, e que é possível, de algum modo, chegar a um conceito popular de acrasia, e se for manifesto que as pessoas normalmente associam a ideia da fraqueza da vontade a uma fraqueza moral, pensa que isso pode ser relevante para a formulação da questão filosófica da acrasia? Pensa que pode introduzir novos dados para pensar a questão?

AZ: Estou a ver: a sua ideia é tentar encontrar um Mauro-effect! Bem, acho um pouco duvidoso que faça sentido tentar replicar a experiência que levou à descoberta do Knobe-effect no âmbito da discussão em torno do problema da acrasia. Repare: não há qualquer dúvida que existe uma tradição de refinamento do conceito de acrasia na literatura filosófica, refinamento esse que teve precisamente como objectivo libertar o conceito de conotações morais, por oposição ao modo como ele era entendido pelos autores cristãos da Idade Média, período em que a acrasia ou fraqueza da vontade está claramente associada à incapacidade para agir moralmente, à tendência para sucumbir à tentação, etc. Creio que ninguém porá em causa que, de então para cá, houve um grande esforço filosófico no sentido de isolar aquilo que é determinante no conteúdo

do conceito e separá-lo de outros conteúdos que lhe são laterais; e que só no âmbito desse esforço é que se tornou claro que o conceito refere um problema que é independente de questões morais. Neste sentido, creio que é uma novidade teórica relativamente recente que possa defender-se, por exemplo, que é perfeitamente possível que um indivíduo aja de acordo com os preceitos morais aos quais se sente obrigado e que, não obstante, a sua acção seja incontinente. Ora bem, se este é o caso, e supondo que existe de facto um conceito popular de acrasia, o que não é nada claro, então o facto de esse conceito popular, supondo que existe, estar mais perto do modo como os moralistas cristãos da Idade Média se referiam ao problema do que do modo como os filósofos contemporâneos se referem a ele, é absolutamente natural. Se esse não fosse o caso, não teria sido necessário um dispêndio de esforço intelectual de alguma magnitude para separar os diversos elementos que estavam misturados no conceito original. Ora, se essa energia intelectual foi despendida, e se essa separação se operou, então o putativo conceito popular actual deve, obviamente, estar mais perto do conceito que existia no âmbito da discussão teórica antes de esse esforço conceptual ter tido lugar. Qual é então a grande diferença em relação à situação posta em evidência pelo Knobe-effect? Neste último caso, o problema diagnosticado apenas ganha sentido quando considerado à luz de uma visão específica acerca do estatuto epistémico da chamada 'psicologia popular'. Esta visão é aquela que defende que as pessoas têm em si próprias um dispositivo de carácter puramente cognitivo para se relacionarem com o comportamento das outras pessoas e, mais em particular, para o explicarem e fazerem previsões acerca dele, e que esse dispositivo corresponderia *grosso modo* à chamada 'psicologia popular'. Deste ponto de vista, um dos elementos do *pack* cognitivo em questão seria precisamente o conceito de acção intencional. Ora, nas situações experimentais criadas por J. Knobe, a avaliação de intencionalidade que as pessoas fazem muda de acordo com variáveis de carácter moral, as quais têm basicamente que ver com a avaliação do comportamento do sujeito a ser objecto de consideração como sendo digno de louvor ou censura. Segundo Knobe, o facto de tais flutuações ocorrerem mostra que o mecanismo que nos permite relacionarmo-nos com as acções e com os comportamentos sociais dos outros não pode ser um mecanismo de carácter puramente cognitivo, cuja finalidade maior seria produzir explicações e previsões acerca do seu comportamento. Mesmo que *também* sirva esses propósitos, este mecanismo serve primeiramente outros propósitos, sobretudo de carácter avaliativo. Portanto, a conclusão que ele quer tirar é

que a competência em causa não é puramente cognitiva, nem sequer predominantemente cognitiva. Ora, esta conclusão contradiz o ponto de vista mencionado acima, de acordo com o qual a psicologia popular seria uma espécie de teoria científica ingénua. Como disse, se esse fosse o caso, ela teria como seu objectivo descrever, explicar e prever o comportamento dos outros, mesmo que apenas de forma ingénua, e não avaliá-lo moralmente. A tese positiva que Knobe extrai desta discrepância é então a de que o conceito popular de acção intencional é um conceito multi-usos, e não um conceito teórico ou pré-teórico cujo objectivo seria o de fazer ciência espontânea ou pré-ciência. Ora bem, não é de todo claro como é que o conceito de acrasia poderia ser relacionado com a crítica de Knobe a esta visão da psicologia popular, uma vez que este não é um conceito psicológico tão básico quanto os conceitos de crença, desejo, intenção ou acção intencional. A própria dificuldade em fixar um rótulo para o referir atesta o seu carácter pouco ‘popular’. Neste sentido, e como já o disse há pouco, não é preciso criar situações experimentais para verificar que o conceito que é usado pelos filósofos contemporâneos é extremamente depurado. Basta ler os textos dos filósofos de há pouco tempo atrás para ver como o conceito se encontra aí misturado com elementos de outra ordem, nomeadamente de ordem moral. Deste modo, uma conclusão *à la* Knobe sobre o conceito de acrasia seria tudo menos surpreendente.

CM: Na primeira parte da entrevista, disse claramente que razões não conscientes estão em jogo quando uma pessoa age. Mas de um modo geral, na filosofia da acção, os filósofos não tratam das razões não conscientes. Broome diz claramente que as razões não conscientes são também causa da acção, mas o que lhe interessa especificamente são acções cujas razões são conscientes e fruto de deliberação. É possível fazer esta distinção? É possível dizer que existe um tipo de acção que pode ser investigado cujas razões sejam claramente conscientes para o agente?

AZ: Acho estranha a forma como coloca a pergunta porque, precisamente, a grande expansão, a partir dos anos 50, do modelo teórico no qual a teoria da decisão está integrada faz-se em associação com o ponto de vista de acordo com o qual o conceito de racionalidade que está a ser considerado é um conceito dissociado do conceito de

consciência. Uma das teses cruciais deste modelo é a de que os agentes são racionais, ou melhor, o comportamento dos agentes tende a ser racional, independentemente da realidade fenomenológica subjacente – este é, aliás, um princípio básico em Economia. E foi esta a ideia que se generalizou.

CM: Claro, o próprio von Mises diz que o agente é racional independentemente de as razões serem ou não conscientes, mas não sei se, na filosofia da acção, essa forma de pensar é assim tão comum.

AZ: É evidente que existe mais do que uma linha de pensamento na filosofia da acção. Suponho que uma linha de pensamento que venha directamente da tradição fenomenológica talvez proceda dessa forma, mas a linha de pensamento à qual eu estou, de alguma forma, ligado não é essa. É a linha que vem de Ramsey, Davidson, Jeffrey, etc., no contexto da qual é um pressuposto que o conceito de racionalidade do agente é dissociado da ideia tradicional de pensamentos enquanto conteúdos presentes à consciência no momento da deliberação.

CM: Pensando no caso da racionalização em Davidson. Quando o agente racionaliza a acção num sentido específico em que fica clara uma acção incontinente, aceita que aquela racionalização corresponde ao que de facto aconteceu. Mas se eu levar em consideração razões que não são conscientes, não posso dar à racionalização a força que ela tem. É nesse sentido que me parece que também Davidson dá alento à ideia de que as razões relevantes são as razões conscientes.

AZ: Isso é uma questão de exegética davidsoniana... Do meu ponto de vista, não creio que se possa dizer de Davidson que ele só conta como uma razão para a acção aqueles conteúdos que tenham estado presentes na consciência do agente nos momentos anteriores à acção. Aliás, creio até que seria bastante estranho dizer-se isso acerca do pensamento de Davidson em matéria de filosofia da acção. Mas, e independentemente do que Davidson diz ou não diz, parece-me também que uma tal afirmação contradiz a nossa própria experiência fenomenológica dos nossos processos de escolha. Não creio que seja muito frequente surpreendermo-nos a nós próprios em

exercícios de ponderação das diferentes opções, em que as consideramos a todas de forma consciente e explícita. Evidentemente, há situações nas quais fazemos isso, mas não me parece de todo que sejam a maioria.

CM e SC: Como responder a este tipo de objecção, típica de um discurso behaviorista?

“(...) representações mentais de eventos futuros são exemplos de mentalismo, vítimas de todos os problemas que discutimos (...). Onde está essa intenção interna? De que ela é feita? Como essa porta aberta fantasmagórica poderia causar meu comportamento de abrir o trinco? Como uma representação interna do Moby Dick poderia causar sua procura? Isso não é uma explicação; serve apenas para obscurecer os factos relevantes sobre o ambiente: abrir o trinco normalmente leva a uma porta aberta, e procurar uma mercadoria normalmente produz a mercadoria. Esses factos naturais explicam o comportamento sem nenhuma necessidade de introduzir intenções internas.” (W. M. Baum, *Compreendendo o Behaviorismo: comportamento, cultura e evolução*, Porto Alegre: Artmed, 2006, p.10). Essa é uma afirmação comum do behaviorismo contra a ideia de que estados mentais podem ser causalmente relevantes. Queíamos que comentasse.

AZ: Eu não conheço o contexto do qual esta citação foi retirada. Tipicamente, objecções behavioristas deste género são objecções a uma investigação psicológica anterior à Segunda Guerra Mundial, na qual se considerava que os estados e processos mentais deveriam ser considerados de um ponto de vista fenomenológico, de acordo com a definição tradicional, cartesiana, de o que é um estado mental. Basicamente, tratava-se de conteúdos de consciência. Como é sabido, este ponto de vista confronta-se com um problema grave, que tem uma enorme dificuldade em resolver, que é o de perceber como é possível que os estados e processos mentais, concebidos desta forma, possam ter alguma causalidade no mundo empírico. Foi a consciência destas dificuldades que motivou o género de objecções behavioristas que trouxe agora à colação. Ora bem, a partir dos anos 60 do século XX foi possível apresentar uma interpretação do que são estados e processos mentais que é não só diferente da interpretação fenomenológica tradicional mas que tão-pouco se deixa reduzir à ideia

behaviorista de que os estados e processos mentais se deixariam caracterizar completamente pela indicação dos seus *input* e *output* fisiológicos e comportamentais típicos. A nova ontologia de estados e processos mentais que então surgiu considera-os basicamente como estados computacionais do cérebro humano, ou de outros órgãos cognitivos. Neste sentido, representações mentais, crenças, desejos, intenções, etc., todos esses estados ou processos mentais de que habitualmente se fala no discurso psicológico, seriam materialmente realizados no cérebro humano, ou noutros órgãos de processamento cognitivo, de uma forma análoga àquela por meio da qual os procedimentos inferenciais, por exemplo, são realizados pelas máquinas computacionais que nós temos à nossa disposição. A partir do momento em que é introduzida esta forma de entender o problema, creio que a objecção behaviorista que referiu perde muito do seu alcance. Com efeito, se nós considerarmos que as representações mentais são, de facto, fisicamente realizadas no cérebro dos agentes, então deixa de haver qualquer obstáculo à ideia de que elas possam desempenhar um papel causalmente relevante no nosso comportamento. A ideia de que os termos para tais entidades seriam ou vazios ou inertes do ponto de vista explicativo deixa assim de exercer qualquer atracção teórica. Evidentemente, de um ponto de vista ainda mais recente, esta visão computacional da mente pode ser posta em causa como sendo algo inadequada; mas o género de objecções que acabou de mencionar não é o género de objecções que é tipicamente apresentado no âmbito de empreendimentos contemporâneos de crítica à chamada visão computacional da mente. Portanto, se este conjunto de objecções é apresentado contra o ponto de vista fenomenológico acerca da mente, o que há a dizer é que o behaviorismo não é a única alternativa à fenomenologia. O computacionalismo, ou cognitivismo, se quisermos, é também uma alternativa à fenomenologia, e tem sobre o behaviorismo a vantagem de nos permitir continuar a falar em estados e processos mentais como tendo uma existência real e como tendo poderes causais, ao mesmo tempo que tem sobre a fenomenologia a enorme vantagem de, ao defender isso, ter também uma teoria explicativa que permite situar no mundo físico as capacidades causais desses estados e processos. Deste modo, quando, no âmbito da teoria de decisão, interpretada de um ponto de vista cognitivista, se alude a crenças e desejos como causas da acção, a ideia subjacente é que, de facto, elas se encontrariam realizadas fisicamente no cérebro do agente.

CM e SC: Em que sentido pode considerar-se que razões (pares crença-desejo) explicam causalmente acções? Uma explicação causal no pleno sentido da palavra não dependeria do estabelecimento de leis?

AZ: Se nós considerarmos a interpretação behaviorista da teoria da decisão, nos anos 50, por exemplo, isso é verdade. Nesse âmbito, não poderia propriamente afirmar-se que, por seu intermédio, se estaria a explicar uma acção; o que poderia dizer-se seria que se estariam a estabelecer correlações estatisticamente relevantes entre diferentes padrões comportamentais. Mas a partir do momento em que a teoria da decisão é sujeita a uma interpretação de carácter cognitivista, isso já não é o caso. A explicação dada em termos de crenças e desejos remete, de facto, para uma cadeia causal fisicamente realizada; deste ponto de vista, deixa, portanto, de haver qualquer objecção. Evidentemente, pode contra-argumentar-se dizendo que não temos, até ao momento, acesso a uma caracterização independente de tais objectos, isto é, independente da sua postulação teórica. E isso constitui, de facto, uma objecção à cientificidade da teoria. Mas como o conhecimento do cérebro humano é ainda bastante incipiente, essa objecção também se aplica à maior parte das teorias que explicam efeitos comportamentais a partir de causas geradas no cérebro. Continua a não existir conhecimento suficiente que permita, de facto, ter um acesso independente à maioria dessas causas.

CM e SC: Então como entender a ideia davidsoniana de que não existem leis psico-físicas?

AZ: Mas o pensamento de Davidson tem os dois aspectos. Quer dizer, creio que ele teria aceite tudo o que eu disse até ao momento. Só que, além do que eu disse, ele diz outras coisas. Mais em particular, ele acrescentaria ao que eu disse, por um lado, que as leis são entidades que pertencem ao domínio linguístico e, por outro lado, que o discurso psicológico e o discurso físico subjacente não são congruentes. É da conjunção destas duas teses que resulta a conclusão davidsoniana de que seria impossível estabelecer leis psico-físicas. Mas esta é uma tese linguística, se quisermos chamar-lhe assim, a qual, de acordo com o ponto de vista de Davidson, não entra em contradição com a tese, que ele também defende, de que a teoria psicológica, tal como

se encontra expressa no vocabulário psicológico, refere, de facto, os eventos físicos que produzem as cadeias causais que dão origem aos comportamentos e acções de que a teoria psicológica trata. Isto é, do seu ponto de vista, Psicologia e Neurofisiologia falam, de facto, acerca da mesma coisa (física), só que fazem-no através de modos discursivos que obedecem a princípios de organização conceptual diferentes e inconciliáveis. É por isso que não é possível proceder a uma redução inter-teórica que reconduza efectivamente os conceitos e as generalizações de uma aos conceitos e generalizações da outra. Mas isso não significa que não descrevam ambas a mesma realidade material subjacente.

CM: Mas não poderia ser dito que num caso teríamos uma explicação e noutra uma descrição da acção, em termos causais? Essa é uma crítica recorrente – a ideia de que apelar a razões não constitui uma explicação da acção no pleno sentido do termo.

AZ: Isso poderia ser dito, mas não é essa a posição de Davidson. Essa ideia tem que ver, de facto, com a interpretação que se dá ao vocabulário utilizado pela teoria. Mas, depois, pontos de vista como esse ficam com um problema entre mãos, que é o de explicar a utilidade da teoria – se ela é meramente interpretativa, se é apenas uma forma de descrever as coisas, sem apresentar qualquer explicação para as mesmas, qual é a sua utilidade? Mas, na realidade, é, de facto, possível, ainda que com alguma falibilidade, formular um certo número de previsões e explicações usando a teoria com conteúdo psicológico. Ora, se ela fosse meramente interpretativa, se fosse apenas uma forma de falar acerca das coisas, ficaria um pouco difícil dar conta de como é que seria possível explicar o facto de, apesar de tudo, sermos capazes, por seu intermédio, de prever e explicar algumas coisas.

CM e SC: Então qual seriam as alternativas na filosofia da acção a uma teoria causal da acção? Temos apenas uma teoria behaviorista ou há outra hipótese?

AZ: Para responder a esta pergunta é preciso clarificar primeiro o ponto de vista teórico que depois a teoria da decisão, interpretada de um modo cognitivista, veio

formalizar de um modo bastante preciso. Creio que é necessário caracterizar esse ponto de vista recorrendo a dois aspectos e não apenas a um. Um desses aspectos é, de facto, o da causalidade – trata-se de um ponto de vista que pretende indicar as causas das acções; mas há também um outro aspecto que convém não esquecer, que é o do individualismo. De acordo com este ponto de vista, então, a causa da acção encontra-se no interior do agente individual. Ora, estes dois aspectos não têm que estar obrigatoriamente associados. Neste sentido, existem pontos de vista que são também causalistas mas que não são individualistas. É o caso, por exemplo, do ponto de vista estruturalista, o qual não aceita que as causas reais das acções, sobretudo daquelas que são socialmente relevantes, estejam no interior do agente, mas que, no entanto, defende que existe uma explicação causal para essas acções. Para os estruturalistas, os elementos das cadeias causais relevantes são entidades como, por exemplo, as estruturas sociais, e são estas que obrigam, por assim dizer, o agente a agir de uma determinada maneira e não de outra. Este ponto de vista é um ponto de vista alternativo àquele do qual temos estado a falar, mas também é causalista. Eu creio que um tal ponto de vista se depara com escolhos teóricos dificilmente superáveis, porque o modo como o conceito de causalidade opera no interior das estruturas que se pretende ter identificado e os critérios de identidade para individualizar estas últimas são tudo menos claros, mas essa já é outra questão. Para responder à sua pergunta, o que é importante salientar é que não podemos colocar a questão das opções teóricas disponíveis apenas em termos da oposição teorias causalistas *versus* teorias não causalistas. Há diferentes tipos de teorias causalistas. E todas elas pretendem ser explicativas, no sentido em que se fala da ciência como um empreendimento explicativo. Do mesmo modo, também há diferentes tipos de teorias não-causalistas. De facto, no seio deste outro ponto de vista, pode igualmente distinguir-se entre teorias individualistas, de uma forma geral associadas à tradição fenomenológica, e teorias holistas. Posso apresentar como exemplo destas últimas as tentativas de formalizar teoricamente algumas das ideias expressas pelo segundo Wittgenstein a propósito de jogos, regras, papéis, etc. A ideia aqui é a de que, não sendo de carácter causal, a influência da regra sobre o agente é a instância adequada para a compreensão da acção deste último no âmbito das suas interacções com os outros agentes. Isto é, estas acções devem ser compreendidas como lances de jogos determinados por regras impessoais e não como resultado da operação das motivações

particulares dos agentes, concebidas como sendo independentes da trama discursiva que transforma comportamentos em acções.

CM e SC: Também a questão da previsão. Uma das grandes vantagens da teoria da escolha racional é que permite a previsão.

AZ: Se uma explicação é de carácter causal, ela introduz, evidentemente, a possibilidade da previsão; se conhecermos a cadeia causal, podemos prever os seus efeitos.

CM: Essas teorias não causais não vão contra a ideia básica de uma *folk psychology*, a ideia de que as pessoas conseguem prever as acções dos outros, de que conseguimos interagir suficientemente bem tendo como base as crenças e desejos que imputamos aos outros?

AZ: Isso depende de como definirmos essa ideia de *folk psychology* ou psicologia popular. Este conceito encerra uma grande ambiguidade porque, por um lado, nos remete para o adjectivo ‘popular’, o qual sugere um entendimento ingénuo do mundo mental e, por outro lado, remete-nos para o substantivo ‘psicologia’, o qual sugere um empreendimento conceptual de carácter teórico ou explicativo. Neste sentido, quando usamos este termo convém tornar claro se estamos a falar do modo como a generalidade das pessoas se relaciona com o comportamento e a vida mental seus e dos outros, ou se estamos a falar de um género particular de teoria psicológica que, mesmo que relativamente elementar, não deixa de ser uma teoria. Ou, ainda, se defendemos o ponto de vista filosoficamente substantivo, de acordo com o qual a generalidade das pessoas se relaciona com o comportamento e a vida mental seus e dos outros de um modo essencialmente teórico ou explicativo. Só se pressupusermos que este último ponto de vista filosófico substantivo é correcto, é que se levantariam as dificuldades que mencionas aos pontos de vista não-causalistas. Mas reparem que é também este ponto de vista filosófico substantivo que J. Knobe considera ter sido decisivamente posto em causa pela sua descoberta do Knobe-effect...

CM: Referindo-nos ao primeiro sentido – como explicar que as pessoas conseguem fazer boas previsões em relação ao comportamento dos outros? A tese de que razões não são causas não inviabiliza essa ideia?

AZ: Talvez. Mas tanto num caso como no outro – isto é, tanto no caso do interpretativismo de raiz fenomenológica como no caso do interpretativismo holista – aquilo de que estamos a falar é acerca de teorias filosóficas relativamente sofisticadas e, portanto, o facto de elas poderem contradizer o entendimento popular da questão acerca da qual se debruçam não é, em si mesmo, uma objecção a essas teorias. Não seria a primeira vez que uma teorização ‘popular’ se viria a revelar estar errada. Mas convém também salientar aqui que, se há correntes interpretativistas de raiz fenomenológica, também existe um ponto de vista de raiz fenomenológica que é causalista – nomeadamente, o ponto de vista cartesiano, de que já falei há bocado. Aliás, o ponto de vista cartesiano é, na realidade, a forma mais conseguida de dar uma profundidade teórica à ideia de uma psicologia popular, tal como ela se encontra subjacente à sua pergunta anterior. Mas, como disse há pouco também, o cartesianismo psicológico soçobra quando se defronta com a questão de saber como é que é possível compreender que entidades não físicas possam ter eficácia causal num mundo físico. Mas repare que a própria colocação desta questão já supõe uma sofisticação conceptual não muito ‘popular’. Neste sentido, é natural que perspectivas que surgiram a partir da necessidade de encontrar uma saída para o impasse teórico e filosófico criado pelo cartesianismo psicológico tenham tido que se distanciar bastante do tal ponto de vista ‘popular’ que criou esse mesmo impasse.

CM e SC: Qual é a sua opinião geral acerca da filosofia experimental?

AZ: Diferentes coisas podem ser ditas acerca da filosofia experimental. Se pensarmos na declaração de princípios que se encontra no manifesto, aquilo de que lá se fala é de um fenómeno que é recorrente na história da filosofia. Aquelas objecções à filosofia percebida como dominante na academia americana são do género das objecções que encontramos na crítica que o jovem Marx fez à análise kantiana da prova ontológica da existência de Deus. Segundo Marx, aquilo que interessaria seria compreender porque é que as pessoas têm o conceito e não demonstrar que, do ponto de vista

teórico, não haveria uma forma racional de demonstrar a existência do objecto referido por esse conceito. Portanto, as críticas do manifesto não são propriamente novas. Eu creio que a crítica do jovem Marx a Kant é bastante injusta; a análise crítica que Kant faz na *Crítica da Razão Pura* às diferentes provas da existência de Deus é não só intrinsecamente pertinente como teve sem qualquer espécie de dúvida um impacto extremamente importante na evolução da filosofia e do pensamento europeu subsequentes. Sendo a crítica injusta, isso não significa evidentemente que a questão levantada pelo jovem Marx não seja em si mesma relevante. É importante não apenas discutir a validade conceptual intrínseca dos argumentos como também compreender as razões pelas quais, independentemente da validade conceptual intrínseca dos argumentos, as pessoas aderem a certos conceitos e ideias, e deixam a sua vida ser conduzida por elas. Sem dúvida que também é importante compreender isso, o que, aliás, tem vindo a ser feito. A fundação de empreendimentos teóricos bem organizados, como, por exemplo, a antropologia cultural, teve precisamente como finalidade criar uma área disciplinar na qual se procura alcançar o género de compreensão de que o jovem Marx falava. O que, por sua vez, não quer dizer que, no interior da filosofia, não seja também relevante focar esses aspectos. Mas são os próprios fundadores do movimento da filosofia experimental quem defende que o ponto de vista que eles apresentam não deve ser visto como alternativo ao ponto de vista que eles tomam por dominante (nos EUA), mas que tem um carácter apenas complementar. Basicamente, o que Knobe e os seus colegas defendem é a criação de um espaço institucional para eles, o que me parece ser perfeitamente legítimo. Todavia, parece-me ser importante chamar aqui a atenção para um aspecto. É claro que eu acho que é intrinsecamente relevante conhecer quais são as intuições que as pessoas de facto têm em associação com os conceitos que usam, em vez de, pura e simplesmente, como alguns filósofos têm uma certa tendência para fazer, se dizer que “a intuição obviamente é esta”. Mas convém sublinhar, mais uma vez levando em conta o que os autores do manifesto escrevem, que o objectivo que eles pretendem alcançar não se esgota nessa investigação. Esse objectivo vai mais além e consiste na tentativa de compreender o modo como a mente humana funciona, o que, evidentemente, é mais que louvável e relevante. Convém chamar a atenção para este aspecto porque, mesmo supondo que os filósofos x, y ou z dizem que a intuição associada ao uso de um termo é obviamente w quando, de facto, para a generalidade das pessoas não é, é preciso deixar claro que a razão pela qual uma certa intuição é

aceitável ou razoável não é o facto de ser mais difundida ou menos difundida – não é isso que deve estar em causa. Portanto, os entusiastas deste movimento devem ter em mente, o que nem sempre parece ser o caso, que o objectivo último das investigações efectuadas na filosofia experimental para determinar quais são, na realidade, as intuições que as pessoas de facto têm não se esgota nessa determinação, mas que aquilo que, em última análise, se pretende é encontrar verdades empíricas acerca do modo como a mente humana está organizada. Não creio que seja uma objecção a este trabalho apontar-se para o facto de que esse é o objectivo de um ramo estabelecido da ciência – a Psicologia Cognitiva. Creio que não devem barrar-se caminhos para a descoberta científica com base em justificações de carácter disciplinar ou corporativo. Os programas de investigação justificam-se pelos frutos que dão e não pela sua inserção departamental. Também há filósofos que desenvolvem análises matemáticas rebuscadas no âmbito da interpretação bayesiana da causalidade, por exemplo, apesar de haver matemáticos que fazem a mesma coisa em departamentos de matemática ou estatística. A única coisa que interessa é a qualidade do trabalho.

Da filosofia antiga à filosofia contemporânea da acção – conversa com Ricardo Santos

Ricardo Santos Sofia Miguens e Susana Cadilha

Ricardo Santos é Professor no Departamento de Filosofia da Universidade de Évora e investigador no Instituto de Filosofia da Linguagem (FCSH – UNL). Ensina nas áreas da Filosofia Antiga, Lógica e Filosofia da Linguagem. É autor de *A verdade de um ponto de vista lógico-semântico* (Lisboa, FCT/FCG,) bem como de numerosos artigos. Traduziu e comentou as *Categorias* de Aristóteles (Porto Editora 1995). É actualmente Presidente da Sociedade Portuguesa de Filosofia.

Sofia Miguens (SM) e Susana Cadilha (SC) – O Ricardo chegou à filosofia contemporânea da acção a partir da filosofia antiga – será que nos pode explicar tal trajecto? Ele é interessante para se conceber o que está em causa na filosofia da acção de uma forma mais ampla, relativamente aos debates da filosofia analítica contemporânea. A acrasia foi um importante tema de articulação para si, não foi?

Ricardo Santos (RS) - Foi esse o trajecto, é verdade. Uma parte da explicação prende-se com as contingências da vida. Em 1997, quando fui seleccionado no primeiro concurso que o Instituto de Filosofia da Linguagem (da Universidade Nova de Lisboa) lançou para contratar investigadores – uma coisa inédita na Filosofia em Portugal, naquela altura –, eu tinha acabado de escrever uma tese de mestrado sobre o problema da acrasia na ética de Aristóteles, uma tese muito centrada na interpretação do famoso capítulo VII.3 da *Ética a Nicómaco*, onde Aristóteles, perante a tese socrática de que não existe acrasia, mas somente ignorância, responde de um modo que os comentadores achavam tão surpreendente que diziam ser inconsistente com tudo o que resto que ele escreveu acerca do assunto (Cook Wilson, num livro de 1879, defendeu que Aristóteles não é o autor do capítulo; e David Ross, em 1923, disse que, embora o capítulo seja de Aristóteles, a posição defendida não é dele). Além disso, também já tinha publicado a minha tradução e comentário das *Categorias* de Aristóteles (Porto Editora, 1995). Mas, na preparação da tese, tive os meus primeiros contactos com essa nova disciplina de Filosofia da Acção que surgira no âmbito da

filosofia analítica. Nos trabalhos sobre Aristóteles multiplicavam-se as referências ao famoso artigo de Donald Davidson, “How is Weakness of the Will Possible?” (1970). Este artigo estava publicado na colectânea de Davidson, *Essays on Actions and Events* (Oxford, 1980), e eu consegui comprar o livro. Lembro-me que, na primeira leitura do artigo, houve uma frase que me intrigou: é uma nota em que ele censura os outros filósofos por não reconhecerem que a acrasia é um problema de filosofia da acção e não um problema de filosofia moral. Eu nunca tinha ouvido falar desta distinção e por isso fiquei perplexo. Mas a ideia entendia-se bem: na sua forma mais geral, o problema da acrasia é moralmente neutro, pois diz respeito a um desalinhamento entre o que o agente julga que seria melhor fazer numa dada situação e aquilo que ele faz realmente, de modo voluntário ou intencional – e pouca importa se aquele juízo é moralmente correcto ou não. Então, as minhas leituras nessa época iam seguindo uma via dupla: lia coisas que me ajudavam a compreender como pensava Aristóteles e também coisas que me ajudavam a compreender, de um modo mais independente e com recursos que excediam o que era dado por Aristóteles, o próprio fenómeno do agir intencional e do agir contra o nosso próprio melhor juízo. Na filosofia anglo-americana há um trânsito fácil entre filosofia antiga e filosofia analítica contemporânea e, por isso, fui naturalmente levado a ler autores como Anscombe, Hare, MacIntyre, Nussbaum, David Pears, David Wiggins, Alfred Mele. Mas é justo que mencione um livro em particular, que fez uma grande diferença: *Aristotle’s Philosophy of Action* de David Charles (Oxford, 1984). Era sem dúvida o livro mais difícil de toda a bibliografia que tinha, exibindo um grau de sofisticação conceptual muito superior ao habitual. Nessa obra, Charles discute longamente a acrasia (e defende uma posição que se aproxima bastante da que me parece mais correcta), mas enquadra o tratamento desse tema numa teoria geral da acção, que vai desde a ontologia até à discussão da racionalidade, do conhecimento prático, da causalidade e dos diferentes níveis de explicação da acção. E o mais surpreendente é a forma como Charles põe as ideias aristotélicas acerca destes assuntos em confronto directo com as ideias de autores contemporâneos como G. H. von Wright, Davidson ou Alvin Goldman. Entender este livro pareceu-me um desafio que realmente valia a pena.

Bem, então, tinha acabado há meses a tese de mestrado quando entrei para o Instituto de Filosofia da Linguagem, para me dedicar a tempo inteiro à investigação. Tinha de me integrar num projecto que estava então a arrancar no Instituto, onde iriam ser

analisados os contributos de seis filósofos contemporâneos – e um deles era Davidson (os restantes eram Wittgenstein, Quine, Kripke, Searle e Habermas). Com a preparação que trazia, é claro que me ofereci para tratar da filosofia da acção de Davidson. E, para não perder a embalagem, comecei exactamente pela acrasia. Mas, depois, acabei por colaborar também noutras áreas – como a teoria da decisão, a filosofia da mente e a semântica das línguas naturais – e dediquei-me também às obras de Quine e de Kripke. Aliás, quatro anos mais tarde estava a defender uma tese de doutoramento sobre o significado filosófico da teoria da verdade de Tarski.

SM e SC - Tem trabalhado bastante sobre Donald Davidson, focando questões de filosofia da linguagem e teorias da verdade, e não exclusivamente a filosofia da acção. Mas pensando em Davidson enquanto filósofo da acção, como caracterizaria o que há de específico no seu pensamento relativamente ao debate sobre a acrasia que vem da filosofia antiga? Diria que a inovação, ou a especificidade, da abordagem davidsoniana tem a ver com a forma de conceber desejo e juízos? Está satisfeito com a solução proposta por Davidson, apesar de algumas falhas que se pode apontar-lhe?

RS - A respeito da acrasia, a primeira contribuição importante de Davidson foi na clarificação do próprio problema. Ele formulou a seguinte definição de acção acrática: Ao fazer x , uma pessoa age acriticamente se e somente se (a) faz x intencionalmente, (b) acredita que existe uma acção alternativa y que pode fazer e (c) julga que, consideradas todas as coisas, seria melhor fazer y do que fazer x . Portanto, num caso de acrasia, o agente acredita que tem de fazer uma escolha entre dois géneros de acção: mentir ou dizer a verdade, pagar o que deve ou comprar um livro com esse dinheiro, etc. É natural que o agente tenha razões que favorecem que faça uma coisa e também razões que favorecem que faça a outra. Por isso, ele tem de confrontar essas razões umas com as outras, avaliar o seu peso relativo, etc., até chegar a uma conclusão, aquilo a que chamamos um “melhor juízo”, um juízo do género “Tendo em conta todas as considerações relevantes, é melhor que eu faça y ”. Num caso de acrasia, o agente chega a um juízo destes e, no entanto, faz antes x . E o aspecto crucial é este: a sua acção é livre e intencional, nada nem ninguém o força a agir desse modo. Como diz Aristóteles, o acrático age “com conhecimento daquilo que faz e do fim

para que o faz” (EN 1152a16). Ora, o problema tem precisamente origem no carácter intencional da acção acrática. Isso obriga a que a explicação dessa acção seja consistente com a explicação mais geral da acção intencional. O que é uma acção intencional? Parece que todos temos uma tendência para pensar no agir intencional de uma maneira que, se estivesse correcta, teria aparentemente como consequência que a acrasia não é possível – precisamente a consequência que Sócrates, para surpresa de muita gente, tirou. Davidson formulou um par de princípios que, supostamente, condensariam o que é essencial nessa maneira intuitiva de conceber a acção intencional. Sem entrar nos pormenores, digamos que a ideia é que, quando alguém faz algo intencionalmente, essa pessoa faz aquilo que quer e que julga ser bom à luz dos objectivos que tem e dos princípios que aceita. Por isso, quando a pessoa enfrenta uma escolha entre *x* e *y* e faz uma delas intencionalmente, dir-se-ia que ela faz aquilo que mais quer e que julga ser melhor. Dito de outro modo: parece constitutivo de uma acção intencional ser uma acção internamente racional, quer dizer, uma acção que é ‘racionalizada’ (ou racionalmente explicável) pelos estados mentais – como as crenças e os desejos – do próprio agente. Daí que haja dificuldade em integrar as acções acráticas, na medida em que estas são, ao mesmo tempo, intencionais e (internamente) irracionais.

A solução proposta por Davidson está vinculada à sua recuperação da chamada “teoria causal da acção” – a tese de que “as razões são causas”, ou seja, que quando alguém faz algo intencionalmente, as razões porque o faz são aquilo que causa a acção. Esta tese tradicional (aceite por Aristóteles e muitos outros) tinha sido negada por Wittgenstein e por alguns dos seus alunos, e Davidson começou a sua carreira filosófica (bastante tardiamente) defendendo a teoria causal. Numa acção intencional, o agente tem uma razão para fazer o que faz e fá-lo por causa dessa razão – e Davidson considera que esta relação entre razão e acção é simultaneamente uma relação de explicação racional e uma relação causal. A razão que explica e causa a acção pode incluir um “melhor juízo” (um juízo que resume a ponderação de todas as considerações relevantes), mas pode também não incluir tanto: muitas vezes, agimos com base apenas em juízos parciais (de desejabilidade *prima facie*), juízos que dizem simplesmente que fazer tal coisa seria bom devido a uma certa consideração. Precisamente, a acção acrática tem na sua base um desses juízos parciais (e, por isso, ela é explicável de modo semelhante a outras acções intencionais); mas, ao mesmo

tempo, ela é contrária ao “melhor juízo” do agente. Davidson atribui ao acrático dois juízos em conflito. Porém, devido à sua diferente natureza, não há estritamente contradição entre eles: o juízo que diz que é melhor fazer y tem em conta todas as considerações relevantes, enquanto o juízo que diz que é melhor fazer x tem em conta apenas alguma ou algumas considerações isoladas. O que temos de explicar, na acrasia, não é porque é que o agente fez x – isso é fácil, e faz-se indicando o que podemos chamar a “razão acrática”. O que precisa de ser explicado é porque é que ele não seguiu o seu melhor juízo. Se não fosse para serem seguidos, os melhores juízos (e o processo deliberativo que os gera) não serviriam para nada. Davidson diz que a racionalidade exige de nós que sigamos os nossos melhores juízos – e é isso que o acrático não faz. Porquê? Por causa da tal razão acrática. A solução de Davidson concentra-se exactamente neste ponto, da relação entre a razão acrática e a violação da exigência racional de agir segundo o nosso melhor juízo: ele diz que, neste caso, a relação é causal, mas não racional. A razão acrática causa duas coisas – a acção realizada (por exemplo, comprar o livro em vez de pagar a dívida) e a desconsideração do princípio racional –, mas só explica ou torna compreensível a primeira delas. Pois o interesse em ler aquele livro, por exemplo, não é uma razão para não seguirmos o nosso melhor juízo, ainda que naquele caso tenha provocado exactamente esse efeito. Para explicar esta aparente anomalia de haver uma causa mental que não fornece uma razão para aquilo que causa, Davidson diz então que a mente do acrático está dividida em duas partes que funcionam como se fossem duas mentes distintas.

De facto, esta solução proposta por Davidson não me satisfaz. Fico com a impressão de que o problema não foi realmente resolvido, que o fenómeno não está ainda explicado. A descrição que Davidson faz da experiência mental do acrático parece-me pouco realista, acho que ele – tal como Sócrates – exagera a racionalidade do agente. Isso é desde logo visível no facto de ele aceitar um princípio que diz o seguinte: se um agente julga que seria melhor fazer x do que fazer y , então esse agente quer fazer x mais do que quer fazer y . Ora, isto poderá estar correcto para um agente racional ideal, mas nós não somos assim: por vezes, julgamos que seria melhor fazer x , mas queremos mais fazer y . Ou, então, há diversas formas de querer – umas mais racionalizadas e outras mais elementares – e o princípio poderá ser correcto para umas, mas não para outras. Julgo também que Davidson tem uma concepção

inflacionada da acção intencional – como acção explicável por um conjunto de atitudes proposicionais (desejos, crenças, intenções) à luz das quais ela surge como razoável ou como ‘a coisa que tem de ser feita para quem tem tais crenças e desejos’. Essa é uma descrição plausível da acção intencional de um agente adulto, psicologicamente maduro e cujas emoções estão bem integradas com o seu pensamento prático. Mas é plausível supor que haja já acção intencional em níveis inferiores de desenvolvimento. O problema, em Davidson, prende-se com o holismo: ele defende o holismo do mental, tal como, a respeito da linguagem, defende o holismo semântico (segundo o qual o significado de uma frase depende das suas relações com as outras frases da mesma linguagem). E o holismo a respeito de uma capacidade complexa concilia-se mal com a compreensão da aquisição progressiva das sub-capacidades que a constituem. No caso da linguagem isso é talvez mais claro: as concepções holistas (sobretudo se forem radicais) têm dificuldade em entender a aquisição progressiva de linguagem efectuada por uma criança, estando mais habilitadas para descrever o produto final dessa aprendizagem. Davidson recusa-se a atribuir intencionalidade ao comportamento de quaisquer animais (não-humanos) e julgo que teria dificuldade em dizer onde é que a intencionalidade surge no desenvolvimento de uma criança. Ora, é possível que a acção acrática típica seja intencional num sentido mínimo de acção intencional – um sentido que requer apenas que o agente sabe o que faz e faz o que quer, mas em que este ‘querer’ pode não estar estruturado em atitudes proposicionais completas. Por fim, não descarto a ideia de mente dividida – a qual começou a ser usada para entender a acrasia por Platão e Aristóteles –, mas critico o facto de Davidson encarar as duas estruturas como sendo, cada uma delas, perfeitamente racional: como se a irracionalidade resultasse apenas de elas serem mentes do mesmo agente.

SM e SC - Considera-se um crente ou um céptico, no que diz respeito à possibilidade de existirem acções acrálicas? Vamos considerar Aristóteles e Davidson: sendo ambos crentes quanto a essa possibilidade, o Ricardo diria que está mais próximo da concepção aristotélica (de acordo com a qual nos agentes acrálicos há como que uma luta entre razão e desejo, entre o que os agentes sabem/conhecem e o que eles desejam, sendo que é o que desejam que leva vantagem), ou diria antes que está mais próximo da solução

davidsoniana (de acordo com a qual o que está patente nos agentes acráticos é um tipo de inconsistência interna que se traduz no facto de as “razões acráticas”, chamemos-lhes assim, suplantarem o princípio da continência, apesar de não constituírem razão suficiente para tal)?

RS – É evidente que existem acções acráticas. A acrasia é um fenómeno real, infelizmente. Os prejuízos que daí nos advêm só são recompensados nalguma medida, pelo menos para nós filósofos, porque ganhamos um problema teórico que tem mostrado ao longo do tempo como é difícil de compreender e rico nas suas implicações. Aristóteles e Davidson são, de facto, dois pensadores fundamentais a respeito da acrasia. Já disse antes o que acho da explicação proposta por Davidson. Quanto a Aristóteles, ele escreveu muito acerca da acrasia – e também acerca da qualidade contrária, a chamada “continência” (o termo grego é *enkrateia*), que é a capacidade de uma pessoa se controlar a si própria ou de dominar os seus apetites excessivos – e não é possível concordar com tudo. Há, sem dúvida, coisas muito interessantes no que ele escreveu acerca do assunto, como por exemplo quando aponta a percepção do tempo como um elemento necessário na acrasia ou quando afirma que a continência não é propriamente uma virtude, pois o indivíduo que se controla *tem* apetites excessivos, o que significa que as suas emoções e os seus juízos de valor não estão em harmonia. Mas, por outro lado, devo dizer que me parece que Aristóteles não deu muita importância ao problema que hoje, no seguimento de Davidson, nós mais discutimos, que é o problema da *possibilidade* da acção acrática. Para ele, a existência de acrasia era uma evidência básica, que havia que respeitar. A heterodoxia ou o paradoxo socrático – de que “ninguém age mal voluntariamente, mas apenas por ignorância” – não o impressionou muito. Precisamente, na minha opinião, uma das dificuldades principais na interpretação do capítulo *EN VII 3* tem origem no facto de muitos comentadores terem tentado ler o texto como se fosse uma resposta ao problema da possibilidade da acrasia. Quando, na realidade, o que Aristóteles aí discute é a questão do saber do acrático. O senso comum diz que o acrático *sabe* que o que faz é mau; mas, no diálogo platónico *Protágoras*, Sócrates argumenta que essa opinião implica uma desvalorização inaceitável do poder do conhecimento (*epistêmê*), quer dizer, implica aceitar que pode dar-se o caso de o conhecimento (que é conhecimento do bem e do mal) estar na alma de uma pessoa,

juntamente com outros estados mentais, e não ser ele “o senhor da casa”. Foi a este argumento que Aristóteles foi sensível (talvez porque ele próprio defende uma posição semelhante a respeito da *phronêsis*, da inteligência ou sabedoria prática: quem é *phronimos* não cai na acrasia) e ao problema por ele colocado, que é o seguinte: se o acrático sabe que age mal, mas não pode ter conhecimento, então que género de “saber”, ou que género de juízo correcto, é esse que ele tem e contrariamente ao qual age? Este é que é o problema que Aristóteles discute em *EN* VII 3, e não o problema da possibilidade da acrasia. Mas durante muito tempo tentaram forçar o texto a responder ao problema que nos interessa a nós, e isso acontecia também na própria tradução. E ainda vai acontecendo, aliás. Por exemplo, nas últimas linhas do capítulo, Aristóteles dá por terminada a discussão, considerando ter mostrado “em que sentido é possível saber (*pôs eidota endekhetai*) quando se age acriticamente”. Pela recente tradução portuguesa da *Ética a Nicómaco*, no entanto, o que Aristóteles diria ao fechar o capítulo é que mostrou “como é possível perder o domínio de si, sabendo”.

Ora, colocada a questão a respeito do saber ou não saber do acrático, o que Aristóteles então faz é explorar distinções relevantes – entre conhecimento actual e potencial, entre conhecimento universal e particular – que lhe permitirão, no final, reconciliar a opinião comum (“o acrático sabe”) com a tese socrática (“o acrático não sabe”). Este dado é sólido, portanto: Aristóteles atribui ao agente acrático uma falha cognitiva, uma forma de “ignorância”. Esta falha afecta um conhecimento que o acrático *tem*, mas que, no momento em que age, *não usa*, nem poderia usar. Esta incapacidade de usar um conhecimento que se tem é temporária e é causada pelos apetites ou outras emoções fortes que o acrático experimenta em si; uma vez satisfeito o desejo, ele recupera esse conhecimento. Além disso, Aristóteles considera que o acrático até pode, ao mesmo tempo que erra, dizer coisas muito sensatas acerca do erro que comete e das razões para não o fazer; nisso ele é comparável a um indivíduo que, apesar de estar a dormir ou embriagado, enuncia toda a cadeia de uma demonstração de algum teorema ou comparável a um estudante que faz exercícios de verbalização para decorar uma matéria que ainda não entendeu: as palavras são sábias, mas não o indivíduo que as profere. Por outro lado, o conhecimento relevante é complexo e inclui itens universais e itens particulares, bem como itens que são “acerca do agente” e outros que são “acerca do objecto (da acção)”, e a ignorância acrática não tem de os

afectar todos do mesmo modo. Aristóteles pensa que, quando se encontra emocionalmente afectado, o acrático sofre uma falha cognitiva temporária que tem por objecto “a última premissa” (do seu raciocínio prático), que é um juízo de tipo perceptivo a respeito de alguma coisa particular.

Voltando à vossa pergunta: eu adiro à concepção aristotélica precisamente na medida em que ela representa a acrasia como envolvendo um conflito entre razão e desejo. Mas muitos comentadores julgaram que Aristóteles, ao dizer que o acrático ignora a última premissa, estaria a ser infiel a essa concepção. Pois, se assim fosse, o acrático não chegaria sequer a concluir o “bom silogismo” e não haveria conflito genuíno, uma vez que o elemento racional se encontraria anulado, deixando assim o terreno livre para que o desejo assumira o comando. Isto é uma má interpretação. Em geral, os comentadores têm dificuldade em entender porque é que Aristóteles teve necessidade de aceitar que há ignorância na acção acrática. Se juntar a isto o facto de eles lerem o texto à procura de uma resposta para o problema da possibilidade da acrasia, é fácil ver como se gera o equívoco: a ignorância seria a explicação aristotélica para esta possibilidade, quer dizer, tal como Sócrates, Aristóteles julgaria que a acrasia só é possível porque, no momento em que erra, o acrático não sabe realmente, no sentido mais completo do termo, que o que faz é errado; se ele o soubesse, não o faria. Deste modo, a posição de Aristóteles permaneceria essencialmente intelectualista, salvaguardando a supremacia do conhecimento: na acrasia, o conhecimento não é realmente vencido pelo prazer, pois o conhecimento simplesmente não está lá de forma completa. Mas esta tese é insustentável, pois Aristóteles afirma claramente que a ignorância acrática é *causada pelos apetites excessivos do agente*.

Para se entender porque é que Aristóteles viu necessidade de postular ignorância na acrasia, é importante considerar que ele pensa que na acrasia há um erro de percepção. O acrático percebe mal as coisas, não as vê como elas são, sofre uma espécie de ilusão. A ilusão do acrático consiste em ver como boas coisas que são apenas agradáveis, que lhe dão prazer. Na medida em que as vê como boas, ele deseja-as. Mas ele sabe que elas não são boas, mesmo quando lhe parecem tais. Aristóteles descreve este género de experiência em diversos textos: uma coisa parece-nos de um certo modo ao mesmo tempo que sabemos que não é desse modo. Enquanto este saber se mantiver activo, as falsas aparências não nos enganam: apesar

de estarmos sujeitos a elas, não as aceitamos, não acreditamos nelas. Numa passagem do *De insomniis* (*Sobre os sonhos*), é dado o exemplo de alguém que vê dois objectos, mas não acredita que sejam dois, pois sabe que o seu olho está a ser pressionado por um dedo. Outro exemplo é o de alguém que está a sonhar com Corisco, mas, ao mesmo tempo, há algo dentro dele que lhe diz que é apenas um sonho e, por isso, ele não acredita no sonho que está a ter. A experiência do acrático é semelhante a estas – e parece-me natural descrevê-la como um conflito. A função da ignorância na acrasia é então clara: é pela incapacitação do conhecimento – quer dizer: da percepção correcta das coisas como elas são – que as falsas aparências ganham credibilidade, levando o acrático a aceitá-las e a acreditar por um tempo que aquelas coisas são realmente boas, tornando assim o desejo por elas mais forte e decisivo. É desta concepção aristotélica que eu diria que estou próximo.

SM e SC– É Aristóteles quem primeiro apresenta a concepção de ‘silogismo prático’ de acordo com a qual há uma relação dedutiva entre desejo e crença, por um lado, e a acção, por outro – sendo a própria acção a conclusão do silogismo ou do raciocínio. Mas, se é assim, como pode haver então espaço para a acção acrática em Aristóteles? Devemos pensar numa outra definição para acção acrática que não seja ‘aquela acção que não está de acordo com a crença do agente acerca de qual é a melhor forma de agir’?

RS – Não acho que a ideia segundo a qual a conclusão do silogismo prático é uma acção e não uma proposição seja uma tese aristotélica. Trata-se de uma questão controversa nos estudos aristotélicos, pois há evidência textual em apoio de ambos os lados. No capítulo 7 do *De Motu Animalium* podemos ler: “Que a acção é a conclusão, isso é claro” (701a22-23). Mas também existem outros textos onde parece evidente que Aristóteles reconhece que tirar a conclusão que se segue das premissas e agir de acordo com essa conclusão são duas actividades que, nalguns casos, podem até ser bastante distantes no tempo; ou em que reconhece que o agente pode ser impedido de agir por factores externos, sem que isso o impeça de tirar a conclusão sobre o que deveria fazer. Uma explicação plausível é que, na passagem do *De Motu*, Aristóteles use “conclusão” para referir aquilo – a própria coisa – de que a conclusão

fala. O dicionário de Liddell e Scott regista uma passagem dos *Primeiros Analíticos* (em 53a17) onde o termo é também assim usado.

O silogismo prático é um modelo de como funciona o raciocínio prático de um agente. A premissa maior diz aquilo que o agente deseja, a premissa menor diz aquilo em que ele acredita e as duas juntas implicam uma conclusão que especifica uma certa acção como sendo boa ou devendo ser realizada. Se o agente aceita a inferência e tem realmente aquele desejo e aquela crença, então ele aceitará necessariamente a conclusão e agirá de acordo com ela, se tal for possível e nada o impedir. Mas Aristóteles não considera que todas as acções intencionais (ou voluntárias) tenham de ser precedidas por um silogismo prático. Isso é evidente no caso de muitos animais que, apesar de não raciocinarem, agem intencionalmente. E, a respeito da acrasia, também não me parece que ser precedida por um silogismo prático seja um elemento necessário de toda a acção acrática. Aristóteles distingue tipos diferentes de agente acrático. Por exemplo, distingue o fraco e o impetuoso: enquanto o fraco delibera, mas depois as suas emoções fazem-no abandonar o resultado da sua deliberação, o impetuoso é alguém que é levado rapidamente pelos seus apetites intensos, não esperando sequer para consultar a sua razão e deliberar. E também não tem de haver um silogismo a recomendar a acção que o acrático acaba por realizar. Muitas vezes, a simples visão de algo agradável é suficiente para activar um desejo forte que conduz rapidamente à acção. Em todos estes casos, mantém-se o traço essencial da acrasia, que é o facto de o agente saber que o que faz é errado, apesar de o fazer voluntariamente.

SM e SC - A motivação para a acção é um tópico que aproxima a filosofia da acção de certos debates em filosofia moral. Por exemplo, o Ricardo parece inclinar-se para a ideia de que é possível um agente julgar que tem razões para fazer X ou que X é melhor do que Y e, ainda assim, não estar motivado a fazer X ou desejar mais fazer Y do que fazer X (é isto, aliás, o que torna possível as acções acráticas). Ora, essa ideia é a negação da tese do internalismo motivacional, de acordo com a qual se o agente julga que X é melhor/tem razões para fazer X/deve fazer X, então necessariamente está motivado a fazer X - tese que um filósofo, por exemplo como T. Nagel, terá

que defender para se opor à sugestão humeana de que razões não motivam.

Admite esta passagem?

Se considera que, em última análise, sempre que agimos intencionalmente, estamos a fazer aquilo que mais desejamos (admitindo que o desejar pode equivaler à noção de 'estar motivado'), então é verdade que também quando agimos moralmente estamos a ser conduzidos, em última análise, pelos nossos desejos e não pelos juízos avaliativos, ao contrário do que defenderia uma tese racionalista?

Que papel resta aos juízos avaliativos, se estes não tiverem qualquer poder causal?

RS - É uma ótima questão, que toca no cerne da dificuldade. Até porque, reparem numa coisa: se os juízos de valor não tiverem qualquer papel e formos sempre conduzidos apenas pelo que desejamos, poderá parecer que resolvemos o problema da acrasia, mas agora temos outro problema: como é que vamos explicar a continência ou o auto-controlo? Como é que explicaríamos que certas pessoas consigam resistir ao que desejam e manter-se fiéis ao que pensam ser o melhor? Portanto, se começarmos por representar a situação comum ao acrático e ao agente que se controla como um conflito entre razão e desejo, não podemos depois dar como resposta que estaria na natureza humana que seja sempre o desejo a determinar a acção – o que tornaria o auto-controlo impossível –, nem dar a resposta contrária, segundo a qual seria sempre a razão a comandar – o que tornaria a acrasia impossível. A explicação a dar tem de prever as duas possibilidades.

Um dos maiores obstáculos na compreensão da acrasia – e também na interpretação de Aristóteles – tem origem na aceitação de uma falsa alternativa entre o intelectualismo socrático, por um lado, e o humanismo (enquanto teoria da motivação), por outro. Muita gente julga encontrar uma versão do “as razões não motivam” humeano na afirmação aristotélica de que “o pensamento por si mesmo não move nada” (*EN* 1139a35). Para Aristóteles, o desejo é a força motriz fundamental, que desempenha um papel causal indispensável na motivação e na execução da própria acção. Mas há também um elemento cognitivo essencial na explicação causal da acção. O que a leitura “humeana” faz é reduzir a contribuição da cognição a um

papel puramente instrumental: o seu trabalho seria apenas o de encontrar os meios para realizar ou para alcançar o fim desejado. Mas a cognição, ou as crenças do agente, não desempenhariam qualquer papel na determinação dos fins desejados, pelo que o valor de cada fim dependeria exclusivamente da maior ou menor intensidade com que ele é desejado. Se as coisas fossem assim, a acrasia não teria lugar, pois a suposta acção acrática, sendo a que é mais fortemente desejada no momento da acção, seria também aquela que o agente julgaria ser melhor nesse momento (e, por isso, não seria acrática).

Além disso, numa concepção puramente instrumental da cognição prática, não se entende porque razão se haveria de atribuir uma falha cognitiva ou uma “ignorância” ao agente acrático: o acrático simplesmente segue, como seria de esperar, o seu desejo mais forte e isso é independente de qualquer saber que ele possua ou não a respeito daquele tipo de acção. Quando Aristóteles introduz uma certa ignorância do acrático, em *EN VII 3*, como factor explicativo da acrasia, isso é visto como significativo de uma mudança de perspectiva. A concepção seria agora de tipo intelectualista: a cognição, o conhecimento prático do agente é que determinaria o maior ou menor valor de cada curso de acção possível e, por sua vez, estes juízos de valor reflectir-se-iam directamente na motivação do agente, que desejaria mais aquilo que julga ser melhor. O que se passaria na acrasia, de acordo com esta concepção, é que, devido a uma falha do seu conhecimento prático, o agente acaba por desejar mais fortemente aquilo que, se ele tivesse acesso completo ao seu conhecimento, reconheceria como sendo pior. Esta leitura, como disse antes, é insustentável. Por uma razão simples: a falha cognitiva não pode ser aquilo que explica o facto de o acrático desejar mais fortemente o pior, porque ela é uma consequência desse desejo. São as emoções do acrático, o seu estado motivacional, que por um tempo lhe roubam a sua capacidade cognitiva normal – nomeadamente, a sua capacidade de percepção correcta das coisas e daquilo que elas valem.

Para encontrar espaço para a possibilidade de um desalinhamento entre a motivação e os juízos de valor, precisamos de uma concepção que ocupe uma posição intermédia entre os dois extremos que acabo de apontar. Julgo que Aristóteles tinha uma tal concepção intermédia, embora não seja uma tarefa fácil determinar exactamente qual. A sua definição de decisão como “razão desiderativa ou desejo raciocinativo” é um

bom sinal dessa procura de uma compreensão adequada de como pensamento e desejo se combinam para originar a acção humana. Vou limitar-me a enunciar dois aspectos que me parecem fundamentais nessa concepção intermédia. O primeiro é que, apesar de o desejo ser a principal força motriz, ele é receptivo aos contributos da cognição. Aristóteles repete muitas vezes que o desejo tem por objecto o bom, real ou aparente. O papel da cognição prática não é somente o de encontrar meios para alcançar os fins desejados, mas consiste também em reconhecer diversas coisas como boas, tornando-as assim objecto de desejo. Há por isso uma correlação necessária entre aquilo que o agente deseja e aquilo que ele acha bom: se desejo fazer X, então julgo que seria bom fazer X; e, inversamente também, se julgo que seria bom fazer X, então desejo fazer X. O segundo aspecto fundamental é que, apesar de haver esta correlação, não há garantia de que o agente deseje sempre uma coisa tão fortemente quanto a julga boa. O desejo é receptivo aos factores cognitivos, mas também é influenciado por outras coisas. Por isso, o valor que é reconhecido a um certo curso de acção tem efeitos motivacionais, mas estes podem ser maiores ou menores em agente diferentes. Repare, então, como isto desarma um pressuposto que estava implícito na vossa pergunta. É que, em princípio, não há incompatibilidade entre reconhecer que os juízos de valor são motivadores, por um lado, e admitir que podemos, em certas ocasiões, julgar que X é melhor que Y e, no entanto, estarmos mais motivados para fazer Y.

SM e SC - Escreveu sobre a questão da individuação da acção - de qual das hipóteses de resposta em campo se sente mais próximo, uma vez que, como diz Alfred Mele, não parece haver razões suficientes para preferir uma explicação em detrimento da outra?

RS- Não concordo com Mele, neste ponto. Julgo que Davidson reuniu dados impressionantes, em qualidade e em quantidade, em apoio da ideia de que as acções são, tal como os objectos materiais, entidades particulares, situadas no espaço e no tempo e, por isso, literalmente irrepetíveis. Isto significa que as acções não são universais – propriedades ou relações – que possam ser instanciados em lugares diversos e em múltiplas ocasiões por agentes diferentes; pelo contrário, tal como os agentes, as acções são particulares, que têm propriedades e relações com outros

particulares. Não existe a acção de agradecer, por exemplo. O que existe são muitas acções particulares, cada uma com as suas características próprias, mas que têm em comum poderem ser correctamente descritas como agradecimentos. Note-se que, uma vez que as acções levam tempo a realizar-se, o conceito de existência que aqui empregamos deve ser neutro a respeito do tempo, quer dizer, deve ser entendido de um modo tal que as coisas passadas e futuras existem ao mesmo título que as presentes.

Inicialmente, Davidson usou um argumento semântico em defesa desta perspectiva. Tratava-se de resolver um problema que se manifestava desde logo como um problema de forma lógica. A visão tradicional era considerar que frases como “João é mais alto que Maria” e “João beijou Maria” têm o mesmo tipo de forma lógica. Entre outras coisas, essa visão tinha dificuldade em explicar o facto de “João beijou Maria” ser uma consequência lógica de “João beijou Maria na praia”. Além disso, se os beijos dados foram muitos, qual deles seria o responsável pela circunstância de a frase ser verdadeira? Davidson rompeu com a visão tradicional propondo que a frase seja vista como uma quantificação existencial: podemos parafraseá-la como “Existe uma acção x tal que x foi um beijo dado por João a Maria e x teve lugar na praia”. Esta proposta mostrou-se rica em aplicações e em consequências e foi geralmente bem acolhida, nomeadamente pelos que trabalham em semântica das línguas naturais.

No artigo “Acções, Acontecimentos e Formas Lógicas”⁸⁹, procurei mostrar como é amplo o poder explicativo desta perspectiva segundo a qual as acções e, em geral, os acontecimentos são particulares. Davidson explorou as suas consequências em domínios muito diversos, como a filosofia da acção, a filosofia da mente, a filosofia da linguagem e a filosofia da ciência.

No domínio específico da filosofia da acção, a ideia permitiu lançar uma nova luz sobre as relações entre acção, intencionalidade e causalidade. Elizabeth Anscombe já tinha notado que não faz sentido apontar para uma acção e perguntar se ela foi intencional. Pois a resposta varia consoante o modo como a acção seja descrita. O casamento de Édipo, por exemplo: foi intencional? Bem, Édipo casou

⁸⁹ “Acções, Acontecimentos e Formas Lógicas”, *Cadernos de Filosofia* 6, 1999, pp. 41-83.

intencionalmente com Jocasta, mas não foi intencionalmente que casou com a própria mãe. Uma mesma acção pode ser descrita de muitas maneiras diferentes e será intencional quando descrita de umas maneiras e não o será quando descrita de outras. Isto tem como consequência que não existe o conjunto das acções intencionais de uma pessoa. A ontologia de acções de Davidson veio permitir dar significado literal a esta ideia de uma acção com diversas descrições. Uma outra aplicação diz respeito à distinção, proposta por alguns autores, entre acções básicas e não-básicas. As acções básicas seriam os movimentos que fazemos com o nosso corpo e as não-básicas seriam consequências desses movimentos – coisas como agradecer, cumprimentar, estabelecer contratos, casar, compor sinfonias, etc. Um exemplo típico é o de alguém que move o seu dedo indicador de modo a premir o gatilho e com isso mata um agressor. Certos autores estariam tentados a dizer que o acto de mover o dedo é básico e teve como efeito o acto de matar o agressor, que é não-básico. Davidson conseguiu mostrar que esta suposta distinção assenta numa confusão fundamental, que é a confusão entre as consequências de uma acção e a redescrição dessa acção em termos das suas consequências. Não devemos confundir a acção de matar o agressor com a morte do agressor. Esta última é um acontecimento distinto, causado pelo disparo e que até pode ocorrer muito tempo depois dele. A acção de mover o dedo, a acção de premir o gatilho e a acção de matar o agressor são uma acção só, descrita de maneiras diferentes. Se eu disparei o revólver e isso causou a morte do agressor, então posso descrever o meu disparo como acto de matar o agressor. Num certo sentido, portanto, todas as acções são básicas. Ou talvez seja preferível distinguir entre descrições básicas e não-básicas das nossas acções.

Por fim, é importante sublinhar também que foi a compreensão das acções como particulares que permitiu a Davidson desarmar o argumento wittgensteiniano (de inspiração humeana) contra a teoria causal da acção, isto é, contra a ideia tradicional de que as nossas razões para agir (onde se incluem desejos e crenças) são causas do que fazemos. Wittgenstein julgou ter encontrado uma diferença essencial entre as razões que alguém tem para agir de um certo modo e as causas dessa acção. Ele considerou que uma afirmação causal singular é apenas uma hipótese que precisa de ser confirmada por “um certo número de experiências concordantes”. A hipótese será aceite se essas experiências mostrarem todas que aquela acção é o que normalmente

se segue quando se verificam tais e tais condições. Em contraste com isto, a afirmação de uma razão não teria este estatuto de hipótese sujeita a confirmação repetida.

Mas, se virmos bem, esta ideia das “experiências concordantes” pressupõe que *aquela acção* poderia ser observada diversas vezes (para verificarmos se ela é, em todas as vezes, antecedida por tais e tais condições). Se as acções são particulares, isto não pode estar correcto. Uma vez terminada, uma acção não volta a repetir-se. É um acontecimento único. E o mesmo se deve dizer da sua causa, quer dizer, do acontecimento que a causou – seja ele qual for. A relação causal envolveu aqueles dois acontecimentos e não faz sentido querer observá-la de novo. O que podemos procurar observar são outros acontecimentos semelhantes ou do mesmo género, para vermos se também eles se relacionam causalmente. Davidson concorda com Wittgenstein quanto ao seguinte: a relação entre dois acontecimentos só é causal se for uma instância de uma lei universal (uma lei sem excepções). Mas os acontecimentos são particulares, enquanto a lei relaciona *géneros* de acontecimentos e, por isso, o modo como os acontecimentos instanciam uma lei tem uma estrutura com alguma complexidade. Davidson volta a usar aqui a ideia fundamental de que cada acontecimento, sendo particular, pode ser descrito de muitas maneiras diferentes. Quando fazemos uma afirmação causal singular, ou seja, quando dizemos que um acontecimento *x* causou um outro acontecimento *y*, usamos normalmente alguma maneira de descrever os acontecimentos *x* e *y*, caracterizando-os através de certos predicados, ou em termos de certas propriedades salientes, como “o acontecimento que é *F*” ou “o acontecimento que é *G*”, etc. Se a afirmação causal for verdadeira, ela continuará verdadeira quaisquer que sejam as descrições de *x* e de *y* que usemos (desde que sejam descrições correctas). Davidson dá um exemplo em que as afirmações “o ciclone causou a catástrofe” e “o acontecimento noticiado na página 5 do *Times* de terça-feira causou o acontecimento noticiado na página 13 do *Tribune* de quarta-feira” são igualmente verdadeiras porque se referem exactamente aos mesmos dois acontecimentos. Mas é evidente que não tem de haver uma lei geral a dizer que os acontecimentos noticiados na página 5 do *Times* são seguidos por acontecimentos noticiados na página 13 do *Tribune*! O facto de as relações causais terem de ser instâncias de leis não obriga a que as descrições usadas na afirmação causal singular e na lei relevante sejam as mesmas. E aqui está a chave para o problema de Wittgenstein: se indico a causa de uma certa acção, tem de haver alguma lei que

relacione causas daquele tipo com efeitos deste tipo, mas nada foi dito ainda a respeito do género de lei requerido. Davidson considera que na psicologia não há leis estritas, quer dizer, leis do mesmo género das que se encontram na física, que não admitem excepções e que permitem fazer previsões exactas e fiáveis. Mas não há causalidade sem leis estritas. Por isso, quando usamos o vocabulário psicológico para descrever a razão que alguém teve para agir de certo modo, seria vão querer encontrar, nesse mesmo vocabulário, uma lei estrita a suportar esse nexos causal. Existe uma lei estrita, na opinião de Davidson, mas é uma lei física (que não conhecemos). Aquela relação causal entre uma razão e uma acção pode instanciar essa lei, porque tanto a razão como a acção são também acontecimentos físicos: a acção identifica-se com os movimentos corporais do agente e a razão identifica-se com acontecimentos no seu cérebro. De acordo com a teoria identitativa da mente, os acontecimentos mentais são eles próprios, também, acontecimentos físicos – e aqui, nesta afirmação de identidade, volta a ser fundamental a ideia de que as acções e os acontecimentos são particulares.

SM e SC – A propósito da explicação da acção, como se posiciona em relação à possibilidade de causação mental? Apelar a razões constitui uma explicação da acção no pleno sentido do termo, ou é meramente uma teoria interpretativa? A solução proposta por Davidson – do monismo anómalo – é para si satisfatória? Pensa que essa solução consegue vencer a acusação de epifenomenismo?

RS – Julgo que as razões que temos para agir são causas, e não apenas interpretações, daquilo que fazemos. A teoria proposta por Davidson é monista (por oposição a dualista) na medida em que afirma que tudo é físico. É uma teoria peculiar, que assenta em três princípios: primeiro, que todos os acontecimentos mentais interagem causalmente com acontecimentos físicos; segundo, que sempre que há uma relação causal, existe uma lei estrita da qual ela é uma instância; e, terceiro, que as únicas leis estritas são as leis físicas. Destes princípios segue-se que todos os acontecimentos mentais possuem uma descrição física sob a qual instanciam uma lei estrita, o que implica que todos eles são físicos. A teoria foi acusada pelos seus críticos de conduzir ao epifenomenismo, isto é, a uma visão segundo a qual a vida mental seria apenas um

epifenómeno, dado que o mental seria em si mesmo destituído de qualquer eficácia causal. Quer dizer, a teoria afirma que os acontecimentos mentais causam acontecimentos físicos (por exemplo, as razões causam acções), mas uma vez que esses acontecimentos mentais são também físicos e é enquanto físicos que eles instanciam uma lei estrita requerida pela relação causal, os críticos concluem que é enquanto físicos que tais acontecimentos causam o que causam e que as suas propriedades mentais são irrelevantes e causalmente inertes. A isto Davidson respondeu que não faz sentido, quando um acontecimento x – que é mental e também físico – causa um acontecimento y , perguntar se x causou y enquanto mental ou enquanto físico. A causalidade é uma relação binária extensional, que se expressa em frases como “ x causou y ” e não em frases do género “ x enquanto F causou y ”. É certo que, na frase “ x causou y ”, o acontecimento x pode ser descrito como mental ou como físico, mas, segundo Davidson, isso não faz diferença, pois “redescrever um acontecimento não pode alterar o que ele causa”.

No artigo “Acção e Explicação Causal”⁹⁰, defendi que é possível salvar o monismo anómalo da acusação de epifenomenismo. Apesar de não serem estritas, existem leis psicológicas, com suficiente poder explicativo e que mostram que o mental é causalmente relevante. A relação causal entre dois acontecimentos não é um facto bruto, inexplicável – e o mesmo vale para a relação entre razões e acções. Aliás, quando indicamos a razão porque alguém agiu de certo modo, estamos já a fornecer uma explicação (que é causal) para essa acção. Muitas vezes, essa explicação será considerada insuficiente ou incompleta, pois indica como causa algo que não constitui condição suficiente para a ocorrência daquele tipo de comportamento. Se alguém me diz que Sócrates fez *isto* porque queria *aquilo*, mas eu consigo apontar outros casos de pessoas que também queriam *aquilo* mas não fizeram *isto* (fizeram antes outra coisa diferente), então direi que aquela explicação é insatisfatória – ela até pode estar certa, mas então é incompleta. Passa-se o mesmo noutros domínios. Se houve uma catástrofe e a explicação fornecida é um ciclone, posso ainda objectar que nem todos os ciclones provocam uma catástrofe deste género ou com esta dimensão e perguntar, então, o que tinha este ciclone de especial para produzir tal efeito. Estes pedidos de

⁹⁰ “Acção e Explicação Causal”, in *Analyses / Análises*, ed. por S. Miguens, J.A. Pinto e C. Mauro, Faculdade de Letras da Universidade do Porto, Porto, 2006, pp. 113-122.

completamento são habituais quando se trata de explicações causais. Se nem todos os *F*s causam *G*s, porque é que este *F* em particular o fez? O que é que ele tem, para além de ser *F*, que fez com que produzisse tal efeito? A forma de explicação mais completa será aquela que consiga colocar a relação causal singular sob uma lei geral. No domínio da acção, a forma de explicação mais completa para o facto de, em certas circunstâncias, um agente *s* ter realizado uma acção *a* de tipo comportamental *C* consistirá em indicar como causa uma razão *r*, de tipo mental *M*, e invocar uma lei segundo a qual, em circunstâncias normais, ou semelhantes àquelas, as razões de tipo *M* causam acções de tipo *C*. A limitação que pode ser apontada a esta forma de explicação do comportamento é a incapacidade de tornar completamente explícito quando é que as circunstâncias são normais ou suficientemente semelhantes – incapacidade esta que deixa espaço para que leis deste género possam ter excepções. Mas esta é uma limitação insuperável, se o que pretendemos é uma explicação de tipo psicológico ou intencional. O que não me parece razoável é usar esta limitação como argumento para defender que o mental não tem poder causal. Pois, se temos uma lei destas, podemos mostrar que a propriedade mental *M* produz, em circunstâncias normais, efeitos de tipo *C*.

Isto não significa que uma tal explicação tenha de ser o fim da história. Quando conhecemos uma lei psicológica segundo a qual razões de tipo *M* causam acções de tipo *C*, fica ainda aberta à investigação a questão de saber *como* é que isso acontece, através de que mecanismo físico. Por vezes fico com a impressão de que os acusadores de epifenomenismo quereriam que este mecanismo subjacente fosse descrito ainda em termos psicológicos – e que só assim reconheceriam poder causal ao mental. Mas isso não é razoável. Quando descemos ao nível do mecanismo, deixamos de lidar com agentes, dotados de uma vida mental.

Limites da racionalidade, irracionalidade motivada e emoções – conversa com Vasco Correia

Vasco Correia e Sofia Miguens

Vasco Correia é investigador no Instituto de Filosofia da Linguagem (FCSH/Universidade Nova de Lisboa). Doutorou-se em 2008 em Paris (Université de Paris IV) com uma tese intitulada *La duperie de soi-même et le problème de l'irrationalité*, orientada por Pascal Engel, que deu origem ao livro *La duperie de soi et le problème de l'irrationalité – des illusions de l'esprit à la faiblesse de la volonté* (2010, Éditions Universitaires Européennes). O seu trabalho, actualmente ligado a Jon Elster⁹¹, um dos autores que definiram o domínio de partida do Projecto *Conversations*, tem girado em torno de questões relativas a racionalidade e irracionalidade, irradiando a partir do tópico maior do seu trabalho de doutoramento, o auto-engano.

Sofia Miguens (SM) – Fez os seus estudos de doutoramento em Paris, com Pascal Engel, alguém que trabalhou bastante sobre os tópicos da racionalidade e irracionalidade, além de ser um tradutor e talvez introdutor em França (corrija-me se estou errada), de Donald Davidson, que por sua vez é um autor essencial para os estudos da racionalidade. Os seus tópicos de dissertação relacionam-se de alguma forma com os escritos de Pascal Engel sobre racionalidade e irracionalidade? Ele inspirou-o de alguma forma? Deu-lhe pistas que ainda hoje considera importantes?

Vasco Correia (VC) – Claro que sim, e de que maneira. Pascal Engel teve uma influência absolutamente determinante na minha maneira de encarar a filosofia, e sei que muitos dos seus antigos doutorandos da Sorbonne diriam o mesmo. Acima de tudo, sugeriu-nos um método de clarificação dos problemas, de elucidação dos conceitos e de rigor argumentativo que contrastava imenso com a tradição académica francesa na qual havíamos evoluído. Fiz a totalidade dos meus estudos superiores em França e, sem querer estar a generalizar, sou da opinião que no meio universitário francês poderia haver um maior esforço no que diz respeito à clareza com que se expõem as ideias e à estrutura argumentativa dos textos. O leitor que aborde Lacan, Deleuze ou Derrida pela primeira vez, por exemplo, arrisca-se a não compreender metade do que os autores escrevem e a pensar para si mesmo: “Devem ser coisas

⁹¹ Vasco Correia leva a cabo neste momento um projecto de pós-doutoramento sobre o tema *Emoções e racionalidade*.

demasiado sofisticadas para a minha inteligência”. Foi pelo menos o que me aconteceu a mim, quando era um jovem estudante. Só mais tarde compreendi que é muitas vezes o contrário que sucede, ou seja, que os autores em causa é que carecem frequentemente da clareza necessária para comunicar as suas ideias de forma inteligível. Um episódio que ilustra bem esse aspecto foi a polémica que suscitou Searle ao relatar que Foucault lhe confidenciou numa conversa que considerava o estilo filosófico de Derrida uma espécie de “obscurantismo terrorista”: obscurantismo porque não se percebe ao certo aquilo que ele diz, e terrorista porque quando alguém o tenta criticar, ele replica com o argumento *ad hominem*: “Vous m’avez mal compris, vous êtes un idiot”⁹². É um problema cultural que já vem de longe, embora a situação tenha vindo a melhorar nos últimos anos. Nesse aspecto, Pascal Engel tentou inculcar-nos uma postura que incarna a célebre afirmação de Wittgenstein: “Tudo o que pode ser dito pode ser dito com clareza”. Não sei se conseguiu, mas lá que tentou, tentou. Mas de forma mais geral, ele contribuiu muito para a difusão da tradição “analítica” em França, contra ventos e marés, não só através das traduções dos textos de Ramsey, Davidson e Dennett, mas também graças aos seus livros sobre filosofia da linguagem e filosofia da mente. A um nível mais pessoal, interessaram-me sobretudo as suas reflexões sobre as noções de “crença” e de “aceitação”; assim como, claro está, a sua resposta ao problema do “auto-engano”, ao qual dediquei a minha tese. Foi aliás ele próprio quem me sugeriu que consagrasse a minha tese a esse tema. A minha intenção inicial era trabalhar sobre o tema da *acrasia* ou “fraqueza da vontade”, mas ele chamou-me a atenção para o facto de não haver ainda nenhum trabalho publicado em França sobre o auto-engano, e eu segui o seu conselho. Por sorte, viria a perceber mais tarde que a explicação do enigma do auto-engano contém de certo modo a chave para elucidar o enigma da *acrasia*. Foi pelo menos o que tentei demonstrar na última parte da minha tese.

SM –A influência do pensamento do filósofo norueguês Jon Elster na sua dissertação é perfeitamente notória. O nosso interesse por Elster esteve presente desde o início do projecto *Conversations*; Elster é dos nomes mais interessantes no actual panorama dos estudos da racionalidade prática tanto quanto estes se

⁹² Cf. John Searle, “The World Turned Upside Down”, *New York Review of Books*, vol. 30, nº 16, October 27, 1983.

estendem a questões de ética, política e economia e também em geral ao que podemos chamar uma filosofia das ciências sociais. Não sei se concorda com esta apreciação. O trajecto da obra de Elster é muito particular e é interessante procurar compreendê-lo: como é que, por exemplo, alguém chega do marxismo a questões de racionalidade aplicadas a indivíduos, tais como a acrasia ou o auto-engano? E o que é que no trabalho de Elster pode ser visto como tendo peso por exemplo na filosofia da economia e na filosofia moral tal como estas são praticadas na tradição analítica? Porque de facto uma das coisas interessantes acerca de Elster, pelo menos para mim, é a sua proximidade ao meio académico francês, cuja tradição de ‘estudos da racionalidade prática’ é profundamente diferente da tradição anglo-saxónica.

VC – Sim, estou de acordo, esse duplo enraizamento traz uma grande riqueza ao pensamento de Elster. Por um lado, foi em França que fez o seu doutoramento, sob a orientação de Raymond Aron. Foi aí que se começou a interessar pela teoria da economia, uma vez que a sua tese abordava o tema da produção na obra de Marx. Foi também nessa altura que se começou a interessar pela teoria da racionalidade, que a seu ver tem por missão estabelecer os fundamentos da teoria da economia. Mas, por outro lado, foi nas universidades americanas que evoluiu a partir do final dos anos 1970; foi lá que se associou ao movimento dos “Marxistas analíticos”, e foi também lá que publicou os seus livros mais conhecidos, como por exemplo *Ulysses and the Sirens* (1979), *Sour Grapes* (1983), *Alchemies of the Mind* (1989) ou ainda *Explaining Social Behavior* (2007), para mencionar apenas alguns. Mas desde há uns anos, curiosamente, voltou ao meio académico francês, uma vez que ensina agora no Collège de France e que publicou recentemente três livros redigidos directamente em francês⁹³. É de facto um trajecto interessante, sobretudo para um filósofo norueguês, e creio que tudo isso se traduz na diversidade de referências que alimentam os seus textos. Faria no entanto uma pequena nuance entre essas duas influências, na medida em que as referências contemporâneas de Elster são sobretudo anglo-saxónicas, enquanto que os autores franceses que evoca com mais frequência pertencem regra geral a uma época mais recuada. É claro que existem algumas excepções, como por

⁹³ Elster, *Agir contre soi*, Odile Jacob, 2007 ; *Le désintéressement*, Seuil, 2009 e *L'irrationalité*, Seuil, 2010.

exemplo o sociólogo Pierre Bourdieu, de quem foi amigo íntimo, ou o historiador Paul Veyne, cujo livro *Le pain et le cirque* lhe proporcionou “a maior experiência intelectual da [sua] vida”⁹⁴, ou ainda o filósofo Alexis de Tocqueville, a quem consagrou recentemente um livro⁹⁵. Mas as suas grandes fontes de inspiração dentro dessa tradição são os chamados “moralistas franceses” – autores como Montaigne, Pascal, La Rochefoucauld, Chamfort, Madame de Sablé, La Bruyère – cujas máximas e reflexões antecipam muitos dos aspectos que a psicologia e as ciências sociais procuram demonstrar hoje em dia.

SM – Imagino que nem todos os aspectos da obra de Elster lhe interessem por igual, mas pode certamente, dada a sua proximidade com a obra, dar-nos uma visão panorâmica do trajecto de Elster. Pelo menos o interesse nuclear de Elster por decisões individuais e colectivas (considerando preferências, formação de crenças, emoções) tem pontos comuns com os seus interesses, não tem?

VC – Um dos aspectos mais interessantes da obra de Elster, a meu ver, é a dimensão interdisciplinar das suas análises. Desse ponto de vista, talvez não seja incorrecto dizer que Elster é um filósofo à moda antiga, no sentido em que se move com grande desenvoltura entre as diferentes disciplinas do saber, mas também no sentido em que considera que a filosofia deve estar intimamente ligada às ciências. Creio que qualquer leitor de Elster se apercebe com facilidade de que a única coisa sagrada nas suas análises são os problemas que tenta resolver, e que está disposto a percorrer todos os terrenos e a beber de todas as fontes para encontrar respostas a esses problemas. Para mim, foi e continua a ser um modelo de investigação. É claro que isso tem um preço, que é o estudo sério e aprofundado de cada uma dessas disciplinas, mas o trabalho de Elster ao longo dos anos é a prova viva que se trata de um método filosófico executável e fecundo. Um exemplo desse tipo de análise foi a tentativa de aplicação dos princípios da psicologia e das ciências sociais à obra de Marx⁹⁶. Em lugar de tentar determinar *a priori* se o modelo de sociedade proposto por Marx é bom ou mau, o que é uma questão normativa, Elster começa por questionar se a adopção de tal modelo é sequer possível do ponto de vista dos princípios que regem a

⁹⁴ Elster, *Le laboureur et ses enfants*, Editions de Minuit, 1986, p. 14.

⁹⁵ Elster, *Alexis de Tocqueville: the First Social Scientist*, Cambridge University Press, 2009.

⁹⁶ Elster, *Making Sense of Marx*, Cambridge University Press, 1985.

motivação individual e a causalidade social, sendo que esta é uma questão puramente descritiva. Por outro lado, Elster também é capaz de questionar a universalidade dos axiomas da teoria da escolha racional confrontando-a com fenômenos de irracionalidade bem específicos, como por exemplo a acrasia e o auto-engano, que os filósofos tiveram o mérito de pôr em evidência. E, reciprocamente, não hesita em demonstrar que a explicação filosófica desses fenômenos teria muito a beneficiar com a tomada em consideração dos estudos empíricos dos psicólogos sobre essa matéria, que não são poucos. Existe portanto da sua parte uma tentativa ambiciosa de conjugar as diferentes abordagens e os diferentes métodos científicos no sentido de atingir a resposta mais completa possível aos problemas de que trata.

SM – Ainda acerca dos interesses variados de Elster: aquilo a que ele chama por vezes ‘estudo empírico da justiça ou das motivações da justiça ou equidade na alocação de recursos’, também lhe interessa a si? E mais em geral, em que medida é que o que passa em torno de Elster é importante para as ciências sociais hoje?

VC – Acho esses temas muito interessantes, mas confesso que não sou grande especialista nessa matéria. Em termos gerais, quanto à importância de Elster para as ciências sociais hoje, creio que o seu trabalho de pode ser encarado como uma reflexão crítica sobre os próprios fundamentos das ciências sociais. É a esse nível que se parece inscrever a sua teoria da racionalidade, que é ao mesmo tempo uma teoria da irracionalidade. Mas Elster não tem a pretensão de edificar um sistema axiomático capaz de unificar o conhecimento das várias disciplinas. Pelo contrário, o que lhe interessa é o estudo de um conjunto de *mecanismos* – ou, como ele gosta de dizer, uma “caixa de ferramentas” de mecanismos – que permita oferecer uma explicação a fenômenos psicológicos e sociais complexos. Os mecanismos não são leis, e o estudo dos mecanismos nem sempre permite estabelecer previsões unívocas, ao contrário do que se passa em princípio com as leis. Mas os mecanismos permitem compreender as razões pelas quais as leis se verificam em determinados casos e não noutros. Por exemplo, a lei da procura prevê que os consumidores tendem a comprar determinado produto em maior quantidade quando o seu preço baixa, e menos quando o seu preço aumenta. Mas existem exceções, como nos casos em que o preço acrescido do produto é interpretado como um índice da sua qualidade elevada, ou mesmo como um

símbolo de um certo estatuto social. Nesses casos, um aumento do preço pode traduzir-se por um aumento da procura. Ora, o estudo destes casos põe em evidência uma série de mecanismos que a lei não toma em consideração, como por exemplo o efeito de “racionalização pós-compra” que consiste em raciocinar da seguinte maneira: “Dado que paguei um preço elevado por este produto, ele deve ser de boa qualidade”, que é uma falácia notória. Daí o interesse do estudo dos mecanismos, pois permite explicar o que acontece quando as generalizações se tornam falíveis, obrigando por conseguinte a uma reformulação mais adequada das leis.

Foi precisamente isso que aconteceu no domínio da teoria da decisão a partir do final dos anos 1960, graças sobretudo aos trabalhos de Amos Tversky e de Daniel Kahneman. Até essa altura, os comportamentos irracionais eram considerados como residuais e de certo modo negligenciáveis. Era tido como irracional aquilo que não encaixava nos moldes da teoria da escolha racional. No entanto, a profusão de estudos empíricos que os psicólogos realizaram desde essa altura permitiram compreender que os fenómenos que a teoria da escolha considerava como irracionais são demasiado frequentes para serem negligenciados como meras excepções. Isso obrigou a uma revisão da teoria da decisão racional e ao desenvolvimento de modelos interessados em explicar como é que as pessoas realmente decidem em condições concretas, com restrições de tempo e de informação. Hoje em dia, muitas teorias reconhecem uma racionalidade subjacente a fenómenos que pareciam puramente irracionais. Existem, por exemplo, mecanismos de raciocínio rápido, as chamadas “heurísticas”, que são muitas vezes o método mais racional de que dispomos para formar o nosso juízo, em particular quando o tempo é pouco e a informação disponível incerta. Eis portanto um exemplo de como a abordagem proposta por Elster pode ser frutuosa. Mas poderia evocar outros tipos de análise, como por exemplo a aplicação da psicologia à compreensão de determinados fenómenos sociológicos, muito em linha com o trabalho de Bourdieu. Outro exemplo é a aplicação da filosofia à psicologia, com a introdução de conceitos como *meta-emoção* e *proto-emoção* que me parecem extremamente úteis para as descrições psicológicas da motivação dos agentes. Por outro lado, há uma dimensão propriamente normativa na sua investigação, como sobressai da sua defesa das estratégias de “compromisso prévio” (*precommitment strategies*). A seu ver, a melhor maneira de evitar comportamentos irracionais consiste em partir do princípio que existem mecanismos que nos impelem a agir de forma irracional, e a conceber estratégias indirectas de

auto-controlo. Daí o título do seu livro *Ulysses and the Sirens*, pois a estratégia de Ulisses é um excelente exemplo de compromisso prévio: ele parte do princípio que é incapaz por si mesmo de resistir ao canto das sereias, e pede que o amarrem ao mastro da embarcação.

SM – Foi também porque os fenómenos de racionalidade e irracionalidade são algo sem o qual não é possível pensar em fenómenos sociais que no projecto *Conversations* nos interessou compreender o ‘ponto de aplicação’ desses conceitos: de que dizemos (de que devemos dizer) que é racional ou racional: a acção? a agência? o agente? Num artigo seu intitulado ‘Os limites da racionalidade: auto-engano e acrasia’⁹⁷ diz a certa altura, citando Davidson, que a transgressão da racionalidade é uma transgressão de normas *do agente*. Diz também, na conclusão da sua tese (*Pour une science de l’affectivité – L’irracionalité expliquée*), que Davidson foi alguém que compreendeu perfeitamente que o grande desafio para uma teoria da racionalidade é ser capaz explicar a irracionalidade. Podemos dizer que irracionalidade é irracionalidade do agente? Qual é a sua posição aqui? Gostava que nos explicasse um pouco como é que a opção a tomar neste ponto se pode reflectir, se é que pensa que pode, nas (diversas) posições contemporaneamente defendidas sobre auto-engano e acrasia.

VC - Penso que talvez seja mais propício falar-se de *acção* racional (ou irracional) do que de *agente* racional (ou irracional), e isto por uma razão muito simples: agentes racionais (e irracionais) somos todos nós, embora uns mais do que outros. Ou haverá alguém que nunca fez o contrário daquilo que lhe pareceu ser a melhor opção? Neste sentido, parece-me algo redundante falar de agentes irracionais. É claro que, ao analisar determinado comportamento irracional, somos levados a qualificar o autor desse comportamento de irracional, mas esse epíteto refere-se no fundo a determinadas atitudes do agente. Os agentes racionais e os agentes irracionais não são pessoas diferentes; estou convencido de que somos todos mais ou menos racionais, e mais ou menos irracionais. Na minha opinião, esta hipótese aplica-se inclusive aos agentes que sofrem de distúrbios mentais considerados patológicos, cujos níveis de

⁹⁷ *Disputatio*, vol. IV, no 29, Novembro, 2010.

irracionalidade são sem dúvida mais graves do que o das pessoas ditas normais, mas que também são racionais até certo ponto. Há uma anedota portuguesa que ilustra muito bem esse aspecto. Começa com um condutor que ficou parado em frente ao hospital psiquiátrico Júlio de Matos porque lhe saltou uma roda do carro. Como já não tinha os quatro parafusos da roda, não sabia como havia de colocar outra vez a roda no eixo. Entretanto, sai um paciente do hospital que o interpela e lhe sugere: “Porque é que o senhor não tira um parafuso de cada uma das outras três rodas? Assim fica com três parafusos em cada roda, o que chega perfeitamente”. O condutor exclama-se então perplexo: “Mas... Pensei que o senhor fosse maluco”. Ao que o outro responde: “Posso ser maluco, mas não sou estúpido!”. A ideia aqui implícita é que uma pessoa pode ser extremamente irracional a certos níveis, ou em determinados contextos, mas racional a outros níveis ou noutros contextos. Seja como for, o facto é que a ideia de que existem os “loucos” e os “normais” já foi abandonada há muito tempo pelos manuais de nosologia psiquiátrica. Até mesmo a esse nível, há que distinguir entre diferentes graus de irracionalidade, e aliás o próprio critério que permite dizer que determinado caso de irracionalidade é patológico é em grande parte dos casos um critério puramente relativo. Segundo a edição mais recente do DSM-IV da Associação Psiquiátrica Americana, o nível de impacto que o distúrbio mental tem na vida e no bem-estar da pessoa é um dos principais critérios que permitem distinguir o normal do patológico⁹⁸. Por exemplo, muitos de nós temos medos irracionais, isto é, sentimos medo em situações que não o justificam; mas só quando um desses medos causa tamanha ansiedade que nos impede de viver uma vida normal, de conviver com as pessoas, de trabalhar, etc., é que passa a ser considerado uma fobia. Neste tipo de casos, por conseguinte, existe apenas uma diferença de grau entre o sintoma normal e o sintoma patológico, e a própria origem dos medos irracionais é muitas vezes idêntica num caso como noutro. Ora, se pretendêssemos fazer uma distinção de princípio entre o agente irracional e o agente racional, como faziam os psiquiatras noutros tempos, seríamos incapazes de detectar as continuidades entre uns e outros. É por isso que me parece importante centrar o conceito de racionalidade nas acções individuais e não nas pessoas.

⁹⁸ Cf. *Diagnostic and Statistical manual of Mental Disorders*, DSM-IV-TR, American Psychiatric Association, 2000, p. 8.

Existe contudo uma segunda razão, que é o grau de precisão acrescido que se obtém quando examina a racionalidade das acções particulares em lugar da racionalidade dos agentes enquanto tais. De facto, um agente pode ser irracional a outros níveis que não o da acção, e importa distinguir esses diferentes planos. Quando nos contentamos em qualificar um agente de irracional, ficamos sem saber ao certo o que é que é irracional no conjunto das suas atitudes. Poderá ser a sua acção, sem dúvida, mas também poderá ser uma crença, ou porventura até uma emoção. E até poderá ser uma emoção irracional que causa a acção irracional, como acontece justamente no caso da fobia. Mas isso só é possível compreender quando se fala de instâncias mais específicas do que o agente no seu todo, isto é, quando se fala dos seus estados mentais e das suas atitudes particulares.

Esse problema é patente na forma como os fenómenos de auto-engano e de acrasia têm sido interpretados na tradição filosófica posterior a Aristóteles. No caso da acrasia, por exemplo, que em grego antigo significava simplesmente “falta de controlo sobre si mesmo”, houve uma grande tendência para interpretar o problema em termos de “fraqueza da vontade”. E uma das razões para tal é precisamente essa maneira de centrar a questão no agente, em vez das suas atitudes específicas, que resulta a meu ver numa simplificação excessiva dos termos do problema. Enquanto Platão e Aristóteles propuseram análises sofisticadas do problema, que antecipavam até certo ponto os fenómenos de distorção do juízo e de conflito entre diferentes tipos de desejos, verifica-se a partir de Cícero que a questão da acrasia se reduz para muitos filósofos ao problema da luta entre a vontade e os desejos. Segundo Descartes, por exemplo, o agente cuja vontade é “forte” revela-se capaz de resistir às tentações e de agir em conformidade com a razão, e o agente cuja vontade é “fraca” cede aos caprichos dos desejos. Trata-se de um modelo muito influente que ainda hoje impregna o vocabulário do senso comum. O problema desse tipo de interpretação é que não nos permite compreender porque é que certas vontades são fortes e outras fracas, ou porque é que determinada vontade é forte em certos momentos e não noutros. Fala-se da vontade como se fosse uma espécie de músculo, mais forte em certos agentes do que noutros, mas não se percebe ao certo o que é que isso da vontade significa concretamente. E nestas condições torna-se difícil compreender adequadamente o fenómeno, como aliás se torna difícil encontrar métodos capazes de o evitar. O agente irracional é um agente cuja vontade é “fraca”, e como ninguém

sabe ao certo de que misteriosos princípios depende a “força de vontade”, não há grande coisa a fazer.

No caso do auto-engano, deparamo-nos com um problema muito semelhante, e alguns autores contemporâneos sugerem mesmo que o auto-engano é uma forma de “incontinência doxástica” ou de “crença acrática”, isto é, uma fraqueza da vontade que tem lugar sobre o plano cognitivo. Outros vão inclusivamente ao ponto de declarar que se trata de uma “cobardia epistémica”, ou seja, que as pessoas que se iludem a si mesmas o fazem porque não têm coragem de encarar a realidade. Pessoalmente, não vejo em que é que esta moralização da questão nos permite compreender melhor o fenómeno. Além disso, parece-me arriscado procurar a explicação da irracionalidade nas características idiossincráticas do agente, quando se trata de um fenómeno amplamente partilhado. Tudo isto resulta precisamente dessa propensão para simplificar os termos do problema e para focalizar a atenção na pessoa do agente, em lugar de examinar as relações entre as atitudes concretas do agente, isto é, entre as suas acções e as suas crenças, entre as suas crenças e as suas emoções, entre as suas emoções e as suas memórias, e por aí fora.

SM - De que forma é que as posições defendidas por filósofos neste contexto se cruzam com o que os psicólogos chamam ‘irracionalidade motivada’, um fenómeno que o interessou? Em que contexto é que os psicólogos estudam a irracionalidade motivada e a que psicólogos em particular vale a pena prestarmos atenção?

VC - Aquilo que os psicólogos demonstraram fundamentalmente, com base nos estudos empíricos, é que somos todos vítimas de dois tipos de irracionalidade cognitiva: a irracionalidade motivada, ou “quente”, que resulta da influência das emoções e dos desejos sobre a formação do juízo, e a irracionalidade não-motivada, ou “fria”, que corresponde a erros sistemáticos de raciocínio que cometemos devido à maneira como as nossas faculdades cognitivas se encontram estruturadas. O fenómeno de irracionalidade não-motivada foi posto em evidência por psicólogos sociais como Richard Nisbett e Lee Ross, cujo livro *Human Inference* (1980) teve uma grande influência na interpretação desses casos de irracionalidade, e sobretudo pelos trabalhos de Daniel Kahneman, Amos Tversky, Paul Slovic e Thomas Gilovitch, entre outros, que apareceram agrupados pela primeira de maneira

sistemática vez no livro *Judgment Under Uncertainty* (1982). Foram eles que distinguiram e classificaram toda uma série de “enviesamentos do juízo” (*biases*) e de “heurísticas” que são susceptíveis de afectar o nosso juízo em determinados contextos. O fenómeno de “irracionalidade motivada”, por seu lado, diz respeito a fenómenos de enviesamento do juízo que são induzidos por um determinado motivo do agente, em geral uma emoção ou um desejo. David Pears teve um papel muito importante na introdução deste fenómeno na reflexão filosófica, graças a um livro que se intitula justamente *Motivated Irrationality* (1984). Mas o livro que não posso deixar de recomendar, porque me influenciou mais que nenhum outro, foi o livro notável de Alfred Mele, *Irrationality* (1992). Convém também observar que a irracionalidade motivada é um fenómeno relativamente intuitivo, na medida em que os provérbios populares, os poetas e os filósofos sempre souberam alertar para os seus efeitos. Por exemplo, quando dizemos que “o amor é cego” ou que “o maior cego é aquele que não quer ver”, estamos a evocar um mecanismo desse tipo, que é a influência que uma determinada emoção é susceptível de exercer sobre o processo de formação da crença. O mérito dos psicólogos, a este respeito, não foi tanto o de demonstrarem *que* as emoções afectam o nosso juízo, porque isso já toda a gente sabia, mas o de demonstrarem *até que ponto* somos vítimas desses efeitos, e também de que variedade de maneiras isso acontece. Um exemplo disso é a chamada “ilusão de superioridade”, ou *above average effect*, que é a propensão para sobrestimar os seus méritos e para subestimar os seus defeitos em comparação como os das outras pessoas. Intuitivamente, poderíamos pensar que só as pessoas “convencidas” ou “pretensivas” é que sofrem deste mal. No entanto, os psicólogos demonstraram que a maioria das pessoas pensa que é mais inteligente, mais feliz, mais justa, mais sociável, melhor ao volante e menos preconceituosa do que a média das outras pessoas. Outro exemplo: num inquérito a um milhão de estudantes, os psicólogos verificaram com espanto que a totalidade dos estudantes declarou que se considerava mais sociável do que a média (a totalidade num milhão!), e que 25% estavam convencidos que faziam parte do grupo dos 1% mais populares do liceu. Seria tentador pensar que este tipo de ilusão se deve à imaturidade dos estudantes, mas o inquérito aos professores revelou que 95% se consideravam melhores nos seus trabalhos do que a média dos seus colegas⁹⁹. Tudo isto revela que o fenómeno de irracionalidade motivada é muito mais

⁹⁹ Cf. Thomas Gilovich, *How We Know What Isn't So*, The Free Press, New York, 1991, p. 77.

vasto e frequente do que aquilo que se possa pensar. Ou do que aquilo que se queira admitir... A verdade é que somos quase todos afectados por estas formas de enviesamento do juízo induzidas pelas emoções. No último caso, por exemplo, parece provável que o nosso orgulho próprio seja a causa do mecanismo de distorção cognitiva que produz a ilusão de auto-comprazimento. Mas existem muitos outros mecanismos deste tipo.

SM – Considera portanto que numa certa perspectiva, todos somos, em variadas ocasiões, irracionais – em que sentido então é que fenómenos de irracionalidade como o auto-engano podem ser, de um ponto de vista prático, ‘racionalis’?

VC – O auto-engano pode ser racional no sentido em que a adopção voluntária de uma crença ilusória é susceptível de tornar o agente mais feliz, seja porque proporciona uma visão mais cor-de-rosa da realidade, ou seja porque permite denegar uma realidade psicologicamente devastadora. Por exemplo, um doente em estado terminal que se queira convencer a si mesmo que vai sobreviver à sua doença, não obstante a opinião dos médicos, não parece ser irracional do ponto de vista prático, na medida em que a adopção dessa crença falsa seria uma maneira de aumentar o seu bem-estar durante o pouco tempo que lhe resta. Do ponto de vista *epistémico*, isto é do ponto de vista das “razões para acreditar”, é seguramente irracional adoptar uma crença manifestamente contrária à informação de que se dispõe, já que o desejo de acreditar que p não é uma boa razão para acreditar que p . Mas do ponto de vista *prático*, isto é, do ponto de vista das “razões para agir”, pode ser perfeitamente racional preferir uma doce ilusão à verdade crua e nua, na medida em que a adopção de uma ilusão positiva pode ser uma acção propícia à maximização do bem-estar. Mas é claro que tudo depende do critério de racionalidade prática que se adopte. O critério mais consensual em teoria da escolha é o critério utilitarista, segundo o qual a acção racional é aquela que procura maximizar o bem-estar do agente em função dos constrangimentos da sua escolha e à luz da informação de que ele dispõe. A vantagem deste tipo de critério é que permite demarcar a questão da racionalidade de considerações de ordem moral, uma vez que se trata de um critério puramente formal. Quaisquer que sejam as preferências de determinado agente, e por muito inaceitáveis ou imorais que nos possam parecer, devemos considerar esse agente como racional desde que as suas acções permaneçam coerentes com as suas preferências. Trata-se do

critério de racionalidade mais minimal e tolerante até agora formulado: só se pede ao agente que aja em conformidade com aquilo que ele próprio considera racional; mas o agente é livre de decidir que tipo de normas ou valores pretende adoptar. Nestas condições, como Davidson sublinha, o paradigma do agente irracional é o indivíduo com falta de controlo sobre si mesmo. É o caso, por exemplo, do indivíduo que quer deixar de fumar, que está convencido que seria muito mais feliz se conseguisse viver sem tabaco, mas que é incapaz de resistir à tentação de acender mais um cigarro. Isso sim, é irracional, porque o agente escolhe aquilo que é pior para si à luz das suas próprias preferências. Em contrapartida, o agente que prefira o prazer de fumar à perspectiva de uma vida mais longa não é de todo irracional, se tal é de facto a sua decisão, porque está a agir em coerência com aquilo que prefere. Aquiles também preferiu uma vida curta e gloriosa a uma vida longa sem glória, não há nada de irracional nesse tipo de escolha. Do mesmo modo, para voltar ao caso do auto-engano, o agente que se queira convencer a si mesmo que não vai morrer em breve também não está a ser irracional do ponto de vista prático, porque tem o direito de preferir uma existência ilusória mas cheia de esperança a uma existência realista mas deprimente. Até aí, estou perfeitamente de acordo com a teoria dita “intencionalista” do auto-engano. A questão depois é de saber até que ponto isso é possível na prática. É sobre esse ponto que não estou tão de acordo.

SM - Então porque é que o intencionalismo (acerca do auto-engano) é atraente? Defende no seu trabalho que o intencionalismo conduz a paradoxos – quais são esse paradoxos? E onde entra a necessidade de fazer apelo a uma ‘divisão da mente’?

VC - O “intencionalismo” é a tese segundo a qual o auto-engano é um processo intencional que o agente desencadeia por sua própria iniciativa. É uma ideia implicitamente presente em expressões do senso comum que se empregam para denunciar a atitude de alguém que “enfia a cabeça debaixo da areia” ou que “não tem coragem para encarar a realidade”. Tudo isto subentende que o agente que se ilude tomou de certo modo a decisão de se iludir. Mas um agente só se pode mentir a si mesmo se tiver um conhecimento prévio da verdade, pois caso contrário não faria sentido dar-se ao trabalho de se mentir a si mesmo. Por exemplo, só faz sentido tentar mentir-me a mim mesmo relativamente ao meu estado avançado de calvície se

primeiro reconhecer esse estado. Segundo Davidson, esse reconhecimento prévio da verdade constitui mesmo a causa do auto-engano, pois é a crença que p , conjugada com o desejo que não- p , que conduz o agente a decidir acreditar que não- p . É daí que surge a necessidade de pressupor o postulado da divisão da mente, pois esta análise parece implicar a possibilidade de o agente dispor simultaneamente de duas crenças contrárias, a crença que p e a crença que não- p . No caso da calvície, por exemplo, o intencionalismo preconiza que uma parte da mente do agente sabe que está a perder cabelo, mas outra parte da mente produz a crença que não está a perder cabelo.

Qual é a vantagem desta interpretação? Antes de mais, permite considerar que os agentes são directamente responsáveis pelas suas ilusões. Supondo que de facto os sujeitos se iludem por sua própria iniciativa, mediante um cálculo utilitário que vê mais benefícios na ilusão do que na verdade, o auto-engano deixa de ser um fenómeno apenas um fenómeno simplesmente *irracional* e passa a ser também um fenómeno *imoral*, porque é sabido que as ilusões de cada um podem ter consequências desastrosas não só para si mesmo mas também para os outros. Se decidir iludir-me relativamente à quantidade de vinho que bebi durante o jantar, e me puser ao volante do meu carro, estarei a colocar em risco a vida dos outros. A concepção intencionalista permite portanto responsabilizar os agentes pela racionalidade das suas crenças.

O problema, todavia, é que não podemos dar validade à hipótese intencionalista simplesmente em virtude das suas implicações sobre o plano ético. Como é evidente, não faz sentido responder a um problema puramente descritivo (como é que o auto-engano se produz de facto?) – com base em preocupações de ordem moral. Primeiro há que determinar se o auto-engano é ou não um processo intencional, que é uma questão de facto, e só em função dessa resposta é que se deve colocar a questão ética. Ora, desse ponto de vista, vários autores puseram em evidência o facto de a concepção intencionalista ser pouco convincente e muito paradoxal. Pouco convincente, em primeiro lugar, porque não há indícios comprovados de que seja possível decidir acreditar naquilo que se deseja graças a uma simples decisão. Por muito que me queira convencer de que não estou a perder cabelo, por muito que isso aumentasse o meu conforto psicológico, a imagem da minha careca reflectida no espelho lembra-me cada manhã que estou de facto a perder cabelo. As crenças são em princípio estados mentais involuntários que não se encontram sob o controlo da vontade. Acontece, obviamente, que os nossos desejos

influenciem as nossas crenças, como já referi, mas isso não se passa em princípio mediante um processo intencional. E em seguida, há o problema da concepção intencionalista se confrontar a uma série de paradoxos. Seria fastidioso reconstituir aqui com rigor os termos desses paradoxos, mas um dos mais importantes é precisamente a suposição de que as nossas mentes são capazes de acreditar em duas coisas contraditórias simultaneamente. Por outro lado, como Alfred Mele explica muito bem, não se percebe muito bem como é que a mente é capaz de se deixar enganar pela sua própria intenção e pelo seu próprio plano de auto-engano, sem que a própria consciência dessa intenção e desse plano inibam a adesão sincera à crença em questão. O intencionalista pode redarguir que se trata de intenções inconscientes, como o fazem David Pears e Herbert Fingarett. No entanto, mesmo supondo que tais intenções existem de facto, fica por explicar como é que uma parte inconsciente da mente pode ser suficientemente sofisticada não apenas para prever as consequências positivas ou negativas que cada crença é susceptível de ter no bem-estar psíquico do indivíduo, mas também para “censurar” cada crença aparentemente perigosa. É aquilo a que o Sartre chama o “paradoxo do recalçamento”.

SM – O Vasco defende que alguns paradoxos da irracionalidade são paradoxos de uma concepção intencionalista e que uma concepção não intencionalista pode escapar-lhes – poderia explicar como é que tal coisa é possível?

VC – O que tentei demonstrar na minha tese, é que os pretensos “paradoxos da irracionalidade” resultam no fundo de concepções paradoxais da irracionalidade, ou seja, que são sintomáticos de uma interpretação inadequada dos fenómenos de irracionalidade. No caso do auto-engano, por exemplo, vários filósofos consideram que se trata de um fenómeno intencional, e resignam-se a aceitar um ou outro dos paradoxos a que essa hipótese conduz, quando existe uma concepção alternativa que permite explicar o fenómeno de auto-engano de forma perfeitamente inteligível, sem qualquer tipo de paradoxo. Essa concepção alternativa, proposta por Alfred Mele e Ariela Lazar, é a concepção dita “motivacional” do auto-engano, segundo a qual é de forma *involuntária* e *inconsciente* que se formam as crenças irracionais. Muito concretamente, é devido ao impacto das nossas emoções e dos nossos desejos sobre a maneira como tratamos a informação que acabamos por adoptar crenças que se revelam totalmente contrárias àquilo que a informação parece sugerir. Por exemplo,

quando uma pessoa apaixonada se põe a pensar que o seu amor é correspondido, apesar de tudo parecer indicar o contrário, não é porque decidiu intencionalmente que essa crença poderia aumentar a sua felicidade. É simplesmente porque o seu sentimento afecta a maneira como ela interpreta os dados de que dispõe, focalizando a sua atenção nos índices que parecem confirmar esse amor e desviando a sua atenção nos índices contrários. Para um observador exterior, até pode parecer que a pessoa apaixonada *não quer ver* aquilo que salta aos olhos de qualquer um, ou seja, que é de forma deliberada que se obstina a acreditar naquilo que mais lhe convém. Mas trata-se de uma ilusão de interpretação, que resulta justamente do facto do observador exterior não estar a ler a informação com as lentes deformadas da pessoa apaixonada.

À luz desta hipótese, em todo o caso, dissipam-se os paradoxos do auto-engano na sua totalidade. Em primeiro lugar, não é necessário admitir a possibilidade de acreditar duas coisas contraditórias ao mesmo tempo, porque se trata apenas uma crença que é contradita pela informação disponível, e não por uma crença inicial oposta. Em segundo lugar, também não há paradoxo da estratégia, porque tudo acontece de forma não intencional, mediante os mecanismos motivados de enviesamento do juízo. E precisamente pelas mesmas razões, deixa também de haver um paradoxo da divisão da mente. Na medida em que não se pressupõe a existência de crenças contrárias, deixa de ser necessário recorrer ao postulado segundo o qual a mente se divide em sub-sistemas diferenciados. Isso não significa que a mente não seja susceptível de sofrer tais diferenciações em determinadas circunstâncias, como acontece de facto nos casos patológicos de “dissociação mental”; significa apenas que não é necessário pressupor esse fenómeno para explicar de forma inteligível os casos comuns de auto-engano. Por fim, talvez seja bom salientar que a hipótese motivacional se encontra em perfeita conformidade com aquilo que revelam os estudos empíricos dos psicólogos.

SM - Portanto na sua opinião os paradoxos ‘desfazem-se’ quando passamos de uma concepção intencionalista para uma concepção motivacional - recordando o que li na sua tese, os seis paradoxos do auto-engano com que lida são o paradoxo doxástico, o paradoxo da estratégia, o paradoxo da interpretação intencionalmente enviesada dos dados, o paradoxo económico dos casos negativos, o paradoxo económico dos casos positivos, o paradoxo da identidade pessoal – certo? Quer acrescentar mais alguma explicação que considere

importante sobre qualquer um deles, que sirva para lançar mais luz sobre a concepção motivacional? Exemplos práticos são sempre muito importantes...

VC – Um dos tipos de auto-engano mais enigmático, que a meu ver nenhuma das concepções intencionalistas consegue explicar de forma convincente, é constituído pelos chamados casos “negativos” de auto-engano, no qual o agente se persuade a si mesmo de algo de absolutamente indesejável. É, por exemplo, o caso do indivíduo paranóico que se convence que os outros andam a conspirar contra a sua vida; é também o caso do marido excessivamente ciumento que se convence que a sua esposa o anda a enganar; ou ainda o caso do pessimista inveterado que se convence que a realidade mais indesejável é precisamente a mais provável. Supondo que a hipótese intencionalista está correcta; supondo que a causa do auto-engano é uma intenção que resulta de um raciocínio prático graças ao qual o agente (ou uma parte da sua mente) percebe que tem todo o interesse em adoptar determinada crença falsa, o que é que leva o paranóico, o ciumento ou o pessimista a persuadirem-se a si mesmos de realidades dolorosas ou mesmo insuportáveis? Que interesse têm eles nisso? Do ponto de vista de um raciocínio prático, tal intenção de auto-engano não parece fazer nenhum sentido. Do ponto de vista da concepção motivacional, em contrapartida, este tipo de caso não encerra mistério nenhum. A hipótese mais provável, que é válida para cada um dos casos, é que a crença irracional dos sujeitos é induzida por determinada emoção irracional. Assim, por exemplo, se o marido ciumento se convence que a sua mulher o anda a enganar, não é devido a uma inexplicável intenção de acreditar tamanho horror, mas simplesmente devido a um sentimento excessivo de ciúme, e mais precisamente devido à maneira como essa emoção irracional acaba por deformar a sua interpretação da informação de que dispõe. Mas isto não tem nada de paradoxal, uma vez que as emoções, contrariamente às intenções deliberadas, não têm por vocação maximizar o bem-estar do agente.

Mas mesmo se examinarmos os casos ditos “positivos” de auto-engano, nos quais os sujeitos se persuadem daquilo que mais lhes convém, a explicação intencionalista também revela os seus limites. Jon Elster é um dos raros a sublinhar este aspecto: só faria sentido adoptar a crença falsa que p de forma intencional se essa crença aumentasse de facto o meu bem-estar. No entanto, na grande maioria dos casos, a adopção de uma crença falsa é uma péssima maneira de maximizar o nosso próprio interesse. É fácil perceber porquê. Mesmo admitindo que me fosse possível

acreditar, aqui e agora, que o artigo que estou a redigir está muito bom e que vai ser aceite para publicação, é evidente que a ilusão daí resultante me impediria de tomar as medidas necessárias para fazer com que de facto o artigo tenha possibilidade de vir a ser publicado. Neste sentido, como diz Espinosa, não há nada mais útil ao homem do que a verdade. É a consciência das insuficiências do artigo em questão que me conduz a trabalhar ainda mais e a aumentar a probabilidade de que venha a ser publicado. Quer isto dizer que, se tomarmos em consideração o interesse do sujeito no longo termo, e não apenas no curto termo, a decisão deliberada de se enganar a si mesmo parece absurda e contra-producente. Existe talvez uma excepção, que é o caso do doente em fase terminal que queira manter esperanças ilusórias de convalescença, mas isso deve-se precisamente ao facto de a perspectiva do longo termo deixar de entrar na equação.

SM – Mudemos agora um pouco de assunto, se bem que para um assunto relacionado, i.e. relacionado com as relações da racionalidade e da irracionalidade com a motivação. Falo do tópico a que normalmente se chama ‘ética da crença’. A relação entre as características do nosso pensamento ético e a questão epistemológica acerca do que nos faz ter uma crença interessou-o – será que podia explicar-nos porquê? Onde entra aqui a questão da responsabilidade e porque razão é ela importante? O artigo clássico de W. Clifford é importante para as questões da ética da crença – será que podia recordar o que lá se defende?

VC - As reflexões sobre o tema da “ética da crença” exploram a ideia segundo a qual somos responsáveis não apenas por aquilo que fazemos, mas também por aquilo que pensamos, ou pelo menos pela maneira como pensamos. O debate filosófico em torno deste tema principiou com o famoso texto do matemático e filósofo britânico William Clifford, intitulado precisamente “A ética da crença” (1878), no qual defendia de maneira polémica que é ilegítimo, no sentido moral, adoptar determinada crença quando a informação de que se dispõe é insuficiente para fundamentar essa crença¹⁰⁰. Esta tese pode parecer um pouco radical, mas quando embarcamos num avião, por

¹⁰⁰ W.K Clifford, «The Ethics of Belief» (1878), in E.D. Klemke, A. David Kline et R. Holinger (eds.), *Philosophy : Contemporary Perspectives on Perennial Issues*, 4th edition, 1994, N.Y., St. Martin’s Press, p. 66-71 .

exemplo, partimos do princípio que os técnicos responsáveis pela verificação do aparelho dão os seus pareceres de maneira fundamentada. Mais: temos tendência para considerar que eles têm a obrigação moral, e não apenas profissional, de tomar em consideração todos os elementos pertinentes antes de chegar à conclusão que o avião está em boas condições. Afinal de contas, são as nossas próprias vidas que dependem da maneira como ele forma o seu juízo, e nesse sentido o aspecto moral da crença torna-se quase palpável. Ora, o que Clifford deplora é que não tenhamos sempre esse tipo de cuidado com as nossas crenças, inclusive no que toca a crenças fundamentais, sob pretexto que são questões de ordem privada que não têm consequências para os outros. É evidente que o aspecto polémico da tese cliffordiana, quando se lê nas entrelinhas, é que põe em causa a própria moralidade do acto de *fé* no sentido religioso. Embora Clifford evite cuidadosamente abordar essa questão de forma explícita, tendo em conta o contexto da época, é claro que a fé faz parte desse conjunto de crenças que a informação disponível não permite justificar e que, segundo Clifford, seria errado manter. Foi um dos grandes combates da sua vida. Em jeito de resposta, William James publicou uns anos mais tarde o texto «A vontade de acreditar» (1896), no qual procurou demonstrar, a meu ver de forma pouco convincente, que em determinadas circunstâncias temos o direito de acreditar em realidades que carecem de justificação racional. Mais uma vez, é a questão da fé que surge em pano de fundo, como aliás os exemplos de James deixam transparecer. Para ele, porém, o direito de acreditar por decisão voluntária restringir-se-ia às questões que o entendimento é incapaz por si mesmo de decidir.

Mas o que me parece curioso no debate Clifford-James, é que ambos partem do princípio que é de facto possível acreditar em algo graças a um acto da vontade. Com efeito, na medida em que o *dever* pressupõe o *poder*, só faz sentido dizer que temos o dever de acreditar nisto ou naquilo caso a vontade tenha de facto o poder de determinar o conteúdo das nossas crenças. Ora, como já tivemos a ocasião de ver, tal não parece ser o caso. Contrariamente a Clifford e James, poucos filósofos e psicólogos defendem hoje a tese do voluntarismo doxástico, isto é, a ideia que podemos decidir acreditar naquilo que queremos *hic et nunc*, por simples decreto da vontade. Quando muito, há quem defenda a tese do “voluntarismo doxástico indirecto”, segundo a qual é possível induzir voluntariamente certas crenças mediante estratégias indirectas. Mas ainda assim, os métodos sugeridos parecem tão falíveis quanto difíceis de empregar. Por exemplo, Pascal sugere que quem se queira

converter ao cristianismo pode fazê-lo seguindo os mesmos rituais que as pessoas crentes, indo à missa, repetindo as rezas, fazendo as genuflexões, e tudo o mais. Ao agir como se Deus existisse, diz ele, condicionamos a nossa mente para acreditar que Deus existe realmente. Mas tenho muitas dúvidas quanto à eficácia de tal método, tanto mais que muitas pessoas que receberam educação religiosa dizem ter ido à missa por obrigação sem nunca terem acabado por acreditar nos dogmas. Davidson também dá um exemplo interessante: sugere que me posso enganar a mim mesmo relativamente a um encontro marcado com alguém de indesejável, se marcar deliberadamente uma data errada na minha agenda. Quando a data do encontro chegar, já me terei esquecido do artifício e falto ao compromisso sem sequer ficar com má consciência. Mas como fazer para encontros mais iminentes? E quem me garante a mim que me vou esquecer da minha artimanha? Outro método seria a hipnose, mas é sabido que a hipnose tem efeitos muito temporários, e o próprio Freud abandonou o método precisamente por essa razão.

A meu ver, por conseguinte, não temos nenhum poder directo sobre o conteúdo das nossas crenças, e o poder indirecto que possamos ter é muito limitado. Concordo portanto com a tese do “evidencialismo doxástico” posta em evidência por Hume, que diz que as crenças são em princípio estados mentais involuntários. Não digo que não possamos, por exemplo, evitar pensar em determinadas coisas ou tentar concentrar os nossos pensamentos naquilo que nos interessa. Mas daí a induzir precisamente determinada crença com determinado conteúdo, vai uma grande distância. Posso evitar pensar nos meus problemas de saúde quando me vou deitar, por exemplo, de modo a que essa preocupação não me impeça de adormecer, mas é evidente que isso não chega para me persuadir que não tenho problemas de saúde.

Agora, há um aspecto muito importante que gostaria de salientar. É que nem a tese do evidencialismo doxástico, nem a tese motivacional do auto-engano impossibilitam a possibilidade de um projecto de ética da crença. Com efeito, embora não tenhamos nenhum controlo sobre o conteúdo das nossas crenças, não deixa de ser verdade que temos a possibilidade de controlar até certo ponto as condições sob as quais formamos as nossas crenças. Quer isto dizer que não somos directamente responsáveis pelas nossas crenças, mas talvez sejamos responsáveis pela maneira como raciocinamos e pelos métodos que empregamos no processo de formação da crença. Na medida em que temos a possibilidade de submeter esse processo a um controlo voluntário, torna-

se legítimo argumentar que existe também esse dever, o dever de fazer aquilo que está ao nosso alcance para assegurar a verdade e a racionalidade das nossas crenças.

SM – Isso conduz-nos à questão com que termina o seu artigo recente na *disputatio*: quais seriam, de que natureza seriam, os métodos de raciocínio a que alude, capazes de aumentar a fiabilidade de formação de crenças?

VC –É uma questão fascinante, sobre a qual estou agora a trabalhar. Existe uma longa tradição filosófica sobre este tema. É, por exemplo, o que preocupa Descartes nas *Regras para a direcção da mente*, ou Espinoza no *Tratado da reforma do entendimento*, ou ainda Locke em *Sobre a conduta do entendimento*. Todos estes autores procuram determinar que tipo de métodos e regras de raciocínio são susceptíveis de assegurar a fiabilidade do juízo, e algumas das suas sugestões conservam todo o seu interesse nos dias que correm. Mas creio que os estudos empíricos dos psicólogos nas últimas décadas revelaram uma realidade que ninguém suspeitava, e que deve ser tomada em consideração: em primeiro lugar, que os fenómenos de irracionalidade cognitiva são muito mais frequentes do que se podia antecipar, afectando praticamente o conjunto da população; e em segundo lugar, que a diversidade de mecanismos de ilusão cognitiva susceptíveis de afectar o nosso juízo é, também ela, muito maior do que o antecipado. Se procurar a expressão *cognitive biases* na Wikipedia, por exemplo, verificará que os psicólogos distinguem actualmente dezenas de mecanismos de “enviesamento do juízo” capazes de distorcer a nossa leitura da realidade sem que disso nos apercebamos. Perante esta realidade, creio que uma reflexão sobre os métodos e as estratégias susceptíveis de neutralizar a irracionalidade das nossas crenças deve começar precisamente pelo estudo desses mecanismos. Só com base num “diagnóstico” rigoroso dos limites da racionalidade é que a prescrição de “métodos terapêuticos” poderá ser eficaz.

Muito concretamente, em que poderão consistir esses métodos? Tal como referi, só podemos exercer um controlo *indirecto* sobre a racionalidade das crenças. Por esse motivo, penso que as estratégias mais eficazes são as que obrigam o sujeito a impor determinados constrangimentos às condições sob as quais forma o seu juízo. Isso implica um esforço em vários aspectos. Em primeiro lugar, é possível adoptar estratégias de controlo indirecto. Por exemplo, um jornalista que receie estar a ser parcial no seu relato de um conflito, por saber que tem grande simpatia por uma das

partes, pode prometer a si mesmo rever cuidadosamente todos os elementos informativos de que dispõe antes de redigir o seu texto, de modo a evitar o efeito de “atenção selectiva” que está na origem de muitos casos de enviesamento do juízo. Mas também pode submeter o seu texto à apreciação de um dos seus colegas, de modo a detectar elementos de parcialidade que poderiam escapar ao seu próprio escrutínio. Outro aspecto que me parece importante é a aquisição de certas competências de raciocínio, graças por exemplo ao estudo das regras de inferência lógica e ao estudo das regras de argumentação. Neste momento estou a trabalhar precisamente sobre as relações causais que parecem existir entre as heurísticas e enviesamentos (*biases*), por um lado, e as falácias argumentativas, por outro. Existem muitas correlações. A ideia é que cometemos frequentemente raciocínios falaciosos sem sequer nos darmos conta disso, precisamente por causa do efeito dessas ilusões cognitivas sobre os processos cognitivos. Outro aspecto que muitos filósofos sublinham hoje em dia é a importância das “virtudes epistémicas”, que são hábitos ou disposições propícias à salvaguarda da validade dos nossos raciocínios e à verdade das nossas crenças, como por exemplo a honestidade intelectual, a perseverança e a abertura de espírito. Por fim, existem também “estratégias de compromisso prévio” que podem ser úteis neste plano. Por exemplo, a pessoa que reconheça a sua tendência para tomar decisões irracionais de forma impulsiva, devido ao efeito das emoções fortes, pode adoptar como máxima “Nunca tomar decisões importantes de cabeça quente”. Caso se mantenha fiel a esse princípio, não evita a eventualidade de o seu juízo vir a ser irracional em períodos críticos, mas evita pelo menos o impacto desse juízo irracional sobre as suas acções mais decisivas. Mas haverá sem dúvida muitas outras estratégias capazes de minimizar a nossa vulnerabilidade aos fenómenos de irracionalidade, a julgar pela diversidade desses fenómenos.

SM – Para terminar gostaria apenas de retomar alguns dos assuntos de que falámos até agora. Começemos pela acrasia. No projecto “*Conversations*” focamos a nossa atenção mais na acrasia do que no auto-engano. Uma das nossas questões de partida era precisamente uma questão acerca da natureza – na verdade da própria *possibilidade* – da acrasia, no âmbito de cada abordagem específica da racionalidade prática. Pessoalmente creio que é importante compreender estratégias de explicação, perceber o que fazemos (o que pensamos que estamos a fazer) quando ‘explicamos’ a acrasia. Ora, o Vasco defende na sua

dissertação de doutoramento que a explicação do enigma do auto-engano *contém de certo modo a chave para elucidar o enigma da acrasia*. Afirmou também que pessoalmente, não via em que é que uma moralização da questão da acrasia, que considera bastante comum na história da abordagem do fenómeno, alguma vez nos permitiria compreendê-lo melhor – também essa questão do envolvimento de uma postura moralizante na própria concepção de um fenómeno como acrático foi tratada no projecto, nomeadamente pelo meu colega Carlos Mauro (embora ele pretenda chegar a algo diferente: ele quer defender a *impossibilidade* de acrasia).

Será que podíamos regressar à questão da acrasia, procurar compreender melhor a sua posição específica, em função de tudo o resto que já avançou sobre racionalidade e irracionalidade? Seria também muito interessante que nos falasse da forma como a sua posição se relaciona com a posição de autores que entrevistámos e que conhece bem, em particular Alfred Mele e George Ainslie. Para Mele, a acrasia é auto-controle deficiente, numa situação em que (1) a força motivacional dos nossos desejos não está sempre ‘alinhada’ com a nossa avaliação dos objectos dos nossos desejos, sendo que (2) tipicamente, melhores juízos decisivos são formados em grande parte com base na nossa avaliação dos objectos dos nossos desejos. Para Ainslie a acrasia é uma falha da pessoa, actuando com base nos seus interesses de curto prazo, em proteger um interesse de longo prazo oposto: se Ulisses navega de encontro às rochas porque não se atou ao mastro, ele é acrático. (Concorda?)

VC – Se me parece que a explicação do auto-engano contém de certo modo a chave para a explicação da acrasia, é no sentido em que os dois fenómenos têm precisamente a mesma origem, que é o fenómeno de irracionalidade cognitiva motivada, isto é, a formação do juízo sob a influência de uma emoção. No caso do auto-engano, como vimos, é o efeito de uma emoção forte sobre as faculdades cognitivas do sujeito que o leva a acreditar que *p* apesar de tudo indicar o contrário. E no caso da acrasia, de forma semelhante, é o efeito de uma emoção forte ou de um desejo sobre o juízo avaliativo do agente que o leva a escolher a opção que é contrária ao seu próprio interesse. Esta explicação da acrasia contém na realidade duas teses. Em primeiro lugar, em conformidade com a análise de Ainslie, sugere que a causa da acção irracional é um fenómeno de irracionalidade cognitiva – aquilo a que Ainslie

chama o “efeito hiperbólico de depreciação”, que é *grosso modo* a tendência para preferir pequenos benefícios imediatos a benefícios futuros muito maiores, em particular quando o momento da decisão se torna iminente. Mas em segundo lugar, em conformidade com a análise de Elster, vai mais longe e sugere que essa ilusão não é “fria”, mas motivada pela influência de um desejo forte sobre o juízo do agente. Por exemplo, quando um agente decide fazer uma dieta e acaba por ceder à tentação de comer uma sobremesa suculenta, o que acontece é que o desejo de comer a sobremesa se torna extremamente forte em determinada circunstância, regra geral porque a tentação se apresenta de forma iminente, e que esse desejo acaba por afectar o juízo do agente de forma momentânea. Por um momento, ele tem a ilusão de que o benefício de comer a sobremesa é maior do que o benefício de perder uns quilos. Mas é claro que a ilusão desvanece mal a sobremesa seja consumida, uma vez que o desejo morre no momento em que é satisfeito, deixando portanto de afectar o juízo avaliativo do agente. É nesses momentos que o agente se arrepende de ter violado a sua decisão inicial. Mas Elster sublinha com razão que o agente nunca age ao invés do seu próprio juízo, dado que o seu juízo *no momento da decisão* diz de facto que comer a sobremesa é a melhor das opções. O problema é que o seu juízo é ilusório, pois representa o proveito menor como sendo maior, e é isso que causa uma inversão temporária das preferências.

Deste ponto de vista, é interessante observar que os casos comuns de acrasia não são casos genuínos de irracionalidade prática, uma vez que os agentes não escolhem *conscientemente* a pior das opções. No momento da escolha, o agente age em conformidade com aquilo que lhe parece melhor, o que é perfeitamente racional do ponto de vista prático. Nesse sentido, parece mais rigoroso dizer que se trata de casos de irracionalidade cognitiva com repercussões sobre o plano prático. A acção até pode ser dita irracional, mas somente na medida em que o juízo que a orienta é irracional. Voltamos aqui à ideia segundo a qual a compreensão dos fenómenos de irracionalidade cognitiva contém a explicação dos fenómenos de irracionalidade prática. Talvez a minha análise não se afaste muito da de Carlos Mauro, que pretende mostrar que a acrasia é impossível nos termos em que é frequentemente descrita, isto é, enquanto escolha consciente e deliberada da opção menos proveitosa. Existe, porém, a meu ver, um caso genuíno de acrasia definida nestes termos, que é o caso dos comportamentos compulsivos patológicos. O hipocondríaco que obedece à compulsão de repetir um ritual de lavagem, por exemplo, está muitas vezes consciente

de que essa acção apenas reduz o seu bem-estar geral. E no entanto, não obstante esse juízo avaliativo, é incapaz de se abster de repetir mais uma vez o ritual em questão. Assim, há uma acção que é claramente contrária ao juízo do agente no próprio momento em que é executada.

Como quer que seja, é claro que este tipo de explicação da acrasia se afasta muito da concepção do fenómeno em termos de “fraqueza da vontade”. O que parece haver é uma “fraqueza do juízo” perante a influência de um desejo. Ora, a vantagem deste tipo de análise, no meu entender, é que permite compreender o mecanismo subjacente ao fenómeno de “ceder à tentação”. Enquanto os teóricos da fraqueza da vontade esbarram com a impermeabilidade da noção metafísica de vontade, que não permite compreender ao certo de que depende a dita “força de vontade”, ou de que depende o esforço que é suposto mover o “músculo” da vontade, a teoria proposta por Ainslie e Elster, em contrapartida, permite analisar a questão em termos claros e concretos, referindo-se apenas a noções inteligíveis, como desejos, emoções, juízos, escolhas e preferências, e elucidando em termos causais os mecanismos de interacção entre estes elementos.

Mas existe outra vantagem, que tem a ver com as aplicações práticas desta teoria. De facto, a compreensão do mecanismo subjacente ao fenómeno de “ceder à tentação” torna também possível agir sobre esse mecanismo no intuito de evitar comportamentos indesejáveis. É aí que intervém a reflexão sobre o auto-controlo. Muito concretamente, só o facto de perceber que é o impacto de um desejo sobre o juízo avaliativo que motiva a acção irracional permite desde logo vislumbrar uma série de medidas capazes de evitar esse fenómeno. Por exemplo, a pessoa que decida fazer uma dieta tem todo o interesse em fazer compras no supermercado logo após uma refeição, pois nesse momento o desejo de comer é menor e, por conseguinte, terá um efeito reduzido sobre o seu juízo. Outra medida de controlo indirecto sobre si mesmo seria deixar de levar para casa alimentos altamente calóricos, de modo a não dispor de nenhuma tentação mesmo à mão de semear nos momentos de fraqueza do juízo. O obstáculo de ter de sair de casa para ir comprar um doce ou uma barra de chocolate chega muitas vezes para dissuadir. Isso corresponde àquilo a que Elster chama uma “estratégia de compromisso prévio”: a pessoa reconhece que é incapaz de resistir à tentação, e por isso evita ter à sua disposição estímulos susceptíveis de a tentar. Outro exemplo é a pessoa endividada que corta os seus cartões de crédito à tesourada, de forma a eliminar a opção futura de se endividar ainda mais. Neste

sentido, a melhor maneira de uma pessoa se ver livre de uma tentação não é “cair nela”, como diz Oscar Wilde, mas sim mantê-la fora do seu alcance. Mele insiste, por seu lado, na importância do controlo sobre as emoções. No livro *Autonomous Agents*, em particular, sugere a certa altura que temos a possibilidade de controlar em certa medida aquilo que sentimos, graças à orientação da atenção para determinados estímulos, e que, nesse sentido, parece possível evitar o efeito acrástico de determinada emoção¹⁰¹. Por exemplo, se Medeia decidir matar os seus próprios filhos para se vingar de Jasão, o pai das crianças, mas recear que o seu amor materno a impeça de levar a cabo a sua vingança, diz Mele, Medeia pode esforçar-se por focalizar o seu pensamento em tudo o que Jasão lhe fez de mal, e evitar pensar nos seus filhos, de modo a reforçar o seu desejo de vingança. Este exemplo de Mele é aliás interessante, porque a acção mais racional, aos olhos de Medeia, é a que nos parece a nós mais imoral. Mas existem muitas outras estratégias de auto-controlo que se baseiam numa compreensão adequada do fenómeno de acrasia.

SM – Nas suas respostas, a propósito da relação de Elster com o mundo intelectual francês, referiu que esse interesse visava sobretudo autores do passado, os chamados “moralistas” – autores como Montaigne, Pascal, La Rochefoucauld, Chamfort, Madame de Sablé, La Bruyère, cujas “máximas e reflexões antecipam muitos dos aspectos que a psicologia e as ciências sociais procuram demonstrar hoje em dia”. Isso fez-me pensar numa observação recorrente do meu colega Paulo Tunhas, que trabalha bastante sobre paixões na filosofia moderna e que também se interessa por alguns dos autores que mencionou: o Paulo gosta de chamar a atenção para a mudança que se deu entre esses autores e o nosso tempo, e que nem sempre percebemos claramente como tal, de um discurso sobre *paixões* para um discurso sobre *emoções*, colocadas estas num contexto cognitivo e assim supostamente neutro ou científico-natural. O ponto dele é que isto altera completamente a importância (moral e mesmo política) que o discurso sobre paixões dos moralistas franceses tinha – a forma de falar sobre paixões dos moralistas não nos deixa esquecer que estamos em pleno território moral, não nos permite pensar que podemos pensar em emoções ‘à distância’. Será isso parte do gosto de Elster pelos moralistas?

¹⁰¹ Mele, *Autonomous Agents*, Oxford University Press, 2001, pp. 64, 106 *et passim*.

VC – Acho o ponto de vista de Paulo Tunha interessante, e gostaria de o conhecer com maior detalhe. No entanto, não sei se existe realmente uma dicotomia entre o aspecto moral das emoções, por um lado, e o seu aspecto científico-natural, por outro. Certas emoções são eminentemente morais, como por exemplo a compaixão, o arrependimento, a vergonha ou o ódio, mas outras não. É o caso, por exemplo, do medo, da curiosidade, da desilusão e da esperança, que podem ter repercussões sobre o plano moral, mas que não têm nada de moral em si mesmas. Como quer que seja, parece-me que tanto as primeiras como as segundas podem ser descritas cientificamente. Os neurocientistas, por exemplo, estão interessados em saber o que se passa no cérebro quando uma pessoa sente determinado afecto, que regiões se encontram mais activas, que hormonas e neurotransmissores são produzidos, e por aí fora. E bem antes de a tecnologia permitir fazer scanners, já os cientistas se entretinham a medir o ritmo cardíaco, a frequência respiratória, os níveis de transpiração, os tremores, os calafrios, as lágrimas e outros elementos fisiológicos susceptíveis de variar de uma emoção para outra. Creio que não existe incompatibilidade entre as diferentes abordagens, e que cada uma delas revela uma faceta interessante dos estados emocionais.

No que diz respeito à passagem de um discurso sobre paixões para um discurso sobre emoções, Paulo Tunha sublinha com razão o facto de que nem sempre nos apercebemos até que ponto se mudou de paradigma. Um dos aspectos mais interessantes dessa mudança é palpável na própria etimologia dos termos empregados, uma vez que a palavra “paixão” deriva do termo latim *passio*, que significa “sofrimento”, enquanto a palavra “emoção” deriva do termo latim *moto*, que significa “movimento”. Creio que a palavra emoção é mais feliz porque a conotação de “movimento” exprime tanto o aspecto passivo ou “afectivo” das emoções, isto é, o facto que elas nos *co-movem*, como o aspecto activo ou “volitivo” das emoções, isto é, o facto que elas também nos *movem* à acção. Mas verdade seja dita, este último aspecto já era bem claro para muitos dos “filósofos das paixões”, e nomeadamente Hobbes, Espinoza e Hume, que souberam insistir sobre a dimensão volitiva de certas emoções.

SM – E isto traz-nos ao assunto sobre o qual gostaria de o ouvir mais, o assunto central no seu actual trabalho de pós-doutoramento: as emoções. Qual é

**exactamente o núcleo da teoria "emocionalista" da irracionalidade que defende?
Como é que uma tal teoria nos posiciona quando procuramos pensar sobre a
relação entre as emoções e as crenças?**

VC – O ponto de partida dessa teoria é a distinção que faz Ronald de Sousa entre três tipos de racionalidade: a racionalidade prática (ou estratégica), a racionalidade cognitiva (ou epistémica) e a racionalidade emocional (ou axiológica)¹⁰². Ora, correlativamente, também há que distinguir três tipos de irracionalidade: a irracionalidade prática, que se refere à acção contrária ao melhor juízo do agente, a irracionalidade cognitiva, que se refere à formação inadequada da crença, e a irracionalidade emocional, que se refere à ocorrência de emoções que se produzem de maneira injustificada. Aquilo a que chamo a concepção “emocionalista” da irracionalidade é simplesmente a hipótese segundo a qual cada um desses tipos de irracionalidade é em princípio causado por uma emoção. Quer isto dizer, por outras palavras, que a maioria dos fenómenos de irracionalidade tem origem, em última instância, num fenómeno emocional. No caso da irracionalidade emocional isto parece ser óbvio, visto que se trata justamente de emoções que se produzem na ausência da crença que as deveria justificar. Por exemplo, em princípio só sinto medo em relação ao objecto *p* se acreditar que *p* é perigoso, directa ou indirectamente. O medo irracional produz-se, todavia, quando sinto medo de *p* apesar de saber que *p* é perfeitamente inofensivo. É o que se passa, como já mencionei, no caso das fobias, em particular quando o indivíduo fóbico toma consciência do carácter injustificado do seu medo.

O que será porventura menos evidente é que também os casos de irracionalidade cognitiva e de irracionalidade prática derivam em geral de fenómenos emocionais. Foi contudo precisamente aquilo que vimos, embora no caso da irracionalidade cognitiva nem sempre isso aconteça, pois existem fenómenos “frios” de irracionalidade. De facto, os casos comuns de irracionalidade motivada – não apenas o auto-engano, mas também a confusão dos desejos com a realidade (*wishful thinking*), a denegação, a racionalização e todos os fenómenos “quentes” de enviesamento do juízo – resultam do efeito das emoções sobre a maneira como processamos a informação. E, por fim, também a própria irracionalidade prática parece ter origem nos nossos estados

¹⁰² De Sousa, *Evolution et rationalité*, PUF, 2004.

afectivos, sendo que é em geral uma emoção ou um desejo que perturba o juízo avaliativo do agente e causa uma inversão súbita das suas preferências. A hipótese emocionalista procura, portanto, estabelecer as bases para uma teoria unificada da irracionalidade.

Há contudo uma ressalva importante que gostaria de fazer. É que a concepção emocionalista da irracionalidade não nos deve levar a crer que as emoções são ameaças permanentes para o bom uso da razão. O facto de as emoções estarem tantas vezes na origem das nossas atitudes irracionais não nos deve fazer esquecer que também contribuem de forma significativa para a racionalidade das nossas atitudes. Em particular, as emoções são necessárias para assegurar a racionalidade das nossas escolhas. É o que revela o exame de pacientes cuja ablação de uma parte do cérebro inibe a actividade emocional, uma vez que esses pacientes têm muito mais dificuldade em tomar decisões importantes e em fazer projectos de vida, presumivelmente porque lhes falta o contributo das emoções para a avaliação das opções disponíveis. António Damásio insiste muito sobre este ponto: sem o contributo das emoções, as nossas decisões requereriam um tempo de análise e de cálculo custo/benefício incompatível com as exigências da vida quotidiana.

Mas, reciprocamente, é interessante observar que as próprias emoções parecem requerer uma certa dose de racionalidade, na medida em que os estados emocionais se fundam por sua vez sobre juízos avaliativos puramente racionais. Por exemplo, é a crença que a realização de determinado desejo é provável que justifica o meu sentimento de esperança relativamente à realização desse desejo. Ou, dito de outra forma, se eu não *acreditasse* que conseguirei provavelmente atingir esse objectivo, não *sentiria* esperança em relação a isso. É por isso que, em princípio, somos capazes de especificar *por que razão* sentimos uma determinada emoção relativamente a um certo contexto, pessoa ou acontecimento. Segundo as teorias “cognitivistas” das emoções, cada emoção requer uma crença específica para se produzir, não apenas no sentido causal, mas inclusive no sentido conceptual. Estamos portanto muito longe do preconceito que tende a encarar as emoções como uma simples ameaça para a razão. Na minha opinião, a esfera da afectividade e a esfera da racionalidade encontram-se indissolivelmente interligadas em tudo o que fazemos, sentimos ou pensamos.

A concepção inferencial da acção - Uma comparação entre a interpretação causal e a interpretação normativista (Respostas a Sofia Miguens)

João Alberto Pinto

João Alberto Pinto é Professor no Departamento de Filosofia da Universidade de Porto e investigador no Instituto de Filosofia, onde dirige a área Lógica e Filosofia do MLAG (Mind, Language and Action Group). Ensina Lógica, Filosofia da Mente e Filosofia da Linguagem. Doutorou-se com uma tese intitulada *G. Boole e as origens da lógica moderna*, na qual procurava contrastar a noção booleana de lógica (ou lógicas) como cálculo (s), com a noção fregeana – tão influente no desenvolvimento da filosofia analítica – da lógica como linguagem. É autor, entre outras publicações, de *Superveniência, Materialismo e Experiência – Uma perspectiva sobre o problema da consciência em filosofia da mente* (Porto, Campo das Letras, 2007) e coordenador de *Investigações em Filosofia da Lógica* (Porto, FL-UP, 2011).

1. Algumas distinções (relativamente) básicas

Sofia Miguens (SM) – Na sua opinião, há algumas distinções básicas que convém começar por fazer quando tentamos falar teoricamente de ‘acções’?

João Alberto Pinto (JAP) – Há pelo menos três distinções (relativamente) básicas para tentar falar em acções. Uma delas é a distinção entre acções e movimentos (e ruídos) corporais. À primeira vista, pretende-se apenas assinalar o carácter intuitivo da ideia de que há – de que se deve tentar consagrar ao nível da teoria – uma diferença entre fazer a digestão, o evitamento de um olhar, corar, etc., e acções naquele sentido que é essencial para o problema a enfrentar. De outro ponto de vista, a expressão «movimentos (e ruídos) corporais» remete directamente para a maneira de falar usada por Gilbert Ryle (*The Concept of Mind*) ao defender uma versão filosófica da posição a que se chama, em geral, «comportamentalismo» ou «behaviorismo». O problema surge quando, a partir daqui, se nota que alguém pode fazer duas perguntas muito parecidas. Uma pergunta é por que razão – se há alguma – não é desde logo usado o termo «comportamento automático» (por contraste, presume-se, com «(comportamento) voluntário»). A outra pergunta é por que razão –

se há alguma – não é desde logo usado o termo «acto reflexo» (desta vez por contraste, presume-se, com «acto (voluntário)»). Pode ser que questões acerca da melhor terminologia sejam, muitas vezes, importantes – mas mais importante é ser claro quanto à maneira de abordar o problema. Será que o problema é para ser colocado no âmbito de uma doutrina acerca do exercício – voluntário ou não – de certas faculdades? E qual seria essa doutrina – aquela que tem a sua fonte no que julgamos poder saber (de modo mais ou menos imediato) acerca de nós próprios? Ou as imagens (teóricas) – nem sempre congruentes – dessa mesma doutrina que nos são devolvidas pela Psicologia ou, por exemplo, pela Etologia?

SM – Mas como é que tudo isso se cruza com a forma como falamos vulgarmente acerca destes assuntos?

JAP – Pode-se sempre tentar começar pelas maneiras vulgares de falar. Os verbos cruciais – ou pelo menos mais gerais – parecem ser então o verbo «fazer» e o verbo «acontecer». Parece também haver um contraste entre, por um lado, «o que fazemos» e «o que fazemos acontecer» – embora neste último caso já estejam conjugados os dois verbos – e, por outro lado, «o que nos acontece». «O que nos acontece», aliás, pode ainda ser contrastado (na aparência, pelo menos) com «o que acontece». A dúvida instala-se quando se atenta noutras duas hipóteses. Uma delas é a hipótese de se falar em «o que nos acontece fazermos». A outra hipótese é a de se falar em «o que acontece fazermos». Assim, em vez dos contrastes iniciais, estar-se-ia em condições de assumir a ideia – que é mais ou menos consensual actualmente – segundo a qual acções são acontecimentos. O problema passa a ser o de ir para além dessa ideia. Ter-se-ia, em resumo e muito esquematicamente, que acções são acontecimentos <...> – mas também que o preenchimento de <...>, de maneira suficientemente específica para merecer consideração, passa a ser o objectivo essencial da teoria da acção.

SM – Continuemos então com a exploração das distinções básicas.

JAP – Uma outra distinção (relativamente) básica é a distinção entre acções e actos mentais. Desta vez pretende-se assinalar o carácter intuitivo da ideia de haver – de que se deve tentar consagrar ao nível da teoria – uma diferença entre calcular (fazer contas (de cabeça)), compreender (uma frase, um elogio ou um insulto, por exemplo),

classificar ou categorizar («mentalmente», entenda-se) objectos ou pessoas (atendendo, por exemplo, às suas características ou aos seus nomes), etc., e acções naquele sentido que é essencial para o problema a enfrentar.

A resolução de eventuais dúvidas suscitadas por esta distinção é muito mais directa do que no caso da distinção anterior. Parece – outra vez de modo mais ou menos consensual actualmente – ser suficiente o seguinte: *Y* é uma acção (e não é um acto mental) só se envolve (necessariamente) um movimento corporal (fenomenologicamente saliente); *Y** é um acto mental (e não é uma acção) só se não envolve (necessariamente) um movimento corporal (fenomenologicamente saliente). Pode-se, mesmo assim, notar que se procedeu a uma espécie de abreviatura da expressão «movimentos (e ruídos) corporais» e que se fez ocorrer, outra vez, «movimento (corporal (fenomenologicamente saliente))» em vez de «comportamento».

SM – Considera mesmo que há algo de especialmente problemático com o termo «comportamento»?

JAP – Pode-se tratar de assumir mais uma distinção (relativamente) básica – agora entre comportamentos e acções. O problema é não ser nada claro que se tem aqui uma distinção – ou, pelo menos, que se tem aqui uma distinção útil para a teoria da acção. Estará em causa uma simples opção estilística – mais ou menos terminológica, do género da que foi apontada no âmbito da anterior distinção entre acções e movimentos (e ruídos) corporais? Seria demasiada inocência presumir que sim. Assim como seria também um exagero achar que basta usar «acção» (mas não «comportamento») para fugir ao estigma que indubitavelmente e inocentemente – como notou ainda G. Ryle (*The Concept of Mind*) – afecta quase tudo que for (suspeito de) «comportamentalismo» ou «behaviorismo».

Uma hipótese mais sensata pode consistir em supor que há uma linha contínua que vai dos movimentos corporais (fenomenologicamente salientes) à noção de acção – e que a noção de comportamento, por sua vez, é para ser localizada nessa linha contínua. De umas vezes, «comportamento» – como quando, por exemplo, ocorre adjectivado em «comportamento manifesto» – há-de surgir na vizinhança dos movimentos corporais

(fenomenologicamente salientes). De outras vezes, «comportamento» – como quando, por exemplo, se opta por «comportamento motivado» – tende para muito próximo de «acção». O melhor talvez seja mesmo formular a seguinte pergunta: qual é a noção (mais) robusta de acção – aquela acepção de «acção» que se revela mais polémica ou mais problemática (menos consensual) do que a noção de comportamento?

2. A noção (mais) robusta de acção

SM – Mas tem de haver uma relação entre as questões em torno de movimentos corporais (fenomenologicamente salientes), comportamentos e acções e as questões relativas à natureza do raciocínio prático, o raciocínio acerca de como agir – sem falar nas noções de intenção, decisão e deliberação.

JAP – Há uma resposta muito directa à última pergunta formulada – e essa resposta é importante porque permite efectuar uma primeira apresentação da ideia básica da concepção inferencial da acção.

A noção (mais) robusta de acção – aquela acepção de «acção» que se revela mais polémica ou mais problemática (menos consensual) do que a noção de comportamento – envolve a forma geral de um raciocínio prático:

- (1) X tem um desejo D cujo conteúdo/significado é d.
- (2) X tem uma crença C cujo conteúdo/significado, c, é: fazer Y é a melhor maneira de obter (realizar, assegurar, alcançar) d.

Logo,

- (3) X faz Y.

A expressão «raciocínio prático» não deixa uma grande margem para dúvidas acerca da ideia básica da concepção inferencial da acção – mas vale a pena notar que essa ideia parece ter surgido com Aristóteles (*Ética a Nicómaco*) e, a partir daí, que vale a pena pensar outras duas coisas. Uma destas coisas corresponde ao reconhecimento de que há mesmo uma diferença entre a forma geral de um raciocínio prático e qualquer uma das formas daquilo a que Aristóteles chamou um raciocínio (dedutivo) demonstrativo. A outra coisa envolve o reconhecimento de que apenas (1) e (2) – conjuntamente e independentemente da sua ordem – são capazes de conduzir ao efeito (conclusivo) pretendido na forma geral de um raciocínio prático. Isto parece bastar,

por sua vez, para entender quanto – e de que maneiras – a ideia básica da concepção inferencial da acção claramente se esforça (ou, melhor, se pode esforçar) para evitar qualquer acusação de unilateralidade nos eventuais debates acerca daquilo que se chama (de modo mais ou menos tradicional) «razão» e «paixão».

Quanto ao recurso a «conteúdo/significado» em (1) e (2), vale a pena salientar que se trata de um expediente – um pouco tosco, decerto – para destacar a relação que a ideia básica da concepção inferencial da acção mantém com aquilo que, hoje em dia, se chama (i) «intencionalidade» (simplesmente ou até mesmo um pouco vagamente «intencionalidade»), (ii) «objecto intencional (semântico ou representacional)» – neste caso após a substituição de «conteúdo/significado» por «objecto», de maneira muito similar à que vigora quando se usa «representação (mental)» em vez de «objecto intencional (semântico ou representacional)» – ou (iii) «conteúdo/significado (objecto) proposicional» de uma atitude, tal como seja um desejo ou uma crença. Apesar de a origem (mais próxima) da reflexão sobre as atitudes poder ser brentiana (impondo, dir-se-ia, a predominância de «intencional» com um «c») ou fregeana (impondo, dir-se-ia, a predominância de «intensional» com um «s»), isto parece bastar para fornecer uma imagem muito geral do quadro teórico no qual a ideia básica da concepção inferencial da acção acaba por estar actualmente situada.

O passo seguinte consiste em ver como assim é fixada alguma da terminologia adicional a «acção». Trata-se de terminologia que pode ser vista como «adicional» num sentido relativamente inócuo, mas que também pode ser vista como um conjunto de termos que passam a constituir uma espécie de pontos de vista sobre a acção – ou, melhor, sobre o problema da acção. Assim, sob a forma de uma lista – relativamente ordenada – tem-se: (i) agente (sujeito da acção, X); (ii) crença (C) e desejo (D), mas também conteúdo/significado de uma crença (c) e conteúdo/significado de um desejo (d); (iii) par crença-desejo (C-D), de maneira a salientar a (crucial) ligação que se estabelece, atendendo às premissas da forma geral de um raciocínio prático (e muito especialmente a (2)), entre um certo desejo e uma certa crença; (iv) razão, sem eliminar a possibilidade de distinguir entre «uma razão pela qual X faz Y», i.e., uma razão motivadora (uma autêntica razão, dir-se-ia, correspondente àquilo que muitas vezes se chama «móvil»), e «uma razão para X fazer Y», i.e., um motivo (um mero

motivo, dir-se-ia, por contraste com o que se tem em «uma razão pela qual X faz Y»); e, por fim, (v) racionalização.

Este último termo, «racionalização», serve (primordialmente) para falar de qualquer instanciação da forma geral de um raciocínio prático – independentemente de uma tal instanciação ocorrer em primeira pessoa ou em terceira pessoa. Chama-se então «racionalização» à referência – nos termos de (1), (2) e (3) – a uma acção. O termo «racionalização», além disso, não envolve imediatamente a ideia – que talvez se pudesse encarar como sendo mais ou menos própria da ideia básica da concepção inferencial da acção – de haver sempre uma espécie de apagamento (ao nível da teoria) de qualquer «energia», «força» ou «dinâmica (inconsciente)» no agente. A ideia de a acção envolver uma ocultação ou uma desocultação de certas «forças» ou de uma espécie de «dinâmica (inconsciente)» – isto para não fazer de todo alusão a «energia» – tem ela própria algum cabimento, quer de um ponto de vista (radicalmente) naturalista, quer de um ponto de vista psicanalítico. Está, porém, longe de ser claro que tal imponha o abandono da concepção inferencial da acção – mas o ponto de vista (radicalmente) naturalista e o ponto de vista psicanalítico voltarão, adiante, a ser referidos.

Antes disso, o relevante é apontar algumas variantes – mais ou menos imediatamente plausíveis – da forma geral de um raciocínio prático. Em vez da premissa (1) pode-se ter

(1') X tem um desejo D cujo conteúdo/significado é d e não tem, simultaneamente, desejos conflitantes com D, i.e., desejos cujo conteúdo/significado é contraditório ou contrário a d – quer dizer, inconsistente com d, para recorrer a uma terminologia mais directamente lógica.

Em vez da premissa (2) pode-se ter

(2') X tem uma crença C cujo conteúdo/significado, c, é: fazer Y é – tudo considerado (ou, mais plausivelmente ainda, uma vez considerado aquilo que há para considerar) – a melhor maneira de obter (realizar, assegurar, alcançar) d.

Há uma outra variante para a premissa (2) que é particularmente interessante, não só porque ela parece ser mais fiel à letra do texto de Aristóteles, mas também porque concretiza a hipótese de haver uma espécie de operador lógico (prático) envolvido na forma geral de um raciocínio prático. Assim, em vez da premissa (2), pode-se ter

(2'') X tem uma crença C cujo conteúdo/significado, c, é: *prima facie* fazer Y é a melhor maneira de obter (realizar, assegurar, alcançar) d – sendo o operador «*prima facie*» interpretado por meio de dois componentes: (i) um princípio geral (que se diz ele próprio *prima facie*, i.e., à primeira vista ou, melhor ainda, salvo conflito com algum outro princípio geral) acerca do que é adequado (ou não) fazer; (ii) um juízo que respeita à verificação (ou não) das circunstâncias de aplicação do anterior princípio geral.

Atente-se agora na ausência, até este momento, do termo «intenção» e do termo «decisão» (ou «deliberação»). Trata-se de uma ausência que é, evidentemente, relevante – e que deve ser explicitada pela afirmação de que as premissas de uma racionalização são encaráveis também como uma análise (pelo menos) satisfatória da noção (vulgar) de intenção e da noção (vulgar) de decisão (ou deliberação). A perspectiva assim assumida – mesmo que ela não seja, também evidentemente, a única disponível – possui duas consequências mais ou menos notáveis pela economia teórica que proporcionam. A primeira consequência é a de que todas as acções se tornam então – ou simplesmente por definição – (acções) intencionais. A segunda consequência é a de que intenções e decisões (ou deliberações) não têm de ser vistas (ao nível da teoria) como constituindo estados adicionais àqueles que são as razões – ou, um pouco mais especificamente, como estados diferentes daqueles que são as crenças e os desejos articulados no âmbito das próprias racionalizações. O mais interessante é como esta última consequência também torna claro que há uma maneira de recorrer à forma geral de um raciocínio prático para, dir-se-ia, interpor, entre a própria forma geral de um raciocínio prático e a acção, as intenções – ou pelo menos as intenções, numa certa acepção de «intenção» que pode bem corresponder apenas a uma espécie de versão (parcial) da noção (vulgar) de «intenção». De qualquer modo, para obter o efeito pretendido, em vez da conclusão (3), ter-se-ia agora

(3*) X forma a intenção de (tenciona, pretende) fazer Y.

As anteriores duas considerações económicas, bem como a tentativa de permanecer fiel (pelo menos) ao texto de Aristóteles, talvez sejam razão suficiente para não se optar pela modificação drástica da forma geral de um raciocínio prático – ou, melhor, da própria ideia básica da concepção inferencial da acção – que (3*) constituiria.

Em vez da exploração da perspectiva representada por (3*) retome-se antes a ideia de que há uma diferença entre variantes e modificações drásticas da forma geral de um raciocínio prático. O objectivo imediato é o de observar que se a variante (1') parece não acarretar qualquer alteração no que respeita à conclusão, (3), da forma geral de um raciocínio prático, o mesmo não se passa com as variantes (2') e (2''). Estas duas últimas variantes sugerem que – em vez da conclusão (3) – se tenha, respectivamente, (3') *ceteris paribus*, i.e., mantendo-se tudo (ou, mais plausivelmente ainda, aquilo que há para considerar) igual, X faz Y;

(3'') *prima facie* X faz Y.

Ora, (3') e (3'') não constituem – pelo menos por contraste com (3*) – uma modificação drástica da forma geral de um raciocínio prático. Pode-se até pensar que constituem, pelo contrário, uma espécie de confirmação da flexibilidade teórica da própria forma geral de um raciocínio prático.

SM – De acordo com a sua hipótese sobre a noção (mais) robusta de acção, um agente ser levado à acção envolve sempre um raciocínio prático – e quando não envolve, há lugar para dúvidas, quer sobre ele ser (ter sido) um agente, quer sobre o acontecimento em causa ser (ter sido) uma acção. Para além de (3*) que outras maneiras vê para escapar à noção (mais) robusta de acção?

JAP – O ponto essencial pode ser conferido – ainda que apenas por meio de mais contrastes – de uma outra maneira. Em primeiro lugar, considere-se o resultado de levar a sério a sugestão segundo a qual, em vez da conclusão (3), se deve ter (3**) provavelmente X faz Y – similarmente: assumindo (1) ou (1') e assumindo (2), (2') ou (2''), a probabilidade de X fazer Y é maior do que a probabilidade de X não fazer Y, ou fazer não-Y.

A consequência imediata da sugestão é tornar a forma geral de um raciocínio prático numa espécie de esquema para raciocínios de tipo – apenas ou exactamente – indutivo. Tudo se complica bem mais quando se pensa na concretização desta sugestão pela (actual) teoria dinâmica da acção que tem origem directa na obra, publicada em 1970, por John W. Atkinson e David Birch (*The Dynamics of Action*). No âmbito da (actual) teoria dinâmica da acção, a acção – ou, melhor, a selecção de uma certa acção – é para ser encarada como uma propriedade emergente de sistemas dinâmicos, i.e., governados (ao nível dos modelos envolvidos, inclusivamente para

efeitos de simulação computacional, onde os agentes tanto podem ser utilizadores de uma biblioteca, como simples formas coloridas, digamos) por sistemas de equações diferenciais. A analogia (conceptual e formal) entre o ponto de vista (radicalmente) naturalista acerca da acção que assim seria adoptado e a teoria (propriamente) física dos sistemas dinâmicos é tão forte que uma das variáveis consideradas nas equações da (actual) teoria dinâmica da acção (como uma propriedade básica, portanto) é precisamente a inércia – ou, mais especificamente, a inércia (ao longo de uma dimensão temporal) das tendências de acção. As tendências de acção também são elas próprias concebidas como instigadas ou inibidas – tendo-se, no limite do caso da inibição, aquilo a que então se chama «negacção» – por intermédio de forças de diversos géneros.

Em segundo lugar, considere-se uma outra sugestão. A sua inspiração vem de Otto Neurath (*“The Lost Wanderers of Descartes and the Auxiliary Motive (On the Psychology of Decision)”*) e apela à ideia de que, em condições de incerteza, a melhor maneira (a maneira mais honesta, na formulação original) de obter (realizar, assegurar, alcançar) d pode bem ser aleatória (recorrer a um procedimento, dito «auxiliar», do tipo «cara-ou-coroa», ainda na formulação original). Para especificar as condições de incerteza (de modo endógeno ao agente) basta assumir – contra o espírito, dir-se-ia, da variante (1’) mais acima – a hipótese de serem pouco frequentes as situações em que X tem um desejo D cujo conteúdo/significado é d e não tem, simultaneamente, desejos conflitantes com D. Neste caso, a forma geral de um raciocínio prático corre o risco de ser vista como drasticamente omissa acerca do papel que o acaso pode ter na acção. Aliás, um efeito muito similar ao anterior também é obtido se, em vez da premissa (2), se tiver

(2*) X tem uma crença C cujo conteúdo/significado, c, é: fazer Y é uma maneira de obter (realizar, assegurar, alcançar) d.

O suporte da conclusão da forma geral de um raciocínio prático pelas premissas perde, agora, quase toda a força (dedutiva ou indutiva). Ao mesmo tempo, parece ser devido algum esclarecimento acerca da natureza das maneiras envolvidas, de modo não qualificado, em (2*) – isto se ainda se quiser manter alguma referência à forma geral de um raciocínio prático. Mas, deixando de lado qualquer elaboração acerca do papel que o acaso pode ter (ou não pode?) na acção ou acerca da natureza das maneiras (aleatoriamente?) envolvidas, de modo não qualificado, em (2*), o passo

seguinte consiste em levar a sério o ponto de vista psicanalítico. Trata-se de um ponto de vista – de acordo com Sigmund Freud (“*O Ego e o Id*”) – segundo o qual não são mesmo pouco frequentes as situações em que X tem um desejo D cujo conteúdo/significado é d e tem, simultaneamente, desejos conflitantes com D. O mais apropriado é até pensar que constitui uma pretensão essencial do ponto de vista psicanalítico a ideia de que situações desse género constituem necessariamente – são a melhor descrição teórica – da dinâmica (inconsciente) de um agente ou, mais exactamente, de um agente humano. Ora, se assim for, ter-se-ia – explicitamente contra a variante (1’) – algo como

(1*) (1’) é (necessariamente) falsa.

Quanto à própria premissa (1) da forma geral de um raciocínio prático, ela poderia passar a ser vista, no máximo, como trivial – ou, de outro modo, como resultante de uma «racionalização» no específico sentido que este termo possui a partir do próprio ponto de vista psicanalítico e em que se tem uma (mera) reconstrução (defensiva, reactiva ou, no caso patológico, sintomática).

Uma vez que um apelo a intuições pessoais para concluir que há algo de errado com a forma geral de um raciocínio prático enfrenta o problema de não ser fácil saber se tais intuições – mesmo que elas estejam rigorosamente contadas – são a coisa mais bem distribuída (ou, pelo contrário, mal distribuída) do mundo, qualquer conclusão pode bem ter de ser muito mais modesta. Ela limitar-se-ia a notar que a concepção inferencial da acção permite o tratamento de muitos temas que estão (intuitivamente ou, até mesmo, tradicionalmente) ligados à acção – e, uma vez que nessa concepção está crucialmente envolvido um raciocínio, vale a pena notar que talvez seja mesmo de admitir o uso da palavra «racionalista» em vez de «inferencial». O interessante é, porém, que o mesmo já não se passa com a palavra «racional». A palavra «racional» suscita o problema de saber se o uso dela pode (ou não) ficar reservado para qualificar a própria acção (Y) ou o próprio agente (X) – e também se, deste modo, se corre (ou não) o risco de criar apenas uma espécie de redundância (mais ou menos indesejada). Trata-se de um problema que requer a análise – com maior detalhe – das águas muito profundas em que se encontra mergulhada a noção (mais) robusta de acção.

3. A dúvida

SM – Voltemos então à questão central no nosso projecto, a questão das relações entre acção e racionalidade, e procuremos perspectivá-la do ponto de vista da forma geral de um raciocínio prático – de que falamos exactamente?

JAP – A forma geral de um raciocínio prático está sujeita a duas interpretações. Trata-se de duas interpretações que têm a sua origem em – ou que podem tomar por quadro de referência com vista à sua explicitação – algumas posições relativamente bem determináveis.

A primeira interpretação é aquela interpretação para a qual as instanciações da forma geral de um raciocínio prático são explicações causais – ou explicações de tipo causal – de acções. Cabe, assim, falar na existência de

(I) uma concepção inferencial causal (realista) da acção.

A palavra «realista» pode ser encarada como uma mera variante, mas também pode ser encarada como uma especificação – não apenas de cariz terminológico – para a formulação da posição que se tem em vista. «Realista» quererá então dizer, mais exactamente, que para uma concepção como (I) há-de haver pelo menos uma racionalização que constitui uma explicação causal – ou uma explicação de tipo causal – de uma acção. A seguir, pode-se também perguntar qual é o problema específico que, de acordo com uma concepção como (I), tende a requerer (primordialmente) uma solução. A resposta à pergunta é clara – e remete muito directamente para o trabalho de Donald Davidson (“*How is Weakness of the Will Possible?*”). O problema que requer (primordialmente) uma solução numa concepção como (I) é

o problema da irracionalidade (ocasional) das acções ou dos agentes – problema esse que se concretiza (mais ou menos intuitivamente) nos fenómenos de fraqueza da vontade ou incontinência («akrasia», para usar a palavra grega) dos agentes.

Na base do problema está uma determinada suposição. Trata-se da suposição segundo a qual a acção envolve um certo princípio de continência interno aos agentes:

(PC) X realiza a acção que julga, com base em todas as razões relevantes disponíveis, ser a melhor.

Assumindo que X fala, nalgum sentido, consigo próprio – ou que X procede nessas ocasiões de maneira (mais ou menos) similar às situações em que fala com (ou de) outros agentes:

(PC') realiza a acção que julgas, com base em todas as razões relevantes disponíveis, ser a melhor.

Ainda assim, os pormenores envolvidos na explicitação de uma concepção como (I) encontram-se associados à forma geral de um raciocínio prático. Ou, mais precisamente, a determinada interpretação da forma geral de um raciocínio prático.

Tomando, outra vez, a forma geral de um raciocínio prático por referência, deve-se pensar que, na base de uma concepção como (I), estão dois aspectos. O primeiro aspecto é constituído pela ideia segundo a qual

(I.a) a conclusão de uma racionalização é destacável (relativamente às premissas em causa nessa racionalização).

Este aspecto encontra-se, por sua vez, articulado com um outro. Aquele que é constituído pela ideia – primeiramente avançada também por D. Davidson (*“Actions, Reasons and Causes”*), em 1963 – segundo a qual

(I.b) há (ou há no máximo) uma conexão (contingente) causal ou empírica – com regularidade (observacional) e conjunção (mais ou menos) constante, à maneira humeana – entre razões e acções.

Deste ponto de vista, uma conexão (contingente) causal ou empírica deve sempre assumir carácter nomológico, i.e., estar articulada com uma lei ou, de maneira similar, articular (agora no sentido de «descrever») acontecimentos sob a forma de uma lei. Mas a noção de lei que aqui está presente remete, antes de tudo, para uma legalidade natural – ou, de maneira menos drástica, para uma legalidade tal e qual como a legalidade natural. Trata-se de uma legalidade que pode perfeitamente encarar-se como sendo de tipo estatístico e, para além disso, como absolutamente isenta de excepções – no sentido específico em que, dado um enunciado singular de causalidade relacionando dois acontecimentos, não tem cabimento supor-se que há (necessariamente) uma lei que se refere a esses acontecimentos de acordo com a descrição que deles consta no próprio enunciado, mas apenas tem cabimento supor-se que há uma lei que se refere a esses acontecimentos de acordo com alguma descrição deles.

A conjugação dos dois anteriores aspectos – no âmbito de uma concepção como (I), na qual as razões são para ser encaradas como causas e as acções são para ser encaradas como efeitos, embora as acções também sejam destacáveis das razões – acaba por tornar proeminente o seguinte condicional:

(I.c) se, dadas instâncias de (1) e (2), X não faz Y, ou faz não-Y, então X (não-Y) é para ser encarado (encarada) como irracional.

A ocorrência de «irracional», imediatamente acima, obriga a proceder a algumas observações. Pode-se começar por dizer que «irracional» significa, em geral, «sem razão» – mas deve-se acrescentar que «sem razão», por sua vez, aponta para (pelo menos) três ideias novas.

Uma das ideias novas consiste em pensar que a situação a que se chegou sugere a hipótese (empírica) de que há um problema com X ou, talvez melhor, um problema com o processamento cognitivo, dir-se-ia, do conteúdo/significado das crenças e desejos de X – isto na suposição de que X se tinha revelado já, noutras ocasiões e de maneira fidedigna, como um agente racional. Há, no entanto, uma versão ligeiramente diferente desta ideia. De acordo com esta versão, a hipótese a fazer assume natureza muito mais conceptual do que a anterior hipótese (empírica). Suponha-se que a racionalização, aquela que se tornou questionável, ocorre em primeira pessoa – embora, no caso de se assumir que não há qualquer assimetria relevante entre racionalizações em primeira pessoa e em terceira pessoa, a suposição não seja realmente essencial. A hipótese que tem cabimento é agora a hipótese de que X não fazer Y, ou fazer não-Y, tem origem num específico fenómeno mental (e não apenas cognitivo, dir-se-ia) que ocorre ainda – quer dizer, tal como na versão inicial da ideia – com X. O termo proposto por D. Davidson (“Deception and Division”) para o fenómeno mental em causa é «auto-engano» – e para lidar com a possibilidade de auto-engano, X (a mente de X) acaba por ter de ser concebido (concebida) como dividido (dividida). A analogia mais profunda que deste modo se revela é evidentemente a analogia entre, por um lado, alguém enganar (acerca das suas razões) outrem e, por outro lado, alguém se enganar (acerca das suas razões) a si próprio.

Uma outra ideia nova consiste em pensar que a situação a que se chegou sugere que um certo – note-se – comportamento de X, acima dito «não-Y», não possui afinal uma razão como sua causa e que, sendo assim, esse comportamento de X simplesmente não pode contar como uma acção (um acto) de X, muito menos como uma acção (um acto) racional de X. A sugestão, no seu essencial, é então a de que pode ser relevante manter uma noção de comportamento para além – ou aquém, claro – da noção de acção (acto). O problema com a sugestão é o de que parece haver uma espécie de instabilidade – uma espécie de jogo aparentemente apenas terminológico – na

equação que assim seria proposta entre «comportamento» e «acção (internamente) irracional».

Mas há, ainda, uma terceira ideia nova. À primeira vista, a ideia consiste em pensar apenas que algo – aquilo mesmo que, de acordo com a anterior ideia, seria um comportamento, mas não uma acção (um acto), de X – permanece (ou não pode senão permanecer) inexplicado. No entanto, atendendo aos dois aspectos que estão na base de uma concepção como (I) – (I.a) e (I.b), mais acima – o que há para considerar é diferente. O que há para considerar é a possibilidade de um acontecimento ser pura e simplesmente inexplicável à luz de razões – ainda que não o seja no âmbito de um apelo a causas de um género diferente daquele que está presente quando se fala em razões ou quando se fala, dir-se-ia também, à luz (anómala, para usar um termo caro a D. Davidson) de razões. O que se passa é que essas causas podem, finalmente, ser encaradas como cognitivas – de acordo com a primeira hipótese (empírica) esboçada – ou como físicas (neurofisiológicas).

Tudo isto permite esclarecer, em resumo, três pontos. O primeiro ponto assinala que no âmbito de uma concepção inferencial causal (realista) da acção as razões são pensadas como – identificadas com – causas específicas (de acordo com a última das ideias novas atrás apresentadas). O segundo ponto assinala que no âmbito de uma concepção inferencial causal (realista) da acção é reivindicada a distinção – mais ou menos nítida, mais ou menos interna aos agentes – entre racionalidade e irracionalidade. O terceiro ponto consiste, por último, na ideia de que para uma concepção inferencial causal (realista) da acção a pergunta mais geral pela natureza da acção tem a seguinte resposta:

[A.I] o que faz um acontecimento (um movimento corporal (fenomenologicamente saliente)) ser uma acção é extrínseco ao acontecimento, i.e., é ele ser causado por razões do agente de uma maneira apropriada.

Note-se apenas que a cláusula «de uma maneira apropriada» é crucial se, como parece apropriado, não se quiser contar como uma acção (de X), por exemplo, o tropeção (de X) numa pedra daquele caminho que determinadas razões (de X) conduziram – de maneira apropriada, claro – X a percorrer.

SM – E quanto à segunda interpretação da forma geral de um raciocínio prático?

JAP – A ideia geral da segunda das duas interpretações da forma geral de um raciocínio prático é a ideia de que as racionalizações – se são (ainda) explicações da

acção – são-no (apenas) num sentido de «explicação» bem diferente do sentido predominante no âmbito de uma concepção inferencial causal (realista) da acção. De modo a determinar, pelo menos um pouco mais, a formulação da posição que se tem em vista, cabe falar na existência de

(II) uma concepção inferencial normativista (não-causal) da acção.

Uma breve revisão de alguns pontos da literatura deve tornar mais clara a diferença entre uma concepção como (I) e uma concepção como (II).

Os primeiros e principais defensores de uma concepção como (II) são Charles Taylor (*The Explanation of Behaviour*) e William Dray (*Laws and Explanation in History*). Ambos estiveram empenhados, logo a seguir a 1963, na polémica contra a posição que tinha sido assumida por D. Davidson (“*Actions, Reasons and Causes*”) e que corresponde – no seu essencial – a uma concepção como (I). A formulação mais drástica do princípio condutor que subjaz a uma concepção como (II), no caso de C. Taylor, é a de que há mesmo acontecimentos – as acções, precisamente – que constituem resultados não causais (de outros acontecimentos, presume-se). Uma origem próxima deste princípio condutor talvez se encontre nalgumas passagens de Ludwig Wittgenstein (*Investigações Filosóficas*; §357, por exemplo) onde parece ser feita uma distinção entre duas questões diferentes. Por um lado, há uma questão acerca da génese de acções a partir de algo que poderia ser encarado como comportamento (eventualmente pessoal) sujeito a observação. Por outro lado, há uma questão acerca do papel que esse mesmo comportamento (eventualmente pessoal) sujeito a observação necessariamente desempenha na semântica das atribuições de acções (ou, na expressão de L. Wittgenstein, em atribuições de «alma»). A primeira questão seria, no máximo, uma questão empírica – «no máximo» porque parece duvidoso que ela tenha cabimento tanto de um ponto de vista intuitivo, como do ponto de vista conceptual.

Atendendo ao caso de W. Dray, o princípio condutor que subjaz a uma concepção como (II) admite outra formulação. Nesta formulação, tem-se – talvez mais directamente – um princípio acerca da heterogeneidade entre ciências:

(PHC) de um lado, as ciências humanas (ou (propriamente) hermenêuticas); de outro, as ciências naturais (de interpretação (não hermenêutica, presume-se) da natureza).

A ideia básica remonta evidentemente (pelo menos) a Wilhelm Dilthey (*Origens da Hermenêutica*). A suposição crucial para a defesa dessa ideia foi a de haver uma diferença entre compreender e explicar («*Verstehen*» e «*Erklären*», com os termos

originalmente usados por W. Dilthey). Tratava-se, num caso, de compreender manifestações (signos ou expressões, um termo que é manifestamente importante também em L. Wittgenstein) psíquicas e, no outro caso, de explicar – causalmente – fenómenos naturais (outra vez com os termos usados por W. Dilthey). Note-se como, deste ponto de vista, se entende bem que, se uma racionalização puder ser (ainda) vista como uma explicação de alguma acção, ela também tem de ser encarada como explicação de um género muito peculiar.

A peculiaridade em causa é melhor determinada ao considerar-se, mais uma vez, a posição de D. Davidson. O ponto de partida pode ser a constatação de que D. Davidson (“*Mental Events*”) acabou por incorporar na defesa da sua posição a referência (mais ou menos continuada) a um ideal de racionalidade constitutivo. No entanto, para esclarecer a diferença entre uma concepção como (I) e uma concepção como (II), há que tratar de perceber melhor o que é o ideal de racionalidade constitutivo a que D. Davidson se refere. Por um lado, ele parece não coincidir com o princípio de continência – uma vez que este princípio é concebido, quase sempre, como interno aos agentes. Por outro lado, atendendo à própria formulação usada por D. Davidson, parece também correcto pensar-se que o ideal de racionalidade constitutivo governa – ou governa especificamente – qualquer tarefa (desenvolvida em primeira ou em terceira pessoa) de atribuição de atitudes, i.e., de crenças e desejos, mas não governa – ou não governa exactamente – a forma geral de um raciocínio prático na sua totalidade. Por isso mesmo, a conclusão de uma racionalização foi – sempre – encarada por D. Davidson como destacável (relativamente às premissas em causa nessa racionalização).

O passo seguinte consiste em ver se o ideal de racionalidade constitutivo admite um uso – ou um efeito (ao nível da teoria) – mais restrito ou menos restrito. No caso afirmativo, uma concepção como (II) poderia caracterizar-se precisamente pelo uso – ou pelo efeito (ao nível da teoria), claro – menos restrito do ideal de racionalidade constitutivo. Em vez de «menos restrito» pode mesmo ter cabimento pensar que o ideal de racionalidade constitutivo acaba por ter um uso – ou um efeito (ao nível da teoria), outra vez – absolutamente irrestrito (generalizado) apenas no âmbito de uma concepção como (II). A origem mais próxima deste último ponto parece remontar a 1956. Nessa altura, Wilfrid Sellars (“*Empiricism and the Philosophy of Mind*”) defendeu – a partir do contraste entre descrições empíricas e razões (no sentido epistemológico mais geral) – que qualquer forma de empirismo estava comprometida

com um mito do dado (na terminologia de W. Sellars), acabando por ter de ignorar a normatividade que essencialmente constituía o espaço lógico das razões (ainda com a expressão usada por W. Sellars, onde aliás se pode pensar que «razões» ocorre em vez de «factos» como acontecia na primeira fase do pensamento de L. Wittgenstein). Uma vez aceite este diagnóstico do empirismo, bem como a (contrastante) caracterização normativa de um espaço lógico das razões, a ideia ter-se-ia estendido da epistemologia para a teoria da acção – quer naquele uso mais restrito do ideal de racionalidade constitutivo que D. Davidson parece preferir, quer naquele uso absolutamente irrestrito (generalizado) do ideal de racionalidade constitutivo que caracteriza uma concepção como (II). Aliás, deste ponto de vista, torna-se completamente expectável que uma concepção como (I) e uma concepção como (II) coincidam quanto a uma espécie de tom racionalista – pelo menos por contraste com um outro tom, mais empirista, que permaneceria empenhado em salientar o envolvimento nas racionalizações de processos de género compreensivo (não (exactamente) explicativo) ou, até mesmo, empático.

Cabe agora encontrar – de um modo similar (mas não mais do que similar) ao utilizado para lidar com uma concepção como (I) – o problema para o qual uma concepção como (II) tende a ser encarada (primordialmente) como uma solução. Eis uma tentativa de resposta à pergunta:

o problema da racionalidade (generalizada) das acções ou dos agentes – problema esse que se concretiza (mais ou menos intuitivamente) nos fenómenos de desaparecimento (potencial, pelo menos) da irracionalidade das acções e dos agentes.

Deixando de lado o envolvimento nas racionalizações de processos de género compreensivo (não (exactamente) explicativo) ou, até mesmo, empático, os pormenores envolvidos na articulação de uma concepção como (II) continuam a encontrar-se – tal como no caso de uma concepção como (I) – associados à forma geral de um raciocínio prático. Ou, mais precisamente, a determinada interpretação da forma geral de um raciocínio prático.

SM – Uma questão constantemente presente no nosso projecto dizia respeito a razões não apenas na relação com o raciocínio prático mas também com explicação – será que pode explicar-me como é que a sua posição o faz ver as coisas quanto a este ponto?

JAP – Tomando, mais uma vez, a forma geral de um raciocínio prático por referência, deve-se pensar que, na base de uma concepção como (II), estão dois aspectos. O interessante é como estes dois aspectos constituem praticamente uma imagem invertida daqueles que estavam na base de uma concepção como (I). O primeiro aspecto é constituído pela ideia segundo a qual

(II.a) a conclusão de uma racionalização não é destacável (relativamente às premissas em causa nessa racionalização).

O segundo aspecto concretiza-se pela substituição da conexão (contingente) causal ou empírica entre razões e acções que figurava numa concepção como (I). Em vez dela tem-se agora a ideia segundo a qual

(II.b) há uma conexão (necessária) conceptual ou lógica (com regularidade constitutiva e dependência recíproca) entre razões e acções.

Note-se que está disponível a recusa pura e simples do termo «explicação» – e, nesse caso, pode-se preferir «descrição» (mas não «descrição empírica») para falar daquilo que uma racionalização (em primeira pessoa ou em terceira pessoa) torna possível efectuar relativamente a uma acção. Atendendo também a que no âmbito de uma concepção como (II) é a própria articulação entre razões e acções (e não apenas entre crenças e desejos) que se encontra submetida ao ideal de racionalidade constitutivo, tem-se o seguinte condicional:

(II.c) se, dadas instâncias de (1) e (2), X não faz Y, ou faz não-Y, então não-Y é para ser encarada como explicada (descrita) por uma razão (de X) diferente daquela que ocorre nas anteriores instâncias de (1) e (2).

Nesta situação, uma racionalização não é derrotada (primordialmente) pela intervenção de causas – cognitivas ou físicas (neurofisiológicas). O que se passa é que o próprio ideal de racionalidade constitutivo está a ser encarado como essencialmente iterável – e que uma nova aplicação desse ideal pode fornecer sempre uma explicação (descrição) de X não fazer Y, ou fazer não-Y, por intermédio do apelo a outra razão (de X), i.e., no âmbito de uma outra racionalização. Assim, enquanto numa concepção como (I) a distinção entre «racional» e «irracional» é atingida de modo relativamente directo, numa concepção como (II) é a própria distinção entre «racional» e «irracional» que acaba por ser encarada como (extremamente) problemática. O uso irrestrito (generalizado) do ideal de racionalidade constitutivo parece apenas poder conduzir a situações em que vigora uma espécie de conflito entre várias versões de uma acção – as quais serão todas racionais no sentido de terem sido geradas de acordo

com o ideal de racionalidade constitutivo. Uma outra maneira de entender a diferença entre uma concepção como (I) e uma concepção como (II) é pensar que, no primeiro caso, o conflito assume uma forma predominantemente interna a agentes submetidos ao princípio de continência, uma vez encontrada a articulação entre as crenças e os desejos que figuram nas premissas da forma geral de um raciocínio prático. No segundo caso, i.e., no âmbito de uma concepção como (II), o conflito pode extravazar para o exterior dos agentes. No primeiro caso, as anomalias da acção (ou do «mental» no seu todo) tendem a ser vistas como individuais ou como o produto de uma teoria que, com uma ideia ainda devida a D. Davidson, simplesmente não pode ser aperfeiçoada (ao contrário do que acontece com a Física (ou com a Neurofisiologia)); no segundo caso, as anomalias da acção (ou do «mental» no seu todo) podem ser vistas mais facilmente como possuindo natureza social – nomeadamente se não se quiser invocar uma autoridade do que ocorrer (apenas) em primeira pessoa – e não há qualquer problema com teoria nenhuma (nem com as da compreensão, nem com as da explicação, para voltar aos termos de W. Dilthey).

Parece, enfim, admissível resumir o contraste entre uma concepção como (I) e uma concepção como (II) de modo paralelo ao usado, mais atrás, para falar de uma concepção como (I). No âmbito de uma concepção inferencial normativista (não-causal) da acção as razões não são para ser encaradas como – não podem ser identificadas com – causas; a distinção entre racionalidade e irracionalidade perde a sua rigidez, designadamente no que respeita ao termo «irracionalidade», pois a própria relação entre razões e acções é gerada (constituída) à luz de um ideal de racionalidade (e não de irracionalidade, claro); por último, em resposta à pergunta mais geral pela natureza da acção vem:

[A.II] o que faz um acontecimento (um movimento corporal (fenomenologicamente saliente)) ser uma acção é intrínseco ao acontecimento, i.e., é ele ser um tipo especial – normativo – de acontecimento.

Note-se apenas que a ideia segundo a qual há um tipo especial – normativo – de acontecimentos parece imediatamente ajustada aos casos, mais ou menos frequentes, em que o julgamento (até mesmo na acepção jurídica do termo «julgamento») de uma acção como «racional» ou «irracional» oscila muito mais do que uma concepção como (I) parece esforçar-se por consagrar.

SM – E é tudo?

JAP – Talvez não. Pode ter de se considerar (pelo menos) uma outra resposta à pergunta mais geral pela natureza da acção.

[A.III?] O que faz um acontecimento (um movimento corporal (fenomenologicamente saliente)) ser uma acção é intrínseco ao acontecimento, i.e., é ele ser um tipo especial – epistémico (mas não dependente de observação) ou (pura e simplesmente) fenomenal – de acontecimento.

Bibliografia

- Aristóteles. *Ética a Nicómaco*. (Trad., int. e notas: António de Castro Caeiro.) Lisboa, Quetzal, 2009.
- Atkinson, John W.; Birch, David. *The Dynamics of Action*. New York, Wiley, 1970.
- Davidson, Donald. “Actions, Reasons and Causes” (1963). Em *Essays on Actions and Events*. New York, Oxford University Press, 2001, p. 3-19.
- Davidson, Donald. “How is Weakness of the Will Possible?” (1969). Em *Essays on Actions and Events*. New York, Oxford University Press, 2001, p. 21-41.
- Davidson, Donald. “Mental Events” (1970). Em *Essays on Actions and Events*. New York, Oxford University Press, 2001, p. 207-225.
- Dilthey, Wilhelm. *Origens da Hermenêutica*. Em *Textos de Hermenêutica*. Ed.: R. Magalhães. Porto, Rés, 1984, p. 147-203.
- Dray, William. *Laws and Explanation in History*. Oxford, Clarendon, 1957.
- Freud, Sigmund. “O Ego e o Id” (1923). Em *Textos Essenciais da Psicanálise – III – A Estrutura da Personalidade Psíquica e a Psicopatologia*. Ed.: A. Freud. (Rev. e notas: José Gabriel Pereira Bastos.) Lisboa, Europa-América, 1989, p. 10-68.
- Neurath, Otto. “The Lost Wanderers of Descartes and the Auxiliary Motive (On the Psychology of Decision)” (1913). Em *Philosophical Papers 1913-1946*. Ed.: R.S. Cohen; M. Neurath. Dordrecht, Reidel, 1983/1913, p. 1-12.
- Ryle, Gilbert. *The Concept of Mind*. London, Hutchinson, 1949.
- Sellars, Wilfrid. “Empiricism and the Philosophy of Mind”. Em *The Foundations of Science and the Concepts of Psychoanalysis, Minnesota Studies in the Philosophy of Science, Vol. I*. Ed.: H. Feigl; M. Scriven. Minneapolis, University of Minnesota Press, 1956, p. 253-329.
- Taylor, Charles. *The Explanation of Behavior*. London, Routledge & Kegan Paul, 1964.
- Wittgenstein, Ludwig. *Investigações Filosóficas*. Em *Tratado Lógico-Filosófico. Investigações Filosóficas*. (Trad. e pref.: M. S. Lourenço.) Lisboa, Fundação Calouste Gulbenkian, 1995, p. 159-611.

Parte II

Ética, Juízos e Objectividade

What some philosophers wouldn't dream of counting as part as their job – Cora Diamond, a filosofia moral e a literatura¹⁰³

Sofia Miguens

O título deste artigo (“What some philosophers wouldn't dream of counting as part as their job”) refere-se simplesmente à literatura e é em parte uma citação, extraída de um artigo de Cora Diamond intitulado «*Having a rough story about what moral philosophy is*» (ela fala em ‘*subject*’, disciplina). Um problema que desde logo se põe é que uma das coisas que Diamond defende nesse artigo é que não chegaremos a uma visão consensual [*widely agreed*], da disciplina, i.e. da filosofia moral, nem presume-se, do trabalho do filósofo moral¹⁰⁴. No entanto Diamond tem a sua própria “história em traços largos” [*her own rough story*], e é essa concepção que ela põe a trabalhar num conjunto de artigos sobre ética¹⁰⁵. Ora, na história que Diamond tem para contar acerca de filosofia moral definitivamente a literatura está envolvida. Neste artigo gostaria de procurar compreender exactamente como é que a visão wittgensteiniana da ética professada por Diamond pode fazer uma diferença tão grande no que diz

¹⁰³ Este artigo corresponde à comunicação apresentada em Setembro de 2010 no *Colloque d'hommage de la communauté philosophique européenne à Cora Diamond* (Université de Picardie - Jules Verne/École Doctorale en Sciences Humaines et Sociales, avec le soutien du CURAPP (CNRS) et de l'Institut Universitaire de France (IUF)). A tradução do original inglês é da minha própria autoria. Cora Diamond é uma filósofa americana cujo trabalho, que se estende da filosofia da lógica e da linguagem até à ética, é inspirado por Frege e Wittgenstein. É um dos nomes centrais do chamado “New Wittgensteinianism” (outro seria por exemplo James Conant, da Universidade de Chicago). O neo-wittgensteinianismo parte de uma interpretação (dita ‘austera’) da noção de *nonsense* na filosofia do autor.

¹⁰⁴ Contrariamente ao que alguém como Martha Nussbaum, por exemplo, assume ser possível (Diamond 1991h: 380), e a faz ter uma concepção menos problemática do que a de Diamond das relações entre filosofia e literatura (cf. Nussbaum 1990, *Love's Knowledge – Essays on Philosophy and Literature*). Seria no entanto injusto atribuir a Nussbaum qualquer coisa como uma proposta de substituição da filosofia por literatura quando se discute moralidade: «I am not proposing, as some recent attackers of traditional philosophy might) that we *replace* the works of Plato, Hume, Kant, Mill and Sidgwick, in our ethics curricula, by the Works of Henry James and Proust. I find that idea ludicrous and highly distasteful. These are great and wonderful works whose contribution to the subject could not possibly be replaced by that of novels. I am only modestly suggesting that we might add the study of certain Great novels to these other studies in order to be sure that we are fair to all the alternative conceptions of goodness.» (Nussbaum 1983: 204)

¹⁰⁵ Diamond 1991d, 1991e, 1991f, 1991g, 1991h, 2006.

respeito aos materiais da filosofia moral, e portanto para o trabalho dos filósofos morais. Para isso, vou considerar antes de mais (i) as críticas de Diamond às concepções ‘limitadas e limitadoras’¹⁰⁶ (e muito espalhadas) do trabalho do filósofo moral, (ii) em seguida as questões que Diamond coloca a si própria acerca da interpretação de Wittgenstein sobre ética que subjazem ao papel que é proposto para a literatura¹⁰⁷, e finalmente (iii) vou sugerir que se é certo que ela tem um caso sólido contra a ‘cegueira à própria cegueira’ [*blindness to blindness*] revelada por alguns filósofos morais, sendo em parte contra esta cegueira que ela sugere o uso da literatura, o facto é que nem todos os aspectos da sua forma de ver o papel da literatura na filosofia moral são igualmente convincentes – é este nomeadamente o caso no que diz respeito ao ‘assombro, maravilhamento e reverência’ [*the awe, the wonder and the reverence*] que ela pensa estarem em falta no pensamento moral. O que eu quero sublinhar é, então, apenas que uma vez que Diamond recruta a literatura para propósitos diferentes (contrariar a ‘*blindness to blindness*’, mas também evocar ‘*the awe, the wonder and the reverence*’) os filósofos que nem sonhariam em contar a literatura como parte do seu trabalho [*wouldn't dream of counting literature as part of their job*] podem estar a rejeitar coisas completamente diferentes quando rejeitam a entrada da literatura na filosofia moral.

I Concepções do trabalho dos filósofos morais

A concepção de Diamond dos materiais admissíveis em filosofia moral¹⁰⁸ necessariamente parecerá a muitos filósofos analíticos demasiado lata: muitos filósofos, certamente muitos filósofos analíticos, nem sonham contar materiais literários como ‘parte do seu trabalho’. É claro que isto presume uma concepção da natureza de tal trabalho, e é quanto a essa concepção que Diamond exprime dúvidas de vários tipos. A minha primeira finalidade aqui é procurar as razões para tais dúvidas nos artigos sobre ética que se encontram reunidos em *The Realistic Spirit*. Creio que é importante ter em mente que esses artigos de Diamond não são apenas artigos críticos: nas suas próprias palavras, eles reflectem o que ela pensa ser ‘a concepção de filosofia de Wittgenstein’¹⁰⁹. Tudo isto importa portanto para se

¹⁰⁶ Diamond 1991h: 376.

¹⁰⁷ Diamond 2000.

¹⁰⁸ Diamond 1992, Diamond 2006.

¹⁰⁹ Diamond 1991 a: 14

conceber o lugar da ética, e a natureza da ética, numa concepção fregeana-wittgensteiniana da natureza da filosofia como a de Diamond.

O filósofo moral cuja concepção de trabalho Diamond quer criticar aparece sob vestes tão diferentes quanto são diferentes entre si pessoas como Peter Singer e Tom Regan (em «*Eating Meat and Eating People*»), Onora O'Neill (em «*Anything but Argument*») e William Frankena (em «*Missing the Adventure*»)¹¹⁰.

Sob as vestes de Singer e Regan, temos o filósofo moral a discutir o que fazer (e não fazer) na nossa relação com os outros animais em termos do direito destes a terem os seus interesses levados em conta. A ideia é que qualquer criatura com interesses – e aqui ter interesses depende da capacidade de sofrimento e prazer – deve ver esses interesses respeitados. Parte do trabalho do filósofo moral aqui é argumentar a favor de imparcialidade: o tom (surpreendentemente) irritado de Diamond quando fala de Singer e Regan não tem nada a ver com a correcção [*rightness*] das intuições destes quanto à ‘forma cruel [*callous*] como nós tratamos os outros animais’¹¹¹: ela descreve-se a si própria como ‘sendo ela própria vegetariana (mas) perturbada por ver quão obtusos são os argumentos [*dos filósofos morais a favor do vegetarianismo*]’. A sua irritação tem tudo a ver com a forma como o trabalho da filosofia moral é visto por pessoas como Singer e Regan, o que se reflecte no ‘tom’ não acidentalmente ‘moralista’¹¹² das suas posições, e que se encontra resumido no próprio uso que Singer faz do termo ‘especisimo’, o qual é para Diamond ele próprio uma marca da forma errada de conceber as nossas relações morais com outros animais, que começa com uma forma errada de conceber as nossas relações morais com outros humanos (voltarei a este ponto).

Sob as vestes de O'Neill temos a acusação feita por um filósofo moral a outro filósofo moral de um tipo específico de falha a fazer filosofia (moral): em *Anything but argument* o filósofo Stephen Clarke é criticado por, no seu livro *The Moral Status of Animals*, ‘não dar o passo para além da asserção e até ao argumento’ [*not reaching beyond assertion to argument*], falhando assim a fazer o trabalho do filósofo moral. O seu caso a favor dos animais apoia-se em última análise numa visão, e num apelo ao coração, e é assim capaz de persuadir apenas ‘aqueles que já estão incinados a aceitar tal visão’. Neste caso a irritação de Diamond diz respeito à suposição de O'Neill de

¹¹⁰ Diamond 1991f, Diamond 1991d, Diamond 1991e.

¹¹¹ Diamond 1991f: 334.

¹¹² Diamond 1991f: p. 323

que os argumentos são a única maneira de ir para além da asserção, a qual depende de ela ter qualquer coisa como uma concepção da filosofia moral ela própria – uma certeza cheia de si acerca da natureza do pensamento moral, a qual torna incompreensível o apelo moral de qualquer coisa outra que não sejam argumentos (regressarei a isto).

Sob as vestes de Frankena temos o filósofo moral a ensinar outras pessoas a fazer filosofia moral. Analisando o *Crítion de Platão*, Frankena decompõe didacticamente os argumentos avançados por Sócrates: ele tem três argumentos para mostrar por que não deve quebrar as leis da cidade; indica as premissas (um princípio e uma declaração de facto) e a conclusão de tais argumentos (por exemplo ‘Devemos obedecer e respeitar os nossos pais e professores’, se Sócrates fugir ele estará a desobedecer aos seu pai e professor, por isso Sócrates não deve fugir). Assim procede pois, o pensamento moral: nele lidamos com argumentos acerca do que é bom, mau, justo, injusto; se concordarmos com as premissas e se formos racionais, segue-se a aplicação do princípio aos factos que estão à nossa frente e uma conclusão surgirá.

Estas formas de ver podem parecer perfeitamente naturais: o que é que pode possivelmente estar errado com concepções do trabalho do filósofo moral que envolvem a discussão de direitos, o uso de argumentos, e conclusões acerca daquilo que qualquer ser racional deve fazer? Segundo Diamond muitas coisas podem estar erradas: na verdade ela está a usar os três casos acima como casos de filósofos morais cegos à sua própria cegueira.

O ideal de não enviesamento ou imparcialidade de Singer e Regan fá-los não prestar atenção à natureza dos próprios conceitos envolvidos na discussão, em particular o que é ser humano, ou ser um humano.¹¹³ Isto manifesta-se a forma mais directa possível num exemplo que Diamond considera particularmente revelador (e que a indigna): a questão colocada por Singer, no contexto da sua discussão com adversários que sublinham a importância da racionalidade e da linguagem para considerar as diferenças entre humanos e animais, acerca das razões por que não usamos humanos a quem faltem tais capacidades da mesma forma que usamos outros animais (nomeadamente em experiências laboratoriais). Se Singer está interessado, ao colocar a sua questão, em forçar a revelação do nosso especismo como um

¹¹³ Acerca desta questão, cf. sobretudo *Experimenting on Animals* (Diamond 1991g) e *Losing Our Concepts* (Diamond 1988).

preconceito não fundamentado, (é essa a razão por que não o fazemos, porque somos especistas) aos olhos de Diamond ele está a revelar a sua cegueira ao tipo de questão que é a questão ‘O que é ser humano?’. Singer simplesmente assume que se trata de uma questão a ser respondida pela ciência natural, da mesma forma que questões idênticas acerca de qualquer outra espécie – a ciência natural dir-nos-á o que são os humanos, os chimpanzés, as amibas. A questão muito gráfica, muito directa, de Diamond que dá o título ao seu artigo de 1975, é: *’Is eating people really like eating animals?’* – será que comer pessoas é mesmo como comer animais? Com ela Diamond pretende chamar a atenção para os limites da analogia entre humanos e animais. Não é como marca de respeito pelos seus direitos e interesses que nós não comemos os outros humanos, da mesma forma que não é uma marca do nosso respeito pelos seus direitos e interesses o facto de nos envolvermos em práticas como enterrar os mortos e não nos envolvermos noutras (como práticas sexuais com pais ou irmãos). Estas são coisas que entram na determinação do tipo de conceito que é o conceito de ser humano; dizem respeito à *invenção do humano*. E é isto que Singer simplesmente não vê.

Quanto a O’Neill, a sua concepção de filosofia moral propriamente dita deixa-nos de mãos vazias quando de trata de fazer algum sentido do facto de as pessoas verem as suas respostas ao mundo completamente alteradas por meios outros que não a argumentação racional (por exemplo por um livro ou um filme) – para ela, tais conversões simplesmente contrastam com aquilo que está à altura de standards apropriados, e respeita as nossas capacidades de humanos pensantes; portanto tais coisas não são the *“business of moral philosophy”*, ficam aquém do trabalho do filósofo moral. Que possa ser um problema que sejamos assim deixados de mãos vazias é um problema que O’Neill não vê.

Para Frankena, a filosofia moral lida com argumentos que dizem respeito à aplicação de regras gerais a estados de coisas factualmente descritos. No entanto se olharmos mais de perto para alguns dos pressupostos factos no exemplo acima (por exemplo ‘fugir seria desobedecer ao nosso pai e professor’) temos que perguntar a nós próprios de que é que estamos a falar. As leis da cidade não são o pai e o professor de Sócrates em nenhum sentido literal: Sócrates está *a ver as coisas de uma forma nova*, e no entanto este trabalho da imaginação moral de Sócrates é algo a que Frankena é cego. Além do que, pensando, como ele pensa, que factos são factos e que na descrição dos factos o trabalho do filósofo moral ainda não começou (tudo aquilo sobre que

podemos discordar são ‘valores’) ele pressupõe uma distinção facto-valor (‘herdada da idade de ouro da meta-ética’) que nos impede de ver o que pode estar envolvido no desacordo moral. Porque o facto é que duas pessoas podem nem sequer ver os mesmos factos (temos mesmo que pensar que ‘fugir é desobedecer ao nosso pai e professor?’) e ‘não ver os mesmo factos’ é um tipo de distância¹¹⁴ totalmente diferente de discordar acerca de valoração de algo sobre cuja natureza se concorda.

Os casos de Singer-Regan, O’Neill e Frankena convergem numa forma de ver o trabalho do filósofo moral que na verdade constituiu o mainstream: uma forma de olhar para o pensamento moral a que podemos chamar forense, centrada na acção e na decisão manifestas, perante possibilidades estabelecidas, e com a qual se lida em termos de argumentos. Sabemos que para Diamond uma tal forma de ver as coisas é presa de um ‘ideal metafísico’ [is ‘*in thrall of an metaphysical ideal*’] que torna os filósofos cegos à sua própria cegueira¹¹⁵. É contra isto que ela acentua o aspecto não forense do pensamento moral [‘*non-forensic aspect of moral thought*’¹¹⁶], a importância, no pensamento moral, de ‘ver coisas’, da sensibilidade e da responsividade, (ser avassalado, em contraste com agir), da atenção em contraste com a desatenção [‘*thought vs. thoughtlessness*’] perante a experiência, do foco em problemas tais como ‘que tipo de bens são os conceitos’ (como ser humano), o papel, no pensamento moral, da imaginação moral, (‘ver o estado como pai e professor’), bem como o esforço envolvido em ver possibilidades que não estão já desenhadas perante nós (o ‘risco’ e a ‘aventura’ na vida moral, que podem perfeitamente passar-nos despercebidos, cf. *Missing the Adventure*¹¹⁷). Finalmente, a importância de disparidades na forma de ver o que as coisas são, uma distância bem mais profunda do que a mera discordância na avaliação (cf. *Disagreement and Distance*¹¹⁸). Sabemos que é aqui que os materiais literários entram em acção, entram no trabalho do filósofo moral: eles marcam a mudança na forma como queremos fazer filosofia moral, e é suposto que consigam coisa tão distintas como a deslocação de conceitos, uma crítica do emotivismo, mesmo uma crítica à confiança que alguém como Martha Nussbaum tem em que teremos uma história consensual acerca do que é filosofia moral. Uma vez tendo entrado os materiais literários, a discussão moral é, como

¹¹⁴ Diamond 2006

¹¹⁵ Diamond 1991: 317.

¹¹⁶ Lovibond 2007: 305

¹¹⁷ Diamond 1991e.

¹¹⁸ Diamond 2006.

Sandra Laugier diz em *Concepts moraux, connaissance morale*¹¹⁹, reconfigurada, recentrada de argumentos para formas de vida e inteligência moral. Isto muda a nossa situação, uma situação ‘em que a permissibilidade do aborto se tornou o paradigma da discussão moral’, como diz Diamond, a qual pressupõe uma visão do mundo como sendo compreensível, um assunto sem mais problemas, que podemos sem mais problemas deixar para a epistemologia e a filosofia da ciência.¹²⁰ Sobre um tal plano de fundo, a ausência de simplicidade [*unplainness*]¹²¹ das obras literárias aparece como totalmente irrelevante (da mesma forma que a personagem Elizabeth Costello, em *The Difficulty of Reality*, não sendo simples, é mal interpretada por tantas pessoas¹²²). O mesmo não acontecerá, no entanto, depois da mudança. Mas como, exactamente, está tal mudança ligada a Wittgenstein?

II *Nonsense* ético e questões para si própria

A concepção de ética de Diamond resulta, como se sabe, da sua muito particular leitura do *Tractatus* – na verdade ela crê que a concepção de ética do Wittgenstein tardio não se alterou no essencial relativamente ao Wittgenstein do tempo do *Tractatus*: mantém-se a forma de separar o dizer ético (*ethical talk*) do dizer comum (*ordinary talk*), a rejeição de um tipo único de avaliação (para cadeiras e morangos como para o bem e o mal), mantém-se o ênfase no ‘visar para além do mundo’ (*reaching beyond the world*). Se a reflexão moral vai continuar, como obviamente continua nos artigos a que me referi mais acima, a inspiração Tractariana não pode possivelmente significar que nada de ético pode ser dito: na verdade ela diz acerca dos artigos sobre ética reunidos em *The Realistic Spirit* que eles são, eles próprios, tentativas de fazer trabalho em ética nos termos do espírito realista, i.e. sem de deixar enfeitiçar por requisitos metafísicos («attempts to think of ethics in a

¹¹⁹ Laugier 2006.

¹²⁰ Diamond 1991h: 376-378

¹²¹ É esta a palavra usada por Diamond.

¹²² Elizabeth Costello é a personagem do romance com o mesmo nome de J.M. Coetzee. Costello é uma escritora já de uma certa idade, que vive dilacerada, ferida, com o seu conhecimento da forma como os animais são tratados nas nossas sociedades. Esse conhecimento, de uma situação que para a maior parte das pessoas nada tem de especial mas que ela compara ao Holocausto (de forma extremamente ofensiva para outros), isola-a, assombra-a. Mas ela não é uma vegetariana ou uma moralista militante, que pretenda persuadir os outros das suas razões através dos melhores argumentos: tudo o que Costello quer é salvar a sua alma; ela sabe perfeitamente quão contraditórios são as suas atitudes (por exemplo usar couro) e os seus sentimentos. É por Costello ser assim que ela é ‘*unplain*’: não é possível substituir Costello por um argumento a favor do vegetarianismo ou de um igualitarismo radical; ela não é um dispositivo que esteja apenas a substituir ideias e argumentos.

realistic spirit, i.e. ‘not in the thrall of metaphysical requirements’¹²³»). Nós sentimo-nos inclinados a dizer coisas éticas (coisas acerca do bem e do mal, da vontade, do mundo como totalidade, da vida) e dizêmo-las. Diamond também as diz. É verdade no entanto que ela sente necessidade de clarificar o estatuto destas coisas que diz e esse é precisamente o propósito, em *Ethics, Imagination and the Tractatus*¹²⁴, das muitas questões para si própria acerca do estatuto das suas articulações da concepção wittgensteiniana da ética. De forma característica, Diamond pergunta: ‘se eu uso uma expressão que é, pura e simplesmente, *nonsense* quando estou supostamente a articular a posição de Wittgenstein – como é o caso por exemplo de uma ‘atitude em relação ao mundo como totalidade – o que é que posso pensar estar a conseguir?’ [‘If I am using a phrase which is simply nonsense in supposedly articulating Wittgenstein’s views – such as ‘attitude toward the world as a whole’ - what can I think I am achieving?’¹²⁵] Devo fazer isso? Posso fazê-lo sem recuar relativamente à visão austera do nonsense? E também, de forma crucial: como pode uma leitura de Wittgenstein acerca de ética retratá-lo como não sendo nem um místico (comprometido com uma visão inefabilista do *nonsense*) nem um positivista (eliminando o nonsense ético através de uma versão ou outra de emotivismo? É bastante claro que Diamond não está a ir pelo mesmo caminho que outras abordagens da ética também inspiradas em Wittgenstein, que tomam a ética como um ramo do pensamento, que se ocupa com as coisas certas ou erradas a fazer, (...) e com aquilo que significa falar acerca de tais assuntos¹²⁶. Mas o que pode então ser a ética?

A resposta wittgensteiniana de Diamond é que a ética é uma atitude perante o mundo como totalidade, ligada ao sentido da vida, que pode penetrar qualquer tipo de discurso ou pensamento, e não um ramo da filosofia com um tópico particular¹²⁷. Mas não será que qualquer tentativa de falar acerca do mundo como totalidade nos condena a pronunciar nonsense? O nonsense é de facto a chave aqui. No entanto, o principal ponto é que este não é o tipo de nonsense de que devemos abster-nos (obviamente neste ponto a questão de Diamond a si própria torna-se: será que isso significa que há tipos de nonsense? E sabemos à partida que ela não quer afirmar tal coisa). O seu primeiro passo em direcção à sua própria visão wittgensteiniana de ética

¹²³ Diamond 1991b: 23

¹²⁴ Diamond 2000.

¹²⁵ Diamond 2000: 155.

¹²⁶ Diamond 1991: 10

¹²⁷ Diamond 1991 a: 9

é então a interpretação do nonsense ético, no contexto da visão austera do nonsense. É claro que isto traz imediatamente uma nova questão para si própria: ‘Se eu aceito a visão austera do nonsense no *Tractatus* – “não pode haver pensamentos nonsense expressos por frases nonsense”, não pode haver diferença entre o nonsense de ‘Bláblabláblá’ e o nonsense de ‘Deus existe’ – como é que posso pretender manter o nonsense ético?’ A razão de ser do artigo *Ethics, Imagination and the Tractatus*¹²⁸ é em parte dar uma resposta muito elaborada a esta questão, uma resposta que envolve toda uma interpretação do *Tratatus* e de Wittgenstein. Isto é feito defendendo antes de mais uma diferença entre nonsense ético e nonsense metafísico. O nonsense metafísico é para eliminar – é isso que faz o *Tractatus*, é assim que, por assim dizer, o *Tractatus* se dirige aos filósofos: é como se se estivesse a falar da perspectiva do metafísico, para fazer com que o que ele diz venha a ser percebido como nonsense. Não apenas o metafísico fala sem sentido, mas a própria maneira de dizer que ele está a fazer tal coisa – i.e. a totalidade do *Tractatus* –, vem também a ser nonsense. Por isso temos o nonsense número um (o do metafísico) e o nonsense número dois (o do *Tractatus*), e depois há o nonsense número três: o nonsense ético, que não é para ser eliminado, antes permanece atraente. Há muita coisa a explicar aqui, mas a ideia principal é que podemos compreender alguém que ‘fala nonsense’ e que o *Tractatus* nos dá o enquadramento e as ferramentas para isso. A chave é (*Tractatus* 6. 54): «Aquele que compreende que estas proposições são nonsense compreende-me a MIM», no sentido em que faz sentido a partir do nonsense que eu pronuncio. No entanto, se não vamos aceitar uma distinção entre bom e mau nonsense, como é que vamos perspectivar, conceber, alguma distinção entre nonsense a ser atirado fora e nonsense a manter em uso? A proposta de Diamond é que uma actividade imaginativa está aqui envolvida, tal como está envolvida em entendermos o *Tractatus*, e podemos usar circunstâncias externas (o contexto) para distinguir o carácter da actividade imaginativa envolvida e aquilo que está a ser conseguido ao compreender o nonsense. Compreender nonsense não é colocar-se a partir de fora do pensamento do enunciador, descrevendo o que se passa de um ponto de vista empírico, mas também não é estar ‘dentro’, como quando se compreende alguém que faz sentido usando frases que nós próprios pronunciaríamos. O que está em jogo não é exprimir o inexprimível ou exprimir emoções mas fazer (novo) sentido: em contraste com a

¹²⁸ Diamond 2000.

ilusão, que é o que está em jogo no nonsense metafísico e que é para ser eliminada, o auto-entendimento é aqui a finalidade.

E eis-nos de volta aos materiais literários: é enquanto explicita tais conexões que Diamond nos diz como vê e como não vê a literatura. Ela não vê a literatura como uma ilustração de princípios (racionais, gerais, éticos) ou como oferecendo-nos personagens e situações como evidência para questões éticas (é precisamente isto que ela diz em *Missing the Adventure*, ao caracterizar a oposição entre Raphael e Nussbaum quanto ao uso da literatura na filosofia moral). Ela não está a propor que se aborde os textos filosóficos como literatura ou os textos literários como textos de filosofia moral; o que ela está a fazer é apontar a literatura como sendo ela própria uma forma de levar a cabo a reflexão moral dada a visão da natureza da ética que obtemos do *Tractatus*: «Assim como se pode fazer matemática com demonstrações mas também desenhando alguma coisa e dizendo ‘Olha para isto’, da mesma forma o pensamento ético decorre em argumentos mas também em alguma coisa que não é argumento, mas sim histórias e imagens [‘Just as mathematics can be done by proof but also (...) by drawing something and saying, “Look at this”, so ethical thought goes on in argument and also not in argument, but (e.g.) stories or images’]¹²⁹]. Mas será este o fim da história?

III Literatura: ‘cegueira à cegueira’ [*blindness to blindness*], e ‘espanto, maravilhamento e reverência’ [*the awe, wonder and reverence*’]

Como ficamos então quando uma perspectiva wittgensteiniana nos faz mudar completamente de ideias quanto àquilo que queremos fazer em ética? Ficamos com abordagens como a de Diamond, nos seus artigos sobre ética de *The Realistic Spirit*, e em geral, deixamos entrar materiais literários no campo da ética. No caso específico de Diamond, não é de todo indiferente que os seus exemplos de materiais literários sejam Henry James, Charles Dickens, W. Wordsworth, N. Hawthorne, L. Tolstoi, i.e. autores que procuram profundidade e detalhes na alma humana, em contraste com obras que podem facilmente ser acusadas de serem superficiais e de não nos dotarem de nenhuma visão (tal como, por exemplo, na recente discussão passada nas letras britânicas, quando o Prof. G. Josipovici, respeitado crítico literário, se referiu da seguinte forma ao trabalho de autores contemporâneos célebres como Julian Barnes,

¹²⁹ Diamond 1991 a: 9.

Salman Rushdie e Ian McEwan: eles ‘não nos oferecem nenhuma visão’, são superficiais, e considerar que retratam o espírito dos tempos não é justificação suficiente). Claro que poderíamos ainda perguntar: mas porquê a literatura? Será que apenas a literatura pode fazer tal trabalho? Em particular e dado o exemplo que acabei de referir, será que qualquer tipo de literatura pode fazer tal trabalho? Mas a verdade é que é suficientemente claro de que forma é suposto que estes materiais funcionem como antídoto à ‘cegueira à cegueira’ da filosofia moral *mainstream*: como diz Diamond em *Losing One’s Concepts*¹³⁰ eles ‘desenvolvem a nossa sensibilidade e formas de vida conceptual’; ou (voltando às críticas que ela fez nos anos 70 às discussões em torno dos direitos dos animais e ao que ela pensa serem os indefensáveis pressupostos de tais discussões); ou: «Os escritos de Tolstoi não têm peso sobre o conceito da espécie biológica *Homo sapiens*; os escritos dos escritores podem iluminar, elaborar e aprofundar o nosso conceito de ser humano, assim como o papel do mistério e do risco nas histórias da nossa vida moral »¹³¹. Mas a verdade é que determinadas questões permanecem em aberto acerca da forma como Diamond vê o papel da literatura: o facto é que, além de apoio para o seu diagnóstico de cegueira à cegueira, Diamond recruta a literatura para sublinhar o espanto, o maravilhamento e a reverência [*the awe, the wonder and the reverence*¹³²] que considera estarem em falta no pensamento moral. É para isso que ela quer manter a linguagem que se encontra no *Tractatus*, e também nos *Notebooks* e na *Lecture on Ethics*¹³³, onde o ético é marcado ou delimitado como envolvendo valor absoluto (*absolute value*), como procurando visar para além do mundo (*reaching beyond the world*). Como diz Diamond, «Há um princípio que guia todo o trabalho de Wittgenstein: aquilo sobre que se está a falar é dado na forma como se fala disso (...) se o assunto é o que é escuro e sinistro no coração humano podemos marcar a forma de falar sobre isso através da característica lógica de separar tal linguagem completamente da linguagem comum sobre o que se passa, não a deixar entrar aí. Essa característica lógica pode ser observada, argumentei eu, nos contos de fadas, e de uma forma diferente, na recusa de Kant e de Wittgenstein, de uma psicologia da vontade má. E não há razão nenhuma para pensarmos que a filosofia tardia de Wittgenstein nos impede de

¹³⁰ Diamond 1988.

¹³¹ Diamond 1991 h: 371.

¹³² Diamond 1991d: 306.

¹³³ Wittgenstein 1961, Wittgenstein 1993.

reconhecemos um tal contraste» («There is a guiding principle in Wittgenstein's work, all of it: what you are talking about is given in how you talk about it (...) If the dark and sinister in the human heart is the subject we may mark our talk about it through the logical feature of cutting such talk off from ordinary talk about what goes on, not giving it entry there. That logical feature may be seen, I argued, in the fairy tales, and, in different ways, in Kant's and Wittgenstein's refusal of an empirical psychology of the evil will. And there is no reason to think that Wittgenstein's later philosophical thinking precludes us from recognizing that kind of contrast»).¹³⁴

O interesse de Diamond por aquilo que é “escuro e sinistro nos corações humanos”, pelos “fenómenos morais extremos que desafiam a imaginação” e que não são articuláveis, não é apenas uma crítica da cegueira à cegueira: o que Diamond está a fazer é ligar o espanto, que Wittgenstein de facto quer na ética, com a escuridão nos corações humanos, a infabilidade do mal, a dificuldade da realidade. No entanto se consideramos o próprio Wittgenstein, nas passagens, que na verdade não são assim tantas, em que ele introduz a linguagem que Diamond pretende manter e usar, temos de perguntar a nós próprios: mas será todo o espanto moral? [*is all awe moral?*] Ou formulando a questão na linguagem do *Tractatus*: será todo o visar para além do mundo, toda a atitude perante o mundo como totalidade, ética? O que é que essa linguagem delimita exactamente, para Wittgenstein? Se lermos essas passagens como uma concepção sobre a fonte do valor (o valor não está no mundo, todos os factos são factos, não há nada de bom ou mau acerca deles), o ênfase na receptividade, em contraste com a acção e decisão, a ênfase em alguma coisa que nos avassala [*something that comes over us*], e que não é alguma coisa que fazemos, na sua conexão com o valor (absoluto), tudo isso está certamente lá. Mas é também precisamente neste contexto que Wittgenstein diz de forma célebre que a ‘ética e a estética são Um’, e em conjunto com o desvio da atenção da acção para a resposta (sentir-se avassalado, marcado pelo espanto: ‘*the awe*’) não encontramos nada que se pareça com o ‘escuro e sinistro no coração humano’, ‘a impossibilidade de articulação (*unspeakability*) ou a dificuldade da realidade de Diamond. O que encontramos é antes felicidade e aceitação no olhar do artista: «Esteticamente o milagre é que o mundo exista. Que exista aquilo que existe. A essência da forma artística de olhar para as coisas é olhar para o mundo com um olhar feliz (*Notebooks*, 20.10.1916) (...)»

¹³⁴ Diamond 2000: 170.

A obra de arte é o objecto visto *sub specie aeternitatis* e a vida boa é o mundo visto *sub specie aeternitatis* (ibidem, 7.10. 1916)». Como Sabina Lovibond comenta, «A resposta estético-contemplativa ao mundo como totalidade tem sido louvada (...) por respeitar a nossa ligação a esses momentos de consciência que nos revelam um valor de que não somos autores (...), em contraste com outros em que o valor é realizado através do bom exercício da nossa vontade em face a alguma resistência natural»¹³⁵. Penso que vale a pena considerar isto se pensamos que a literatura, admitindo o seu papel na ética, tem ainda uma pretensão a ser pensada como arte e não apenas como educação moral. Para os propósitos presentes tudo o que quero é sugerir que a ênfase de Diamond no escuro e sinistro no coração humano e na dificuldade da realidade como inarticulabilidade que a isto se liga pode perfeitamente ser separada, deve ser separada, das suas críticas específicas à cegueira à cegueira, aquelas com que comecei. E porque Diamond recruta a literatura para ambos os propósitos, aqueles filósofos que ‘nem sonhariam considerar a literatura como parte do seu trabalho’ podem simplesmente estar a rejeitar coisas diferentes.

Referências

- Cavell, S., Diamond, C., McDowell, J., Hacking, I. & Wolfe, C., 2008, *Philosophy and Animal Life*, New York, Columbia University Press.
- Conant, James, 2000, «Frege and early Wittgenstein», in A. Creary and R. Read eds, *The New Wittgenstein*, London, Routledge.
- Crary, Alice (ed), 2007, *Wittgenstein and the Moral Life – Essays in Honor of Cora Diamond*, Cambridge MA, MIT Press.
- Diamond, Cora, 1988 «Loosing our concepts», *Ethics* 98, pp. 255-277.
- Diamond, Cora, 1991, *The Realistic Spirit – Wittgenstein, Philosophy and the Mind*, Cambridge MA, MIT Press.
- Diamond, Cora, 1991a, «Introduction I: Philosophy and the Mind», in Diamond, Cora, 1991.
- Diamond, Cora, 1991b, « Introduction II: Wittgenstein and Metaphysics», in Diamond, Cora, 1991.
- Diamond, Cora, 1991c, «Throwing Away the Ladder: How to read the *Tractatus*», in Diamond, Cora, 1991.
- Diamond, Cora, 1991d, «Anything but Argument?», in Diamond, Cora, 1991.
- Diamond, Cora, 1991e, «Missing the Adventure – Reply to Martha Nussbaum», in Diamond, Cora, 1991.
- Diamond, Cora, 1991f, «Eating Meat and Eating People», in Diamond, Cora, 1991.
- Diamond, Cora, 1991g, «Experimenting on Animals: A Problem in Ethics», in Diamond, Cora, 1991.

¹³⁵ Lovibond 2007: 319.

- Diamond, Cora, 1991h, «Having a Rough Story about What Moral Philosophy is», in Diamond, Cora, 1991.
- Diamond, Cora, 2000, «Ethics, imagination and the Tractatus», in A. Crary and R. Read eds, *The New Wittgenstein*, London, Routledge.
- Diamond 2006, «Differences et distances morales: quelques questions», in *Étique, littérature et vie humaine*, Sandra Laugier ed., Paris, PUF.
- Diamond 2008, «The difficulty if reality», in Cavell, S., Diamond, C., McDowell, J., Hacking, I. & Wolfe, C., 2008, *Philosophy and Animal Life*, New York, Columbia University Press.
- Laugier 2006, «Concepts moraux, connaissance morale», in S. Laugier ed, *Étique, littérature et vie humaine*, Paris, PUF.
- Lovibond 2007, «In Spite of the misery of the world: Ethics, Contemplation and the Source of Value»
- Nussbaum, Martha, 1983, “Flawed Crystals: James’s *The Golden Bowl* and Literature as Moral Philosophy”, *New Literary History* 15, 25-50.
- Nussbaum, Martha, 1990, *Love’s Knowledge – Essays on Philosophy and Literature*, Oxford, Oxford University Press.
- Wittgenstein, Ludwig, 1993, Lecture on Ethics, in J. Klagge and A. Nordmann eds, *Philosophical Occasions 1912-1951*, Indianapolis-Indiana, Hackett.

Hilary Putnam e a ética – respostas a Susana Cadilha¹³⁶

Sofia Miguens e Susana Cadilha

“O que é um facto? Um facto é uma proposição que é verdadeira”

Este texto teve origem na necessidade de responder a um conjunto de perguntas colocadas por Susana Cadilha após a apresentação das propostas da filósofa americana Cora Diamond acerca de filosofia moral e literatura correspondentes ao artigo anterior deste volume (S. Miguens, *What Some Philosophers Wouldn't Dream of Counting as Part of Their Job – Cora Diamond, literatura e filosofia moral*). As questões da Susana excediam a filosofia de Diamond e o uso que esta faz de Wittgenstein, e em particular do *Tractatus*, na sua abordagem. Fiquei por isso convencida de que as respostas pedidas só poderiam ser dadas uma vez feitos os esclarecimentos mais genéricos acerca da natureza da ética e da história do pensamento ético que aqui constam. Estes esclarecimentos assumem por outro lado uma necessidade própria dentro do projecto “*Conversations on Practical Rationality*”, uma vez que envolvem o problema da separabilidade das questões ‘puras’, ou ‘teóricas’, da racionalidade prática relativamente às questões substanciais. Agradeço muito à Susana por me ter dado o pretexto de esclarecer estes pontos, inclusive perante mim própria

Susana Cadilha - As questões éticas são questões de semântica ou de metafísica?

Sofia Miguens - A resposta que gostaria de dar é que as questões éticas não são nem questões de semântica nem questões de metafísica: são questões éticas. Enquanto questões éticas são obviamente questões filosóficas, sendo que qualquer questão filosófica nos conduz a esclarecimentos semânticos e metafísicos, sem que tenhamos que dizer que ‘as questões éticas são em última análise questões semânticas porque são acerca da forma como falamos’ (acerca de assuntos como o bem e o mal) ou são em última análise questões metafísicas porque são acerca do que há, acerca da existência ou não existência de um tipo especial de factos e propriedades que seriam “morais” – factos morais, propriedades. Mas não quero dar tal resposta de uma forma

¹³⁶ Desde a criação do nosso grupo de investigação Hilary Putnam tem sido para nós uma figura inspiradora de referência, nomeadamente porque a ambição do MLAG é ter uma agenda de investigação englobante em filosofia contemporânea, que vá desde a filosofia da lógica à ética e Putnam é um dos raros filósofos analíticos contemporâneos cuja obra tem tal amplitude. O trabalho aqui presente resulta da transição de ambas as autoras do projecto *Conversations on Practical Rationality* (2007-2011) para o Projecto *The Bounds of Judgement* (PTDC/FIL-FIL/109882/ 2009) (2011-2013) (Task: Judging Morally).

abrupta – penso que precisamos aqui de um excuro através da história do pensamento ético para que ela faça sentido. A ideia básica é que talvez não estejamos condenados a ver a ética como um domínio especial ou à parte em comparação com outros domínios do nosso pensamento, um domínio em que a nossa forma de pensar seria totalmente singular. Nesta concepção haveria uma linha a separar o domínio ético – os juízos éticos e as questões quanto à natureza destes – dos restantes domínios do nosso pensamento. E a verdade é que percursos determinados na história do pensamento tornaram para muitas pessoas natural pensar precisamente assim, assumindo que em ética se trata de valores, que estes seriam totalmente subjectivos, e que proposições éticas não capturariam a forma como o mundo é mas apenas exprimiriam emoções, ou sentimentos, ou imperativos. Mais especificamente, o emotivismo¹³⁷ e o prescritivismo¹³⁸, i.e. não-cognitivismos de tipos vários na história do pensamento ético na filosofia analítica, situam-se aqui: a ideia é que não há uma realidade moral ‘independente’. Isto não significa, obviamente, que nós não façamos juízos morais: o que acontece é que, se estes exprimem desejos, preferências ou emoções, ou outros estados e atitudes subjectivos, então não podemos dizer que eles são verdadeiros (ou falsos) quando e por que captam correcta (ou incorrectamente) a realidade moral. Repito o que disse antes: um pressuposto desta forma de ver é que o que fazemos quando julgamos moralmente é algo de completamente diferente do que fazemos quando fazemos juízos de outra natureza, por exemplo juízos perceptivos ou juízos lógicos. Ora, é precisamente isto, desde logo, que o filósofo que me guiará neste excuro, Hilary Putnam, não aceita. Putnam não aceita tal forma de ver as coisas porque considera que ela está dependente de uma dicotomia factó-valor que por sua vez considera injustificada: para Putnam a dicotomia factó-valor é um dogma, não apenas tanto quanto se aplica aos nossos pensamentos e juízos éticos como aos nossos pensamentos em geral. O que me leva àquilo que eu gostaria de ter dito imediatamente perante a pergunta da Susana: que a formulação da pergunta ela própria, tal como a Susana a fez, pressupõe de alguma forma, subtilmente, em *background*, a aceitação dessa dicotomia factó-valor, para termos uma noção ‘pura’ de *factó* com que lidar quando falamos daquilo que torna proposições éticas

¹³⁷ Cf. Stevenson (Ethics and Language, 1944) ou A. J. Ayer (Language, Truth and Logic, 1936). A partir daqui para um mapeamento da ética, seus autores e questões, tomo como referencia Galvão, Ética (A filosofia por disciplinas)

¹³⁸ Cf. Hare (The Language of morals 1952, Freedom and Reson, 1963)

verdadeiras. Conseguimos formular a alternativa que a Susana formula se já pressupusermos de alguma forma a dicotomia. Talvez isso se torne mais claro se virmos que esta dicotomia está na verdade por trás da separação (canónica na filosofia analítica) da ética em partes (meta-ética, ética normativa, ética aplicada), a qual reserva as questões teóricas da racionalidade prática (as ‘questões semânticas e metafísicas’ da pergunta da Susana) para a meta-ética – nesta trata-se dos ‘problemas metafísicos, epistemológicos e semânticos suscitados pela ética’. E a ideia é que a meta-ética pode ser separada da ética ‘substancial’, onde se discutem ‘questões substantivas sobre aquilo que se deve fazer e aquilo que é bom ou valioso’¹³⁹.

Consideremos então a dicotomia facta-valor: sustentá-la é afirmar que uma coisa são factos, outra completamente diferente são valores. Há questões acerca de factos e questões acerca de valores, e estas são questões totalmente diferentes entre si. Mas será realmente esse o caso? Para muitas pessoas esta distinção é perfeitamente pacífica mas para Putnam trata-se de uma dicotomia tão dogmática como a célebre distinção analítico-sintético, e de resto relacionada com ela¹⁴⁰. Para compreendermos esta posição de Putnam temos que recuar até questões gerais relativas à nossa concepção de mente, conhecimento, pensamentos, juízos, na relação destes com o mundo, já que é esse o quadro no qual se colocam tais questões e tais distinções fazem (ou não fazem) sentido. De qualquer forma, e isto pode ser tomado para já como uma mera observação acerca de ‘sociologia académica’, muitos filósofos trabalhando hoje no quadro da filosofia analítica, e configurando aquilo que fazem, de um ponto de vista metodológico¹⁴¹, como uma consequência da crítica à distinção analítico-sintético, não parecem incomodar-se de todo com a dicotomia facta-valor. Não é esse o caso de Putnam: para ele as discussões estão interligadas, e portanto essa é uma situação paradoxal.

Começemos com a distinção analítico-sintético: esta pode e deve ser criticada e como se sabe Quine é aqui a referência incontornável. Mas parece que nos afastamos muito da ética com esta discussão – a discussão acerca de analítico e sintético não é uma discussão epistemológica e de filosofia da linguagem? Desde logo uma observação – não é nada claro que possamos legitimamente instaurar tais separações de disciplinas

¹³⁹ Cf. Pedro Galvão, “Ética”.

¹⁴⁰ Putnam 1992, *The Collapse of the Fact-Value Dichotomy and Other Essays*. Este é o principal tópico do livro.

¹⁴¹ Nomeadamente reclamando-se de um *naturalismo* que é para Quine uma consequência do colapso da distinção analítico-sintético.

filosóficas, a ética para um lado, a epistemologia, a filosofia da linguagem, a metafísica para outro, como se fossem assuntos completamente estanques. Há o risco de tais separações pressuporem elas próprias decisões específicas acerca da natureza dos nossos pensamentos e juízos na relação destes com o mundo. Também aqui, - particularmente aqui - para compreendermos a natureza e a história das relações mútuas entre domínios da filosofia ao longo da história da filosofia, e em função de visões determinadas da mente e do pensamento, Hilary Putnam pode dar-nos uma valiosa ajuda.

Mas porquê apelar especificamente a Putnam, se ele é mais conhecido como filósofo da ciência (da física, da lógica), da mente e da linguagem? Será que Putnam tem alguma coisa de particularmente importante a dizer sobre ética? Bom, uma das razões para evocar Putnam é o facto de ele se interessar pela história da filosofia analítica no século xx e pelos dogmatismos que, na sua opinião, nesta nasceram. Por outro lado, Putnam está particularmente interessado nas repercussões das questões em torno da dicotomia facto-valor na forma como se trabalha noutras disciplinas, em particular na economia, que Putnam considera, como o seu colega de Harvard Amartya Sen, ter uma relação natural estreita com a ética. Ora tais repercussões seriam, por exemplo que muitas pessoas pensem poder trabalhar em economia como se esta não tivesse nada a ver com a ética, e como se essa forma de ver as coisas fosse perfeitamente justificada. A questão é: justificada como? Putnam pensa - e considera essa uma das razões fundamentais para ter escrito *The Collapse of the Fact-Value Dichotomy*, que dedica a um outro amigo e colega economista de Harvard, Vivian Walsh – que a expulsão das questões éticas da economia pelos economistas foi algo que aconteceu numa constelação intelectual específica que importa compreender do ponto de vista filosófico, algo que não era o caso nas origens da economia, com Adam Smith¹⁴². Foi possível chegar a pensar assim, de uma forma que permite considerar menor, ou de alguma forma ilegítima, ou teoricamente não fundamentada, qualquer tentativa de *welfare economics*¹⁴³ - esta é uma outra repercussão que preocupa Putnam. Porque

¹⁴² Como diz Putnam, “a economia tem-se orgulhado frequentemente de evitar suposições metafísicas ao mesmo tempo que engole inteira a metafísica do positivismo lógico” [“economics has frequently prided itself on ‘avoiding metaphysical assumptions’ while positively gobbling up logical positivist metaphysics”].

¹⁴³ O exemplo de Putnam em *The Collapse of the Fact-Value Distinction*, é Lionel Robins (economista britânico, da London School of Economics, 1898-1984). Uma figura inspiradora central dos seus escritos neste domínio é de resto o seu colega economista Amartya Sen (economista indiano, n. 1933, prémio Nobel da Economia em 1998, professor em Harvard durante a maior parte da sua carreira)

numa área como a economia se trata de políticas públicas, de aconselhar governos e organizações não governamentais acerca do que fazer, a convicção de que de um lado estão ‘a razão e os factos’ e do outro ‘juízos de valor subjectivos’ tem consequências práticas notórias – como diz Putnam, estão em questão questões de vida e morte. Será tal separação fundamentada? Uma subtil separação entre racionalidade prática e ética é bastante para argumentar a favor da legitimidade da separação entre ‘a razão e os factos’ e os ‘juízos de valor subjectivos’. Mas como é que ela funciona? E será que em última análise se sustenta?

Não é vão que seja Putnam, um *insider* quando se trata de pensar na forma como se pensa ao fazer física ou lógica, a defender que não podemos (não devemos) separar o nosso pensamento sobre a natureza dos nossos pensamentos em lógica ou em física, do nosso pensamento sobre os nossos pensamentos éticos. E o que Putnam quer defender antes de mais pensando acerca da natureza dos nossos pensamentos é que as duas dicotomias, a dicotomia analítico-sintético e a dicotomia facto-valor, nos impedem de ver a forma como *no nosso pensamento* (em geral, sem discriminação de domínios, quer esteja em causa a percepção ou a ética) e na relação destes com o mundo *descrição e avaliação estão entrelaçadas e são interdependentes*. Quando se afirma alguma coisa acerca do mundo e se pergunta se essa pretensão é verdadeira, há considerações envolvidas que não são distintas em ética e em descrições-avaliações de outro tipo – desde logo, quando quer que digamos que *p*, que as coisas no mundo são tal que *p*, coloca-se a questão acerca *do que conta como p*. Noutras palavras, é preciso saber de que falamos quando falamos de ‘factos’, e esta não é uma questão restrita à ética, mas sim uma questão que encontramos em qualquer área da filosofia. Ora, para Putnam, como dizia J. L. Austin em *Sense and Sensibilia*, “a verdade e a falsidade são [*também elas*] nomes para uma dimensão de avaliação – como as palavras se sustentam quanto a [*satisfactoriness*] com os factos, eventos, situações a que se referem” (Austin 1962: 148¹⁴⁴). Noutras palavras, não temos aí de um lado o domínio dos puros factos acerca do quais proposições podem simplesmente ser verdadeiras ou falsas e do outro ‘uma dimensão de avaliação’ de que a ética faria parte. A boa forma de ver as coisas é ver que mesmo na possibilidade de falar de verdadeiro e de falso e de factos há uma dimensão de avaliação. E, por outro lado, na

¹⁴⁴ Cf. Travis 2010, *Truth and Merit*: 252.

ética enquanto dimensão de avaliação não há por que excluir verdade e falsidade de juízos na relação destes com factos.

Olhemos então, através dos olhos de Putnam, para a ‘paisagem das disciplinas filosóficas’ e para a forma como a filosofia é praticada. A tese de Putnam é que embora a história da dicotomia analítico-sintético e respectiva queda, sob a crítica de Quine, seja bem mais familiar àqueles que se interessam por epistemologia, metafísica e semântica, do que a história e o destino da dicotomia facto-valor, esta última colapsa de forma análoga à primeira. O facto de ambas as dicotomias colapsarem é para Putnam uma intuição central do pragmatismo. Em *The Collapse of the Fact-Value Dichotomy* Putnam procura por isso contar as duas histórias em paralelo, a mais conhecida e a menos conhecida, indo para além da filosofia do século vinte, até à filosofia moderna, e em particular ao empirismo clássico, ou em geral à discussão em torno de ‘analítico’ e ‘sintético’ que se passa em torno dos empiristas, racionalistas e Kant – isto é, como se sabe, o que Quine faz no início de *Dois Dogmas do Empirismo* – a discussão, reconhecidamente, vem de longe.

Referi o pragmatismo de Putnam – é de um ponto de vista pragmatista que Putnam rejeita as duas dicotomias – no entanto é bem sabido que Quine, na sua célebre crítica à dicotomia analítico-sintético, em *Dois Dogmas do Empirismo*, se reclama também ele do pragmatismo. Acontece que Putnam e Quine são dois pragmatistas profundamente diferentes entre si. Desde logo o interesse que um (Putnam) tem por ética, está totalmente ausente no outro (Quine), e tal situação é um sintoma das diferenças na forma de ser pragmatista. Seguindo aqui Ch. Travis, queria sugerir que para compreender Putnam é preciso compreender que: (i) Quine é crucial na forma como Putnam pensa na filosofia do século xx, (ii) Putnam não é Quine¹⁴⁵. Além do interesse ou não interesse pela ética, uma outra marca ilustrativa da diferença entre o pragmatismo de Putnam e o pragmatismo de Quine são as suas posições totalmente diferentes acerca de naturalização e naturalismo. Recorde-se que a naturalização e o naturalismo que saem da crítica quineana à distinção analítico-sintético, que são concebidas como uma consequência desta, foram extremamente importantes na filosofia analítica nas últimas décadas, e também em ética, e em particular na

¹⁴⁵ Cf. Travis 2010, especialmente os três artigos sobre Putnam, *The Shape of the Conceptual, On Morally Alien Thought e What Laws of Logic Say*.

‘legitimação’, digamos, de programas naturalistas de investigação em áreas várias¹⁴⁶. Ora Putnam pensa que um certo tipo naturalismo hoje muito generalizado só pode assentar sobre uma defesa dogmática da dicotomia facto-valor.

Partamos então de algo a que Putnam chama um “fragmento de filosofia quase generalizadamente aceite como senso comum”¹⁴⁷: a ideia de que juízos de valor como ‘Este homem é cruel’, proferido por mim acerca de alguém que acabou por exemplo de executar uma pena capital sobre uma mulher adúltera, são subjectivos. Ao contrário de asserções de facto eles não podem ser nem verdadeiros ou falsos, nem objectivamente justificados [*objectively warranted*]. Juízos de valor estariam assim de alguma forma fora da esfera da razão; quando se trata de valores não existe espaço para argumentação. Em *The Collapse of the Fact-Value Dichotomy* Putnam procura compreender as condições de possibilidade para se chegar a pensar assim na filosofia dos últimos séculos. Uma consequência da suposição de que juízos de valor são subjectivos é que *valores ou fins não podem ser racionalmente discutidos* - ora esta é sem dúvida uma ideia que se encontra por exemplo em David Hume, um filósofo empirista clássico, usualmente identificado como fundador da corrente sentimentalista contemporânea¹⁴⁸. Vale por isso a pena recordarmos em que contexto e com que pressupostos formulou Hume tal ideia, para podermos avaliá-la melhor – é isso que Putnam propõe em *The Collapse of the Fact-Value Dichotomy*. Isto ajudar-nos-á a ver mais claro, de um ponto de vista histórico o que estamos a deixar entrar quando aceitamos – se é que aceitamos – que juízos de valores são subjectivos e que valores ou fins não podem ser racionalmente discutidos.

Mas antes de avançar quero recapitular algumas das principais ideias que introduzi até agora e que me parecem fundamentais para contextualizar de um ponto de vista da história da filosofia e da história da ética a resposta à pergunta que a Susana fez. Sugeri que uma dicotomia facto-valor estava subtilmente subjacente à formulação da pergunta. Introduzi Putnam para sugerir que não devemos admitir sem mais discussão essa dicotomia facto-valor. Mas talvez uma coisa não tenha sido ainda clara, e espero que o venha a ser agora, falando de David Hume: é que, mesmo que isso não venha explicitamente ao caso, uma opção tão básica em filosofia como ser ou não ser empirista *não é de todo indiferente quando discutimos estas questões* (vamos chamar-

¹⁴⁶ No nosso projecto, o exemplo mais directo disso é a filosofia experimental.

¹⁴⁷ Putnam 2002: Introduction 1.

¹⁴⁸ Cf. Miguens 2004, Racionalidade, Filosofia moral: D. Hume versus I. Kant (p. 106)

lhe questões acerca da natureza dos juízos éticos e questões acerca da natureza da ética como domínio filosófico). Apenas com este pano de fundo completo montado se compreenderá a resposta que quero dar à Susana; que as questões éticas não são nem questões de semântica nem questões de metafísica: são questões filosóficas, e as questões filosóficas éticas não estão separadas por um abismo ontológico das outras questões filosóficas, sendo que questões filosóficas envolvem todas elas discussões semânticas e metafísicas. Ou pelo menos só assumiremos que as questões filosóficas éticas estão separadas por um abismo ontológico das outras questões filosóficas se concedermos à partida determinados pressupostos epistemológicos e de teoria da mente que frequentemente na história da filosofia nos vêm do empirismo. Sublinho de novo que a minha resposta à pergunta da Susana só pode ser compreendida se considerarmos a possibilidade daquilo a que Hilary Putnam chama o ‘colapso da distinção factio-valor’ como estritamente paralelo ao colapso da distinção analítico-sintético e se considerarmos que esse ‘colapso’ tem consequências não apenas na forma como vemos a natureza da ética mas consequências que se estendem por todas as áreas do nosso pensamento, desde a metafísica e epistemologia até às relações da ética com a economia. A moral é: não podemos, não devemos, separar as áreas do nosso ‘pensamento sobre a natureza do pensamento’ umas das outras sem mais problemas. Falar de filosofia por disciplinas é extremamente útil do ponto de vista metodológico, ou de iniciação, ninguém contesta tal coisa, mas não pode fechar os nossos olhos às relações entre os objectos destas – afinal em filosofia estamos sempre a falar da relação pensamento-mundo, estejamos a trabalhar em lógica filosófica ou em ética. Vamos então olhar mais um pouco para a história da ética através dos olhos de Putnam.

Nos dois livros que escreveu sobre ética e sobre o lugar e a importância da ética na filosofia (já referi *The Collapse of the Fact-Value Dichotomy* (2002), o outro é *Ethics Without Ontology* (2004)), Putnam pretende mostrar as raízes históricas (no empirismo e no Positivismo Lógico) daquilo que considera ser uma forma ‘segregacionista’ de abordar as questões éticas na filosofia analítica, começando por classificar domínios, separando meta-ética, ética normativa e ética aplicada (ou colocando estas duas últimas em conjunto, como sendo substanciais, por contraste com o carácter meta-teórico da meta-ética¹⁴⁹). A discussão em meta-ética pode então

¹⁴⁹ Cf. Galvão, *Ética*.

ser direccionada para a semântica e a metafísica, ficando as questões substanciais de lado, reservadas para a ética normativa e para a ética aplicada. O seu ponto é que olhando para as raízes encontraremos opções que podem ser criticadas e que nos farão ver as questões de forma diferente – até porque por razões independentes podemos perfeitamente pura e simplesmente não ser empiristas, e em particular não subscrever o tipo de empirismo dos positivistas lógicos. No caso de Putnam, e mesmo se a sua carreira como filósofo, em particular no seu início, está de alguma forma ligada ao positivismo lógico (Putnam foi discípulo do filósofo alemão Hans Reichenbach, e foi nesse enquadramento que trabalhou em filosofia da ciência, nomeadamente filosofia da física)¹⁵⁰, é em parte o seu passado pragmatista¹⁵¹ que o faz considerar o entrelaçamento factó-valor como natural, uma ideia alheia à recuperação positivista-lógica do empirismo, tal como esta marcou a ética, nomeadamente dando origem aos chamados não-cognitivismos, de vários tipos, emotivismos ou prescritivismo. Estes têm em comum a ideia de que frases éticas não exprimem proposições e não podem ser verdadeiras ou falsas. E recorde-se que apesar da sua profissão de fé pragmatista, a consideração da questão factó-valor quando pensamos na forma como pensamos em ‘factos’ está também ausente em alguém como Quine.

A ideia de Putnam é então que a distinção entre descrições, que seriam factuais, e avaliações, que não seriam factuais, vai em conjunto com distinção analítico e o sintético. Vejamos com que forma aparece esta última remonta em Hume. De que falamos quando falamos de verdades ‘analíticas’ e ‘sintéticas’? Em linguagem contemporânea, ser “analítico” seria ser definicional, verdades analíticas sendo ditas serem verdadeiras em função do significado, por contraste com as verdades ditas sintéticas que enunciariam ‘factos’. Mas já antes, quer num metafísico racionalista como G. Leibniz quer num empirista como Hume, encontramos esta forma de classificar verdades: ela permite-nos situar de um lado por exemplo as verdades matemáticas (‘verdades de razão’ para Leibniz, e como tal verdadeiras em todos os mundos possíveis, *relations of ideas* para Hume) e do outro verdades ‘físicas e morais’ (para usar a linguagem de Leibniz), tais como ‘Júlio César atravessou o Rubicão’ ou ‘Esta parede é vermelha’ (verdades contingentes para Leibniz, *matters of*

¹⁵⁰ A ligação diz respeito ao objecto da prática da filosofia (a ciência, nomeadamente a física), já que Reichenbach não fez parte do Círculo de Viena.

¹⁵¹ E portanto americano, e não europeu, ou de origem europeia (como é o caso do Positivismo Lógico). Cf. o interesse de Putnam por John Dewey.

fact para Hume). É esta distinção que é criticada por Quine, de forma célebre, em *Two Dogmas of Empiricism*, onde alude à história da discussão. Pelas mãos de Quine, as consequências da crítica alargam-se a uma reformulação das relações entre filosofia e ciência, contra a que tinha sido a proposta dos positivistas lógicos (a ‘expulsão’ da metafísica, e a identificação da filosofia com uma teoria da forma – lógica, linguística – daquilo que a ciência produz). Na verdade parte do que está aqui em causa – na ética e não apenas na ética – é a relação entre a ciência e a filosofia, e o que cabe à filosofia ser e fazer – e essa é uma questão que os naturalismos hoje tão difundidos, pelo menos na filosofia analítica, não podem evitar.

Para pesarmos melhor a forma como a *dicotomia relations of ideas – matters of fact* se situa no empirismo de Hume, consideremos a sua relação com a chamada “Lei de Hume”¹⁵². Esta é uma tese generalizadamente atribuída a David Hume segundo a qual que não se pode derivar um ‘deve-se’ [*ought*] de um ‘é’ [*is*], i.e. não se pode derivar um dever-ser da forma como as coisas são. Por exemplo no *Tratado da Natureza Humana* (Livro III, Parte I, secção I) Hume afirma que em todos os sistemas de moralidade que até então encontrou, os autores começam raciocinando de forma perfeitamente usual, descrevendo a sociedade humana ou procurando provar a existência de Deus, e de repente, sem qualquer aviso ou justificação, passam de um ‘ser ou não ser’ para um dever-ser (por exemplo de ‘Deus é o nosso criador’ para ‘Deve-se obedecer a Deus’). É esse passo que de acordo com Hume não pode ser justificado

Em *The Collapse of the Fact-Value Dichotomy*, Putnam começa por notar que a tese de Hume acima não é um simples ponto lógico, que devamos conceder sem problemas¹⁵³. Pelo contrário, ela pressupõe uma metafísica substancial: o que Hume está a dizer não é comparável com afirmar que não se pode inferir ‘p&q’ de ‘p ou q’. Ao formular a sua ‘Lei’ Hume está, antes de mais, a assumir a distinção epistemológica entre *matters of fact* e *relations of ideas* (a sua particular antecipação da dicotomia analítico-sintético) e a dizer que quando um juízo descreve um *matter of fact*, nenhum ‘*ought*’ pode ser dele derivado. Além de chamar a atenção para o facto

¹⁵² Cf. *Tratado da Natureza Humana*, Livro III, Parte I, secção I. Hume afirma aí que em todos os sistemas de moralidade que até então encontrou, os autores começam raciocinando de forma perfeitamente usual, descrevendo a sociedade humana ou procurando provar a existência de Deus, e de repente, sem qualquer aviso ou justificação, passam de um ‘ser ou não ser’ para um dever-ser (por exemplo de ‘Deus é o nosso criador’ para ‘Deve-se obedecer a Deus’). Para Hume este passo não pode ser justificado.

¹⁵³ Putnam 2002:14.

de subjacente à célebre ‘Lei de Hume’ encontrarmos um metafísica de *matters of fact*, Putnam faz notar que tal metafísica de *matters of fact* pressupõe por sua vez uma particular teoria da mente, a que Putnam chama uma ‘*picture semantics*’: a ideia é que conceitos são ideias, ideias são imagens, i.e. representam questões de facto assemelhando-se-lhes. Estas mesmas ideias também podem estar associadas a sentimentos ou emoções – o que lhes conferirá outras propriedades que não semânticas. Por isso quando Hume diz que não se pode derivar um ‘ought’ de um ‘is’ ele está também a dizer que não há *matters of fact*, questões de facto, quanto a ‘certo’ (*right*) ou ‘virtude’ (*virtue*) – ou essa propriedade seria representável pictoricamente tal como na descrição acima se caracteriza. Juízos de valor não são portanto juízos de facto – são as paixões ou sentimentos que nos fazem pensar que tais *matters of fact* existem. A ideia substancial para a história do pensamento ético é que os componentes dos nossos juízos acerca de tópicos éticos como virtude e vício ou outros são sentimentos. Estes sentimentos são despertados pela contemplação das acções correspondentes, dado o particular tipo de mentes que nós temos.

A primeira coisa a notar aqui é que em termos de teoria da mente, de semântica, de epistemologia, ninguém se lembraria hoje de subscrever tal e qual o que Hume aceita. Há coisas inaceitáveis por todos os lados, simplismos que podem ser discutidos por si, e sem qualquer referência à ética. Temos de qualquer forma uma classe de juízos com um traço distintivo: juízos de valor, aqueles que envolvem ser bom, ser mau, virtude, vício, certo, errado, dever– ora a pergunta que Putnam faz é a seguinte: chamar-lhes juízos de valor dir-nos-á, por si, o que é um juízo ético? A sua resposta é negativa, como é negativa a resposta à pergunta acerca de se chamar ‘analíticos’ a uma certa classe de juízos, por exemplo lógicos, nos diz, por si só, o que são juízos lógicos. Chamar a determinados juízos ‘juízos de valor’ ou ‘juízos analíticos’ não nos diz o que os membros da classe têm em comum ou o que é não pertencer à classe. E é isso que é preciso compreender.

Note-se que Putnam não quer negar que existam esses juízos de que Hume fala, mesmo se as suas fronteiras são vagas: como no caso de ‘analítico’, Putnam pensa que a eliminação seria prematura: da mesma forma que no caso da distinção analítico-sintético a posição de Putnam é mais nuanceada do que a de Quine, o crítico empirista radical da analiticidade (a crítica quineana à distinção analítico-sintético termina com a *eliminação* do analítico, e Putnam não dá esse passo e por isso a sua posição em filosofia da lógica, nomadamente, é diferente da de Quine), a sua posição perante a

dicotomia facto-valor é mais nuanceada. O que há de errado em Hume para Putnam é a metafísica de *matters of fact*: sem ela, a distinção que nos permite encarar juízos de valor como não sendo juízos de facto, sendo as paixões ou sentimentos que nos fazem pensar que tais *matters of fact* existem – cai.

Olhemos agora para os juízos ditos de valor de uma outra perspectiva, absolutamente central na história da ética – a perspectiva de Kant. Também Kant é um actor importante na discussão geral acerca da natureza dos juízos, também para ele a discussão envolve a sub-questão da natureza dos juízos éticos. O papel de Kant na história, se esta for contada em termos de factos e valores, é bastante complexo – Putnam sugere que pensemos em Kant como defendendo que juízos éticos têm a forma de imperativos (Kant fala de máximas, regras, imperativos categóricos). Podemos ver isto da seguinte forma: por exemplo dizer ‘Assassinar é errado’ é dizer ‘Não assassines’. Isto não é uma descrição de factos – nesse ponto Kant concordaria com Hume. Juízos éticos não são descrições de factos. Onde Kant discorda decididamente de Hume é no passo seguinte: se um juízo ético não é uma descrição de factos, pensa Hume, então é uma expressão de sentimentos. Para Kant não se trata de nada disso. Para Kant é a razão que tem aqui todo o papel. Estamos assim perante o núcleo do pensamento kantiano sobre a natureza (racional) da moralidade e também perante aquilo que é o mais difícil de justificar e manter no kantismo ético: o papel e a natureza da ‘razão pura prática’.

Desde logo, temos aqui um problema estrutural semelhante àquele que encontramos em Hume – ninguém, por mais kantiano que seja e há muitos e importantes kantianos na filosofia moral contemporânea (por exemplo, e já que atrás referi Amartya Sen, uma outra colega de Putnam em Harvard, Christine Koorsgaard), estaria hoje disposto a subscrever tal e qual a metafísica e a teoria da mente que sustentam a posição de Kant acerca da natureza dos nossos juízos éticos. E sem isso como conseguiremos compreender a ideia kantiana, fundamental de ‘razão pura prática’? Não parece nada simples – mas sem ela nada se sustenta da mesma maneira.

O ponto de Putnam nesta história comparativa de disciplinas filosóficas (epistemologia/ética¹⁵⁴) é que: «assim como o colapso da credibilidade filosófica da noção kantiana de verdade sintética a priori levou os positivistas lógicos a recuar para

¹⁵⁴ Obviamente esta classificação disciplinar não faz muito sentido para alguém como Putnam – mas ele próprio a adopta para propósitos de exposição, e ela parece-me, nesse contexto, bastante esclarecedora.

uma versão vastamente inflacionada da ideia de Hume segundo a qual um juízo ou é analítico (lida com *relation of ideas*) ou é sintético (lida com *matters of fact*), e também os levou a expandir o analítico (uma vez que a tentativa para mostrar que a matemática é sintética a priori não funcionou), assim o colapso da credibilidade filosófica da noção kantiana de ‘razão pura prática’ (e com ela da variedade kantiana de ética a priori fundada sobre tal noção) conduziu os positivistas lógicos a recuarem até uma versão vastamente inflacionada da ideia de Hume segundo a qual os juízos éticos não são asserções de factos mas ou expressões de sentimentos ou imperativos sob disfarce»¹⁵⁵. Putnam nota que para os positivistas estes imperativos não podem, ao contrário, do que acontece com Kant, ser ‘racionais’ – eles reflectem aquilo que seria em última análise um estado volicional.

Só falta mais um elemento para chegarmos à constelação que subjaz ao texto sobre a interpretação por Cora Diamond de algumas ideias de Wittgenstein sobre ética, o texto de onde saem as perguntas da Susana, nomeadamente a questão sobre se questões éticas são questões semânticas ou questões metafísicas. Falta-nos compreender o aproveitamento que os positivistas lógicos fazem da noção wittgensteiniana, tractariana, de *nonsense* e que vai ser o outro componente importante para pensar sobre ética. Repare-se que aí o contexto é outro relativamente a Hume ou a Kant. Não é apenas um contexto em que se lida com juízos analíticos e sintéticos, factos e sentimentos: é um contexto em que a forma de abordar as questões discutidas sob esses títulos mudou por razões metodológicas. A ideia de análise lógica está a mudar a forma como a natureza dos ‘juízos’ é concebida – aliás, tornou-se mais comum falar de ‘pensamentos’ e ‘proposições’. A lógica formal contemporânea e o seu uso na análise de pensamentos obviamente não fazia parte do horizonte de empiristas e racionalistas clássicos ou de Kant. É de qualquer forma no novo contexto que Rudolf Carnap dirá, em *The Unity of Science* “Todas as asserções pertencendo à metafísica, à ética reguladora, (...) têm esse defeito de serem inverificáveis (...) e nós estamos habituados no Círculo Vienense a descrevê-las como *nonsense* (seguindo Wittgenstein)”. E em *The Logical Syntax of Language*¹⁵⁶: “Uma vez purificada a filosofia de todos os elementos não científicos [frases metafísicas, por exemplo, ou frases morais, todos elas visando, sem sentido, referir entidades transcendentais],

¹⁵⁵ Putnam 2002: 17. Como frequentemente, Putnam remete-nos para o seu professor Hans Reichenbach, no caso *The Rise of Scientific Philosophy*, capítulo 17, *The nature of ethics*.

¹⁵⁶ London, Kegan Paul, 1937

apenas a lógica da ciência permanece“. E para a filosofia ficam assuntos de lógica, linguagem, e significação, inclusive acerca da linguagem das asserções éticas, daquilo que de facto dizemos e continuamos a dizer de ético. Assim sendo, a concepção de ética dos positivistas lógicos, para além de uma concepção de analítico e sintético, incorporou a noção wittgensteiniana, tractatária, de *nonsense* – e esse é o ponto de partida no artigo *What Some Philosophers Wouldn't Dream of counting as part of Their Job*. Dizemos que asserções são *nonsense* se elas não são nem analíticas nem empiricamente verificáveis ou falsificáveis – é esse precisamente o caso das asserções éticas. Como havemos então de pensar na natureza deste *nonsense*? A ideia dos positivistas lógicos será que no caso do nonsense das asserções éticas se trata de algo como expressão de emoções, ou prescrição. Essas são análises que decorrerão na meta-ética. Fazer ética será então fazer análise de linguagem e nestas circunstâncias os problemas éticos seriam realmente problemas semânticos, pegando na alternativa da Susana. Ou pelo menos a tarefa filosófica de fazer ética seria em grande parte ocupada por esclarecimentos semânticos deste tipo. Como diz David Wiggins, nas mãos dos neo-positivistas, “tudo o que até então passara por filosofia moral deverá ser reconfigurado [*revamped*] como meta-ética e apresentado como algo de moralmente neutro”¹⁵⁷.

A contra-proposta de Putnam (de resto diferente da proposta de Diamond que explorei no artigo que a Susana comentou) é a seguinte: uma vez desinflacionada a dicotomia factó-valor permanece ainda uma distinção útil entre juízos éticos e outros juízos (exactamente como após a crítica à distinção analítico-sintético queremos manter algo da analiticidade nomeadamente para pensar na natureza de juízos lógicos). Também há uma distinção entre juízos da química e os que não são desse domínio – mas o que é importante é que nada de metafísico se segue daí (ao contrario do que Wittgenstein do *Tractatus* lido pelos Positivistas Lógicos pode levar a pensar). E o mesmo deverá passar-se quanto a juízos éticos.

Em contraste com Quine, Putnam está especificamente interessado em conceber (correctamente, justificadamente, no seio de uma visão filosófica global) a natureza daquilo a que vamos chamar ‘factos’, e é precisamente nesse contexto que ele defende (em geral e não apenas para a ética e os juízos éticos) que *na relação pensamento-mundo a avaliação e a descrição são entrelaçadas e interdependentes*. Perceber como

¹⁵⁷ Wiggins 2006: 318.

chega ele a esta tese obriga-nos a ter em conta toda a filosofia de Putnam, e particular a procurar compreender como, segundo Putnam, de um ponto de vista pragmatista, uma distinção de que podemos precisar não é, não tem que ser, uma dicotomia. A marca do pragmatismo em filosofia é atacar dualismos; mas o pragmatismo não nos obriga a abdicar de distinções: a ideia é que os problemas surgem quando distinções são tratadas como dogmas. A dicotomia analítico-sintético é o exemplo central da diferença de postura de Quine e de Putnam quanto a distinções; na verdade aquilo que acabei de dizer reflecte-se por exemplo na forma como Putnam critica a filosofia quineana da lógica, toda ela construída sobre a rejeição da analiticidade. Ao contrario do que acontece com Quine, a crítica de Putnam à distinção analítico-sintético não redundava em ‘deitar fora’ a analiticidade.

Em suma, não temos que pensar nos juízos éticos como diferindo abissalmente em natureza de outros juízos quaisquer – os problemas éticos são semânticos, metafísicos, epistemológicos, o que quisermos, da mesma forma que quaisquer outros problemas filosóficos, da mesma forma que os problemas relativos ao conhecimento perceptual ou ao conhecimento lógico. Não tem que haver dicotomia alguma, como diz David Wiggins, entre «*getting it right in matters of ethics and getting it right in matters of fact*» (captar as coisas de forma certa em questões de ética e captá-las de forma certa em questões de facto)¹⁵⁸.

Susana Cadilha – Mas o que dizem afinal Cora Diamond e Wittgenstein acerca da dicotomia facta-valor?

Dizem coisas diferentes e arriscamo-nos a ter aqui outra história longa... Mas para perceber qualquer um deles temos mesmo de os ver a partir do contexto que eu estava a descrever (chamemos a esse contexto ‘a forma-como-se-chega-à-concepção-de-ética-desenvolvida-no-positivismo-lógico’); temos que pensar na relação de ambos, Wittgenstein e Diamond, com as propostas acerca do que se faz quando se faz ética, avançadas pelos positivistas lógicos. Ninguém escolhe quem virá a considerar-nos como precursores, e Wittgenstein não escolheu ser seguido da forma que foi pelos positivistas lógicos. O que é indiscutível é que quer Diamond quer os positivistas lógicos desenvolvem as suas posições sobre a ética a partir da forma como

¹⁵⁸ Wiggins 2006: 330.

Wittgenstein no *Tractatus* vê a dicotomia facto-valor. Provavelmente quer Diamond quer os Positivistas Lógicos ‘violentam’ Wittgenstein, mas isso não tem que nos importar demasiado, pelo menos por agora. O que temos de compreender é que o *Tractatus* é a própria origem do critério de significação a que me referi acima: se uma asserção é cognitivamente significativa então tem que ser ou analítica ou sintética. Nada do que possamos dizer de cognitivamente significativo poderá deixar de ser ou analítico ou sintético. Como nada que possamos dizer de ético ou de estético cai sob essas categorias, que subsomem para o Wittgenstein do *Tractatus* respectivamente as proposições lógicas, e ‘as proposições das ciências naturais’, pensamentos empíricos acerca de como o mundo é, então falar de assuntos éticos ou estéticos é dizer coisas sem sentido, pronunciar *nonsense*. Note-se que há uma nova discussão que começa aqui, e que é importante para a interpretação de Wittgenstein, e central no *New-wittgensteinianism* que referi atrás: a discussão que começa aqui é uma discussão acerca da natureza do nonsense, e em torno de saber se há ou não vários tipos de nonsense (temos os critérios do sentido, e temos aquilo que não os cumpre, e que é nonsense; mas por exemplo dizer ‘Blahblahblah’ seria diferente de dizer ‘Deus existe’, embora em ambos os casos se trate, de acordo com o critério em causa, de nonsense). Se em dizer coisas éticas não se trata de fazer sentido, de uma forma que possa tornar-se verdadeira em função quer do próprio significado (proposições analíticas) quer da verificabilidade ou falsificabilidade (proposições sintéticas), então de que se trata? Um opção, como já se percebeu, é dizer que se trata de exprimir algo. Expressão, emoção, prescrição – são as escolhas aqui para os positivistas lógicos. Mas parece claro que para o próprio Wittgenstein as coisas não são assim tão simples. Colocar a alternativa acerca de ser cognitivamente significativo nos termos em que os positivistas lógicos o fazem, implica assumir que o nonsense (por exemplo ético) não é importante. Ora para Wittgenstein o nonsense, nomeadamente o nonsense ético, é importante – é o isto que Diamond explora. De qualquer forma temos a ligação da ética com o nonsense como transgressão dos limites do sentido – quebrar a linha de fronteira do que pode ser dito de forma significativa na linguagem, avançar para além dos limites da linguagem cuja função central (representativa) é explicada – ou posta – pela teoria do *Bild* (modelo) do *Tractatus* tal como esta se conjuga com uma doutrina do analítico e do sintético. Para o Wittgenstein do *Tractatus* a dicotomia facto-valor é formulada dizendo que nada no interior do mundo tem valor. No interior do mundo só há factos. E não há factos éticos. No interior do mundo, no mundo dos factos, nada é

bom ou mau. Tudo é apenas como é. Então quando dizemos o que quer que seja que nos ser pareça ético estamos a transgredir os limites do pensamento e da linguagem com sentido, o domínio das proposições que podem ser verdadeiras ou falsas acerca de factos. Diamond na sua interpretação da ética retém isto – nada no interior do mundo tem valor, só o mundo como totalidade traz a questão de valor – e a ligação que ela propõe da filosofia moral com a literatura apoia-se aí. Recapitulando a forma como Diamond recupera o Wittgenstein do *Tractatus*, o mesmo Wittgenstein que foi incorporado pelos Positivistas Lógicos para proporem a distinção facta-valor como ponto de partida de uma forma totalmente diferente da de Diamond de fazer ética (ou de considerar o que fazemos em ética) – já que este era o assunto do texto que a Susana discute, *What Some Philosophers Wouldn't Dream of Counting as Part of Their Job*. Diamond discorda dos positivistas lógicos – nunca poderia ir pelo lado expressivista. Não é uma empirista, o sentimentalismo de género humeano também não é um caminho. Faz então uma interpretação *sui generis* do que Wittgenstein afirma no *Tractatus*, levado a sério a importância da ética, algo que podemos dizer, de certa forma injustamente, que os positivistas lógicos não fazem. Aceita-se o nonsense ético, não se dá uma explicação não cognitiva (o que é ético não é cognitivamente significativo: o que é cognitivamente significativo só pode ser ou analítico ou sintético), distingue-se o nonsense ético do nonsense metafísico (do realismo metafísico) que Wittgenstein quer criticar, e do nonsense linguístico (o problema da visão austera do nonsense que interessa Diamond enquanto filósofa da lógica e da linguagem), e o que fica é a literatura.

Susana Cadilha – Mas, nestas condições, de que falamos quando falamos de realismo? Uma posição wittgensteiniana que rejeita o realismo metafísico, não nos deixa ao nível das relações humanas, das práticas e das formas de vida (apenas)? Isto não é em última análise verificacionismo?

Bom, quando fala de uma posição wittgensteiniana que rejeita o realismo metafísico e que nos deixa ao nível das relações humanas, das práticas e das formas de vida, já está a falar de outro Wittgenstein, o Wittgenstein das *Investigações Filosóficas*, não o Wittgenstein do *Tractatus*, da dicotomia facta-valor, tal como esta passa para os positivistas lógicos, a dicotomia que marca, pelo menos segundo Putnam, a maneira como a ética ainda hoje se faz, na filosofia analítica. E o Wittgenstein das

Investigações é um crítico do Wittgenstein do *Tractatus*, a quem, como é sabido, acusa de platonismo e dogmatismo. E uma forma de vermos na linguagem que estávamos a utilizar as críticas do Wittgenstein das *Investigações* ao Wittgenstein do *Tractatus* é dizer que ele está a atacar precisamente as dicotomias de que falávamos atrás: a dicotomia analítico-sintético e a dicotomia factio-valor. Como ficamos depois dessa crítica quanto a poder pensar e dizer coisas verdadeiras e falsas (porque podemos perfeitamente pensar e dizer coisas verdadeiras e falsas, segundo o Wittgenstein das *Investigações*) é algo a ser explicado tendo em consideração a queda dos dois dogmatismos. O problema do realismo e do tipo de realismo coloca-se aqui, e é um problema que excede a ética – mas inclui a discussão em torno de realismo em ética, evidentemente. É verdade que a discussão acerca de realismo em ética entra na ética analítica usualmente via a meta-ética, no entanto sem percebermos as ‘colocações históricas’ que estou a procurar descrever creio que a discussão em ética será conduzida descontextualizadamente.

De qualquer forma, mostrar a excessiva crueza da dicotomia realismo-anti-realismo em metafísica não apenas quando se trata de pensar na relação pensamento-mundo no domínio ético mas quando se trata de pensar nessa mesma relação em qualquer domínio que seja (a lógica, a percepção....) é o ponto de nuclear do pensamento de pessoas como Putnam, ou Wittgenstein, ou McDowell, todos eles vendo-se como realistas e sendo considerados por muita gente a partir de forma como não o sendo. Estamos aqui portanto num ponto sensível e nuclear: podemos perguntar a todos eles (Wittgenstein, Putnam, McDowell): és ou não és um realista? O próprio Putnam é, como se sabe, alguém que tem passado a sua vida filosófica a reformular o seu realismo; por isso mesmo, talvez, ele prefere hoje dizer ‘pragmatismo’ – mas quando fala de percepção por exemplo, professa um ‘realismo natural’, ou ‘realismo artitotélico sem a metafísica aristotélica’, contra quer o anti-realismo quer os contra-realismos da ‘concepção absoluta de mundo’.

De qualquer forma e voltando à ética, ser realista como Wittgenstein II, McDowell ou Putnam é pensar que é possível defender a objectividade dos juízos éticos e ao mesmo tempo recusar-se a postular um reino platónico de entidades e propriedades éticas – nos seus escritos sobre ética Putnam utiliza a expressão “objectividade sem objectos”, McDowell fala de “cognitivismo” (dito de uma forma ainda demasiado pouco afinada: os juízos morais têm uma função cognitiva, os juízos morais exprimem

crenças), Wittgenstein ele próprio não escreve assim tanto sobre ética – o que acontece é que hoje há pessoas que desenvolvem posições inspiradas por ele, por exemplo McDowell. Um realismo de tipo putnamiano consagra o que podemos chamar anti-platonismo e a especificidade das mentes, dos pensadores que nós somos – ‘nós vemos isto’ – não devemos dizer que há isto ou aquilo aí existindo independentemente do particular tipo de pensador que nós somos. É assim que Wittgenstein, Putnam, etc se opõem a um realismo platonista – de certa forma esta postura em filosofia começou com Kant. O realismo de Putnam (ou de Wittgenstein) é um realismo que não admite, para usar uma expressão de que a Susana gosta de usar nos seus estudos de ética, um ‘ponto de vista sem ponto de vista’. Trata-se de autores que levam a sério que não estamos em posição de pensar na relação pensamento-mundo a partir do não-pensamento e que pensam que um realismo platonista incorre nesse erro; que há pensadores é algo que tem que estar lá na forma como caracterizamos o realismo e como caracterizamos a objectividade em domínios vários, inclusive o domínio dos juízos éticos – mas não estamos só a falar de ética, estamos a falar de tudo o que pode ser pensado pelos humanos.

Susana Cadilha – Por outro lado, para além de uma posição assim [uma posição wittgensteiniana] nos deixar ao nível das relações humanas e das formas de vida, teremos que pensar que a moral não tem uma base natural?

Parto do princípio que a Susana vê o nível das relações humanas e das formas de vida como sendo superficial e insuficiente para a fundamentação da moral, por contraste, por exemplo, com um reino platónico de factos, factos que tornariam verdadeiras as proposições éticas. Mas essa não é a única forma possível de ver as coisas. O que esbocei acima pretende ser o princípio de uma alternativa. Por outro lado, para falarmos de base natural é fundamental sabermos o que queremos dizer com ‘natural’ e ‘natureza’. Para usar um termo de McDowell desde logo uma base natural não tem que ser concebida de acordo com o “naturalismo cru” (*bald naturalism*) tão espalhado e dominante hoje, em versões que permitem a muitas pessoas, filósofos inclusive, acreditar que questões filosóficas, como as que estamos a discutir, terão respostas científicas. Deixe-me formular o princípio da alternativa que se encontra em Wittgenstein e passa para alguém como McDowell: a gramática (sermos criaturas linguísticas, animais linguísticos) é a história natural dos humanos. Não é menos

natural para nós humanos do que a vida animal para outros animais, mas não é vida meramente animal – é racional. Mesmo se não temos, como diria McDowell, senão as práticas ‘para nos manter nos rails’, ainda assim existem ‘rails’. Não é um reino platónico mas é tudo o que temos, enquanto humanos. E é suficientemente comum para a partir daí ser possível defender um realismo. ‘Natureza’ e ‘natural’ são, como sabe, termos que ‘sofreram’ às mãos de perspectivas reducionistas e cientistas, que nem todos os filósofos partilham: pense-se por exemplo no caso de McDowell contra o naturalismo cru – o ponto é que ‘natural’ não é para ser identificado com o que as ciências naturais nos dizem que há – o que faríamos então das nossas capacidades de pensar, falar, ser morais? Diríamos que são ‘sobrenaturais’? Não, diz McDowell, elas não são nem sobrenaturais nem redutíveis ao que o naturalismo cru pretende; são totalmente naturais e para nós são uma segunda natureza – não chegamos a ver que tipo de seres somos se não as incluirmos na nossa natureza. Mas claro que isto nos obriga a repensar os nossos conceitos de natural e natureza.

Deixe-me tentar formular isto de uma outra maneira, ainda voltando à ética. Uma pessoa pode pensar que os juízos morais são objectivos, mas que um exemplo como “É errado matar pessoas inocentes” é um exemplo demasiado geral para se colocar essa questão da objectividade – por exemplo McDowell pensaria assim. Mas se tenho uma pessoa inocente à minha frente, talvez haja resposta determinada do mundo ao juízo ‘Matar esta pessoa inocente é errado’. Para McDowell essa resposta só estaria disponível com uma sensibilidade apropriada (este é o ponto aristotélico que ele considera essencial para pensarmos na natureza dos juízos éticos) – não há o ver que é assim “do ponto de vista do universo?”, mas há objectividade desse juízo para esse agente com sensibilidade apropriada sendo o mundo como é. Noutras palavras, isso - ‘Matar esta pessoa inocente é errado’ - está aí para ser visto, não sou eu que projecto, ou interpreto, ou crio, ou o que seja: está do lado do mundo. Não se diz portanto que a moralidade é totalmente subjectiva (individual, variável), mas não se pode evocar um ‘ponto de vista sem ponto de vista’ a partir do qual as coisas são assim *tout court*. A propósito, um ponto da teoria das virtudes como alternativa à deontologia e ao consequencialismo explicando o que se deve fazer é precisamente este contextualismo. Podemos formular o contextualismo da seguinte forma: falhar em obter resposta do mundo mostra que alguma coisa não é uma questão objectiva apenas se nós estamos a fazer as questões certas. Se se encontrar as questões certas, então a natureza, o mundo, tem resposta – se não, não há respostas (mesmo para um exemplo como ‘a

relva é verde’ é preciso saber o que conta como tal – aquela relva no quadro pontilhista de Seurat feita de pigmentos de tinta azuis e amarelos e que eu vejo como verde é verde?) Falemos agora em termos de juízos morais e do que os torna verdadeiros ou falsos. A esperança do cognitivismo moral pode ser formulada como a esperança de que haja uma dicotomia entre juízos factuais e juízos éticos. Os juízos éticos estão perante o mundo a serem tornados verdadeiros e falsos pelas formas como as coisas são. Mas o ponto geral acerca de como os juízos estão perante o mundo a serem tornados verdadeiros e falsos pelas formas como as coisas são é que é preciso formular as questões certas - e isto em moralidade é particularmente difícil. Mas admitir isso não nos obriga a pensar é abissalmente diferente de qualquer outra postura perante o mundo (nem sequer ‘A relva é verde’). Quer para a moral quer para outros domínios, a ‘moral’ aqui é que não existem respostas para ‘perguntas em abstracto’ – este é o ponto geral de sensibilidade à ocasião como abordagem da natureza do verdadeiro para pensar nas relações pensamentos-mundo (a propósito, esta é uma leitura de Wittgenstein II – sensibilidade à ocasião do verdadeiro não é de todo relativismo das formas de vida, que consistiria antes em pensar: nós fazemos assim, o povo ao lado faz de outra maneira, e não há mais nada a dizer). Outro ponto além de “não há ponto de vista sem ponto de vista” e da “sensibilidade à ocasião” é o que o Putnam chama o colapso da distinção facto-valor – no facto de o mundo dizer se algo é ou não é verdadeiro está envolvido saber o que conta como verdadeiro numa dada ocasião e as considerações para isso não são essencialmente distintas no caso moral e noutros casos.

Vou apenas recapitular o que diz McDowell acerca de ética, já que fui para lá conduzida, ao procurar falar de Putnam. A ideia é que requisitos morais existem, dão-se-nos. A questão que se coloca é a saber a partir de que perspectiva são os requisitos morais perceptíveis. A ideia wittgensteiniana mas também aristotélica de McDowell é que eles são perceptíveis a partir de dentro de formas de vida determinadas. Para um sistema cognitivo sem visão de cor no nosso mundo não existem cores, no sentido em que não está presente a perspectiva da qual elas seriam perceptíveis. No entanto se pusermos ao lado desse sistema cognitivo, no mesmo mundo, um animal com visão de cor, este vê cores que estão realmente lá – não queremos dizer que as cores são ‘meramente subjectivas’. O mesmo se passa com as nossas respostas morais – elas são respostas de uma sensibilidade ou ponto de vista, mas são respostas a algo que está lá no mundo, não são meras expressões subjectivas. Proposições éticas são

verdadeiras proposições, levantam a questão da sua verdade ou falsidade. Este é o pano de fundo para McDowell ser realista e cognitivista em ética. Outro ponto do cognitivismo de McDowell é que, na sua particular formulação de uma perspectiva wittgensteiniana, ele já está a supôr uma fusão do que é cognitivo e ‘desejante’, conativo, nas nossas repostas ao mundo (*‘besires’* é um termo expressivo que já foi avançado para estes estados cognitivo-afetivos de uma só natureza) – obviamente isto também tem que ser explicitado e elaborado.

Susana Cadilha – Como é que isto nos deixa quanto ao papel da razão na moral? Ficamos condenados ou reduzidos à ideia de que a base do comportamento moral não é o raciocínio? Estará alguém como Diamond do lado da tradição sentimentalista em filosofia moral?

Sofia Miguens – Não, de todo, alguém como Diamond não poderia estar do lado da tradição sentimentalista, da mesma forma alguém como McDowell não poderia estar do lado da tradição sentimentalistas, dado o não-cognitivismo desta (veja-se o célebre artigo de McDowell *Non-cognitivism and Rule Following*), e também pelos laços que já procurei explicar do sentimentalismo com o empirismo: quer Diamond quer McDowell são profundamente anti-empiristas. É precisamente para a formulação do anti-empirismo que Frege e Wittgenstein lhes servem.

Repare que tudo na sua pergunta (Como é que isto nos deixa quanto ao papel da razão na moral?) depende de como concebemos a racionalidade. E da racionalidade ou razão não temos apenas concepções ‘puras’ como a kantiana – aliás, muito provavelmente, sem a imagem de mundo e de mente de Kant, aquela que hoje ninguém sustentaria, não poderemos sequer ter uma concepção pura de racionalidade – a melhor aposta para a racionalidade é concebê-la como a segunda natureza deste animal que também somos, para pôr as coisa nos termos de McDowell, e portanto ‘natural’, e portanto limitada, com todas as limitações características que os estudos empíricos da racionalidade hoje nos revelam, sem as capacidades que Kant lhe atribuiu.

Mas queria dizer também acerca da pergunta da Susana que a questão em Diamond, quando Diamond critica a forma como usamos ‘argumentos em ética’ tem muito mais a ver com o que pensamos que fazemos quando fazemos ética ou filosofia moral do que com a natureza dos juízos éticos, a coisa ela própria, que era aquilo de que

estávamos a falar até aqui. O ponto de Diamond aí é metodológico, e é um ponto crítico relativamente a uma certa forma de fazer ética na filosofia analítica: ela pensa que a fetichização dos argumentos e da argumentação como forma de trabalhar em ética nos leva a deixar escapar por entre os dedos a natureza do ‘material moral’, e este material moral é o pensamento moral das pessoas, o primeiro grau de tudo isto, digamos: nós pensamos moralmente, fazemos juízos morais, e isso é uma forma de pensamento sobre o mundo como outras formas de pensamento sobre o mundo. E que forma é? Como pode existir? Não é função de racionalidade pura, já que racionalidade pura, como Kant a quis, é algo que pessoas como Diamond ou McDowell, não vêm como uma forma realista de falarmos sobre o que nós somos e fazemos. No entanto é de racionalidade que falamos, não se trata de mera expressão de sentimentos. Acontece que essa forma de sermos racionais que é fazermos juízos éticos sobre a forma como as coisas são no mundo tem como base percebermos o mundo de uma certa maneira – noutras palavras: não é pela argumentação que começamos, é, se quisermos dizer assim, pela percepção, pela experiência. É esse em última análise o ponto de Diamond: perceber moralmente é uma forma de perceber, que coloca todos os problemas que usualmente a percepção coloca.

Portanto não se trata de dizer, com Hume, que a base do comportamento moral não é a razão – será que a alternativa é Kant, a ideia segundo a qual a base do comportamento moral é o raciocínio prático puro? Para Putnam, pelo menos, já que continuo aqui a orientar-me por ele, a ideia é: Kant sim, mas Kant não basta. Porquê? Os problemas do kantismo podem ser evidenciados por um autor europeu que interessa muito Putnam, Jürgen Habermas, que Putnam considera um dos mais importantes filósofos europeus actuais e cujas posições discute em *The Collapse of the Fact-Value Dichotomy*. Inspirado em Kant, aliás no quadro da inspiração kantiana da sua filosofia, Habermas tem defendido uma distinção entre normas e valores. O que é esta distinção? Como é que ela sai de Kant? Por que é que Putnam a critica?

A discussão entre Putnam e Habermas evidencia as consequências para a ética de uma certa ideia de racionalidade prática. Vamos chamar à posição de Habermas o lance de um kantiano consequente. Coloquemos a questão em termos de valores e normas: a ideia de Habermas é que as normas são racionais e os valores substanciais; queremos alguns imperativos categóricos, mas não demasiados. O que Putnam diz é: Mas

Jurgen, então na tua forma de ver as coisas, os valores – enquanto opostos às normas – são tão não-cognitivos como para os positivistas!¹⁵⁹ Talvez não seja uma crítica totalmente justa, mas é acríica de Putnam a Habermas. O que temos que perceber aqui é que para Habermas, existe uma única norma universal vinculativa – a da razão comunicacional, com o seu ideal de consenso racional. Mas para Putnam a dicotomia habermasiana norma-valor é extremamente problemática. E mostra que há algum problema em Kant quanto a ‘valores substanciais’. Qual problema exactamente? O facto de a razão kantiana não parecer poder ser guiada por nenhuns fins substanciais; só há um princípio da razão pura prática. Algumas pessoas falaram de naturalismo reducionista quanto a valores, tal como o defendido pela colega de Putnam em Harvard, e influente kantiana, Christine Korsgaard: Kant não pode aceitar nenhum tipo de realismo substancial de valores.

Conclusão

Vou terminar procurando dizer directamente onde é que Putnam, falando de ética, quer finalmente chegar. Putnam quer atacar o cepticismo moral, como sempre quis atacar o cepticismo filosófico em geral. Mas quer atacar o cepticismo moral sem por isso, ao defender o objectivismo (termo que prefere a ‘realismo’), ser interpretado como defendendo qualquer tipo de apriorismo ou ‘concepção absoluta de mundo’. Em termos de história da filosofia, isto traduz-se em pensar que em ética precisamos dos insights kantianos mas também precisamos dos insights aristotélicos. Quais são os insights kantianos de que precisamos? Precisamos da ideia de que não há entidades e propriedades platónicas – não há ponto de vista sem ponto de vista, na expressão da Susana. É isso que o nome de Kant significa na história da filosofia e também na ética. O problema é que Kant via a natureza do ponto de vista ético como consistindo em racionalidade pura; essa era a sua porta de entrada em ética – e isso é hoje traduzida por kantianos contemporâneos como Habermas como uma ética do discurso e com a separação normas-valores. Ora Putnam pensa que tem que haver mais em ética do que apenas a ética do discurso – kantismo não é suficiente porque normas universais kantianas não exaurem aquilo que é objectivo em ética. Daí a necessidade

¹⁵⁹ Putnam 2002:

do insight aristotélico: à necessidade de considerar o *ponto de vista* junta-se a necessidade de considerar a *substância*, sendo que essa só pode ser dada pelo contexto, e aí se colocam questões de objectividade e discussão de valores.

É neste enquadramento que valores, ou valorizações, éticos podem ser racionalmente discutidos e não precisam de ser naturalizados. Esta forma de se posicionar em ética leva-nos directamente ao pragmatista que ele é. Vamos por isso imaginar como responderia Putnam à pergunta “Por que sou um pragmatista?” Ele evocaria 1) o anti-cepticismo, 2) o falibilismo, 3) a crítica à dicotomia factos-valores, 4) a convicção de que a prática é, de um ponto de vista filosófico, primeira ou primária¹⁶⁰. E segundo Putnam, estas orientações valem não apenas para a ética mas para qualquer área da filosofia em que pensamos sobre como pensamos.

Londres-Porto 2011

Referências

- Cadilha, Susana, Susana Cadilha, “Naturalização da ética: pressupostos, compromissos e implicações metaéticas”, in Miguens, S., Pinto, J. A. e Teles, M., *Aspects of Judgement*, , vol. 4 MLAG Discussion Papers, Porto, FL-UP, 2010.
- Galvão, Pedro, “Ética”, in Galvão (coord), *A Filosofia por disciplinas*, no prelo.
- Guisán, Esperanza, 2010, *Introducción a la Ética*, Madrid, Cátedra (4ª edición)
- Kane, Robert, 2002, *The Oxford Handbook of Free Will*, Oxford, Oxford University Press
- Ogien, Ruwen, 2007, *L'éthique aujourd'hui – maximalistes et minimalistes*, Paris, PUF.
- Putnam, Hilary, 1994, “Pragmatism and moral objectivity”, in Putnam 1994 *Words and Life*, Cambridge MA, Harvard University Press, edited by James Conant.
- Putnam, Hilary, 2002, *The Collapse of the Fact-Value Dichotomy*, Cambridge MA, Harvard University Press,
- Putnam, Hilary, 2004, *Ethics Without Ontology*, Cambridge MA, Harvard University Press,
- Shafer-Landau, Russ, 2006, *Oxford Studies in Meta-Ethics – vol. 1*, Oxford, Oxford University Press.
- Wiggins, David, *Ethics – Twelve Lectures on the philosophy of morality*, London, Penguin, 2006

¹⁶⁰ Putnam 1994, Pragmatism and Moral Objectivity, p. 152.

Em defesa da meta-ética

Susana Cadilha

I.

Sofia Miguens - “A resposta que gostaria de dar é que as questões éticas não são nem questões de semântica nem questões de metafísica: são questões éticas, e enquanto questões éticas são questões filosóficas, sendo que qualquer questão filosófica nos conduz a esclarecimentos semânticos e metafísicos, sem que tenhamos que dizer que ‘as questões éticas são em última análise questões semânticas porque são acerca da forma como falamos (acerca de assuntos como o bem e o mal) ou são em última análise questões metafísicas porque são acerca do que há, acerca da existência ou não existência de um tipo especial de factos e propriedades que seriam “morais”.”

O ponto que se pretendia aqui defender é que as questões éticas não são de natureza diferente de outro qualquer tipo de questões filosóficas – e, nessa medida, desencadeariam problemas semânticos, metafísicos ou epistemológicos semelhantes aos que estão presentes em qualquer outro tipo de questões filosóficas.

Em fundo está uma certa reserva em aceitar as distinções programáticas entre metaética e ética normativa, ou entre questões metateóricas e questões substanciais, distinções que ao que percebo lhe parecem escusadas ou superficiais porque assentes numa separação facto-valor que não temos que aceitar. Ora, o problema é que não concebo que a legitimidade da metaética enquanto área de estudo filosófico dependa dessa separação facto-valor.

Parecem-me, de facto, questões diferentes a de querer saber, por um lado, o que devemos fazer, qual a melhor forma de agir ou o que devemos valorizar, e, por outro, a de querer saber o que é isso que fazemos quando pensamos moralmente, e de perguntar pelo fundamento ou ancoragem desse discurso. Não me parece que isto indicie uma forma de ver o domínio ético como separado de outros domínios de pensamento. Simplesmente se concede que de entre as questões filosóficas mais gerais que visam a ligação entre o pensamento e a linguagem e o mundo, as que se ligam ao pensamento e ao discurso éticos levantam problemas específicos (desde logo porque o discurso ético tem a particularidade de guiar/orientar/incitar a acção, o que não se passa com todos os tipos de discurso). Mas nada está ainda a ser dito acerca da separação facto-valor. Aliás, essa separação não é um dogma que atravessa toda a investigação metaética, desde logo porque não somos obrigados a concordar com ela

(e ainda assim estaremos a ‘fazer’ metaética, i.e., a pensar sobre a natureza do nosso pensamento moral). De facto, qualquer questão filosófica pode conduzir-nos à necessidade de chegarmos a esclarecimentos semânticos ou metafísicos, mas isso não significa que todas levem os mesmos problemas semânticos ou metafísicos. Assumir isto é apenas reconhecer um problema filosófico, não é já a assunção de uma solução, que seria a admissão da separação factó-valor.

O que quero dizer é que a Professora Sofia parece fazer uma associação entre a simples admissão de que faz sentido uma investigação metaética e uma certa posição já delineada acerca da natureza dos juízos éticos – como não sendo descritivos de nenhuma realidade moral, como não passando de expressões de atitudes subjectivas, sentimentos ou emoções (o que equivale já a uma certa forma de responder às questões metaéticas, que só faz sentido se se pressupuser a tal separação factó-valor). Como se levantar questões metaéticas fosse já comprometer-nos com um particular tipo de solução, o não-cognitívismo. Ora, historicamente, pode ser o caso que a metaética enquanto divisão do campo de estudos filosófico tenha emergido depois da proposta do positivismo lógico (embora me pareça que as suas origens são anteriores, remontam a G.E.Moore) e de uma particular visão do que pode ser considerado conteúdo cognitivo. Mas uma vez pondo em causa esse critério positivista e a distinção factó-valor que lhe está associada, como fizeram P. Foot ou E. Anscombe, as questões metaéticas não desaparecem, antes recrudescem, porque continuam, quanto a mim, a fazer sentido.

Parece-me, claramente, que a metaética enquanto área de estudo perdeu há já algum tempo a ligação que a prendia a um determinado momento histórico e filosófico – o positivismo lógico – e a uma determinada posição quanto ao estatuto dos juízos éticos – o não-cognitívismo, nas suas variadas formas. Se é verdade que, na senda do positivismo lógico inspirado no Wittgenstein do *Tratatus*, a metaética se instituiu como aquele domínio do pensamento filosófico que lidava com o *nonsense* ético, visto como mera expressão de sentimentos e atitudes (e por isso mesmo, a metaética era sobretudo uma questão de semântica e de análise de linguagem), também é verdade que veio depois a evoluir e passou a identificar-se com um conjunto de problemas

muito mais vasto, desligado dos compromissos iniciais e da vocação semântica original.¹⁶¹

Sofia Miguens - “Não é vão que seja Putnam, um *insider* quando se trata de pensar na forma como se pensa ao fazer física ou lógica, a defender que não podemos (não devemos!) separar o nosso pensamento sobre a natureza dos nossos pensamentos em lógica ou em física, ou crenças perceptivas, do nosso pensamento sobre os nossos pensamentos éticos.”

Parece-me claro que haverá problemas filosóficos comuns que perpassam o debate sobre a natureza do pensamento – seja este moral, lógico ou científico; problemas como a natureza da representação ou da verdade, por exemplo. Parece-me também adequado como princípio não partir desde logo do pressuposto de que esses são domínios do pensamento estanques e sem ligações entre si, uma vez que tal resultará no enviesamento da discussão, desde o início. Mas tais ‘princípios metodológicos’, chamemos-lhe assim, não obviam o facto de que o pensamento moral levanta problemas próprios, diferentes até certo ponto dos que são suscitados por exemplo pelo pensamento científico, e isto não representa desde logo a assunção da distância entre facto-valor, mas apenas a assunção de que para além das questões substanciais da ética prática que temos que resolver (o que devemos fazer?), existem questões de pleno direito que se levantam a respeito das próprias asserções éticas – e podem ser questões de âmbito metafísico, semântico ou epistemológico, na mesma medida em que a filosofia da ciência ou a estética também dão origem a questões de âmbito metafísico, semântico ou epistemológico. Isto é, parece-me que a separação facto-valor pode ser rebatida, como defende Putnam, mas continuamos a precisar de explicar o que estamos a fazer quando emitimos juízos morais, nomeadamente, precisamos de explicar como é que tais juízos, que se assemelham em muitos aspectos aos juízos perceptivos (ou a qualquer outro tipo de juízo), podem ter a força que têm sobre nós – e que não partilham com os juízos perceptivos. Se é verdade que “descrição e avaliação estão entrelaçadas” em qualquer juízo que façamos, que não podemos pura e simplesmente dividir os juízos em duas categorias (descritivos e normativos), também é verdade que fica ainda por explicar porque é que em alguns

¹⁶¹ Cf., a este propósito, por exemplo, o artigo que traça a história da metaética, das suas tendências e viragens temáticas, desde G.E. Moore até aos nossos dias: S. Darwall, A. Gibbard e P. Railton, “Toward *Fin de siècle* Ethics: Some Trends”, in Darwall, Gibbard & Railton (eds.), *Moral Discourse and Practice – some philosophical approaches*, New York: Oxford University Press, 1997, pp. 3-47.

juízos está presente uma ligação directa a uma capacidade motivacional e em outros não (porque é que alguns desses juízos nos motivam a agir e outros não?).

De um ponto de vista abrangente, claro que os problemas filosóficos são transversais a toda e qualquer forma de pensamento humano; mas se focarmos mais o olhar sobre determinadas expressões desse pensamento, descobriremos questões mais particulares. Admitir isto não é rendermo-nos à separação facto-valor, assumindo que o nosso pensamento segue regras diferentes quando pensamos no que devemos fazer ou quando pensamos acerca do funcionamento de uma célula. É apenas admitir que a metaética tem razão de ser enquanto campo de estudos filosóficos – mesmo que possamos discordar quanto às respostas, tudo o que para já estou a dizer é que faz sentido fazer certas perguntas. Perguntas como estas: o que estou propriamente a afirmar quando digo que X é moralmente errado? Alguma coisa acerca do mundo, acerca de mim ou acerca do que as pessoas pensam? Há uma realidade moral, composta por factos ou propriedades morais? Se sim, como chegamos a ter deles conhecimento?

A investigação ética não constituirá um campo de estudos à parte, desligado das conclusões a que possamos chegar acerca da relação entre pensamento, linguagem e mundo, mas coloca questões próprias. Portanto, concordo com a ideia de que não há necessidade de ‘reduzir’ as interrogações éticas a outro tipo de interrogações, mas sim que há espaço para colocar essas outras questões no domínio da metaética. O meu ponto é que mesmo não sendo o pensamento ético uma forma de pensar absolutamente singular e dissociada de outras, há problemas que se colocam a propósito do pensamento moral que não têm razão de ser no caso, por exemplo, do pensamento perceptivo. Mais especificamente: saber o que pode contar como um facto é uma questão filosófica que faz sentido colocar qualquer que seja a área filosófica. No entanto, os factos morais – como o facto de que assassinar é errado – seriam factos *sui generis*, na medida em que têm poderes normativos (obrigam-nos a fazer coisas) que o facto de uma mesa ser quadrada não acarreta. Nesta medida, parece que “Assassinar é errado” não é uma descrição de factos do mundo exactamente do mesmo tipo que a descrição “A mesa é quadrada”. Mesmo que não queiramos ir tão longe, temos pelo menos que admitir que um juízo moral suscita problemas específicos. Na realidade, é o próprio Putnam quem nos diz que é necessário manter alguma distinção entre juízos éticos e juízos de outro tipo – mas com a ressalva de que nada de metafísico se segue daí (e que não podemos por isso

simplesmente assumir que não existem coisas como factos morais). Ora, esta não deixa de ser uma posição perfeitamente possível e distinguível no campo de estudos da metaética.

Em suma, tudo quanto queremos manter é que pensar na natureza dos juízos morais é uma tarefa legítima e que nos coloca problemas semânticos, metafísicos e epistemológicos até certo ponto específicos. Mas defender que os juízos morais suscitam questões metateóricas próprias não é ainda afirmar nada acerca do estatuto dos juízos morais (se são susceptíveis de ser avaliados como verdadeiros ou falsos; se são objectiva ou subjectivamente justificados) nem tão pouco nos compromete com uma particular posição em relação à dicotomia facto-valor.

II.

Outro ponto que me parece merecedor de atenção é o seguinte: a separação facto-valor é interpretada como um pressuposto unicamente ligado a uma posição subjectivista no que toca à natureza e justificação dos juízos éticos (mais especificamente, o não-cognitismo). Ora isto não me parece totalmente preciso. Por exemplo, se pensarmos nas teorias que normalmente se denominam como teorias da razão prática (inspiradas em Kant), elas consubstanciam-se numa descontinuidade essencial entre factos e normas ou valores. A ideia crucial é precisamente a de que o discurso moral não pretende ser descritivo nem reportar-se a uma realidade factual; ele resulta antes de uma vocação racional que se apresenta como solucionadora da questão normativa ('o que devemos fazer?') – é uma actividade de construção ou criação racional, de chegar a soluções para os problemas práticos válidas para todos os seres racionais, é uma actividade essencialmente avaliativa e não descritiva, o que não significa que seja menos objectiva por isso (ou pelo menos é essa a ambição de tais teorias). Portanto, a dicotomia facto-valor, ou descrição-avaliação, não está estritamente ligada a uma teoria subjectivista acerca dos valores nem a uma concepção de que o pensamento moral está fora da esfera da razão e que portanto os fins da acção não podem ser racionalmente avaliados.

Se há pouco dizia que a investigação metaética não nos compromete desde logo com uma posição não-cognitivistica (que assenta numa distinção facto-valor), agora o ponto é ligeiramente diferente: essa separação entre facto e valor que Putnam considera um dogma da filosofia moral não é exclusiva de uma posição subjectivista. Isto é

importante porque, em fundo, me parece que esse é o problema principal aqui em jogo. Apelar a uma descontinuidade entre facto e valor poderia deixar-nos com um problema, o de que o discurso acerca de valores fosse mero confronto de opiniões subjectivas, que nada de objectivo e racionalmente sustentado pudesse ser dito acerca de questões valorativas. Mostrar que tal separação é um dogma injustificado parece ser para Putnam a prova de que a ética não é um domínio à parte, deixado a considerações subjectivistas. Ora, parece-me que essa associação de ideias não é absolutamente necessária – a separação entre facto e valor não nos deixa fadados a uma concepção subjectivista da ética, particularmente se considerarmos que a objectividade não é uma noção metafísica mas epistemológica.

Se pensarmos, por exemplo, em R. Dworkin, filósofo americano professor de filosofia e direito, ou ainda nos filósofos T. Nagel e C. Korsgaard, vemos como a sua defesa da objectividade da ética passa precisamente pela tentativa de afastá-la da inteligibilidade científica, i.e., do que seria uma tarefa descritiva. O ponto desses autores é que não podemos esperar aplicar ao domínio normativo o mesmo “quadro epistemológico” da ciência porque isso conduzir-nos-ia ao que Dworkin designa como a “absurda tese do campo moral”¹⁶²: a ideia de que o universo acolhe, para além das partículas de energia e matéria, umas partículas ‘especiais’ que constituiriam a moralidade ou imoralidade das acções humanas. Para Dworkin, a epistemologia científica, de acordo com a qual se estabelece o que é uma crença científica justificada, por exemplo, adequa-se ao pensamento acerca do mundo físico, mas não se adequa de todo ao pensamento moral (“uma vez que a moral e outros domínios avaliativos não fazem qualquer afirmação causal”¹⁶³). E isto precisamente porque a tarefa da ética não é a mesma tarefa da ciência; não se trata de descrever como as coisas são, mas de dizer como elas devem ser. Para estes autores, o que importa não é o estatuto ontológico dos supostos factos morais, mas a qualidade dos argumentos morais. A ética é uma forma de pensamento objectiva e sustentada porque racionalmente justificada, e não porque descreve propriedades que fazem parte do tecido do mundo.

¹⁶² Cf. R. Dworkin, “Objectivity and Truth: you’d better believe it”, *Philosophy and Public Affairs*, 25 (1996): 87-139.

¹⁶³ Cf. R. Dworkin, *idem*, p. 120. Ele adianta ainda: “a moralidade é uma dimensão distinta e independente da nossa experiência, e exerce a sua própria autoridade” (*ibidem*, p. 127).

O que tais autores defendem, portanto, é, precisamente a autonomia do domínio moral perante outras formas de pensamento e discurso, a independência do que deve ser daquilo que é o caso. A separação factó-valor não é, portanto, um dogma sustentado por uma particular posição na filosofia analítica, ligada ao positivismo lógico e ao subjectivismo ético (sob a forma de não-cognitívismo). O que faz tais autores de inspiração kantiana querer manter a dicotomia factó-valor e a autonomia do domínio moral é precisamente o receio que à ética aconteça o mesmo que à epistemologia após o colapso, com Quine, da distinção analítico-sintético. Depois do colapso da distinção analítico-sintético, o naturalismo viu as suas fronteiras espriarem-se e os que se reclamavam naturalistas viam-se obrigados a rejeitar qualquer espécie de filosofia primeira, de reduto filosófico imune aos avanços da ciência. A vontade de manter a distinção factó-valor será um dogma, talvez, mas defendido também por parte dos que querem manter a ética como esse reduto filosófico não contaminado pela ciência. Porque se descrição e avaliação estiverem entrelaçadas e forem interdependentes, isso significa também que aquilo que sabemos em termos descritivos acerca do mundo, do funcionamento das nossas mentes e até das nossas sociedades terá implicações sobre as nossas avaliações normativas. Manter a descontinuidade entre factó e valor serve, portanto, para impedir que o conhecimento científico possa ter alguma coisa a dizer à ética, para manter a diferença entre o comportamento moral humano e outras formas de comportamento pró-social que não envolvam capacidades de conceptualização; enfim, para cristalizar, de certa forma, uma distância de segurança entre o homem e o animal. Sustentar que as questões de valor não são questões de factó pode entender-se, portanto, como um pressuposto já não dos que querem deixar a ética num pântano subjectivista, mas dos que entendem que a nossa dimensão moral, porque racional, deve ser entendida como um acrescento, um ponto a mais relativamente à nossa dimensão animal, e, por isso mesmo, deve ser compreendida em moldes diferentes.

O próprio McDowell, apesar da inovação trazida pelo conceito de segunda natureza (que nos permite pensar que a nossa dimensão racional é também natural) não deixa de querer manter bem nítida uma diferença entre o ‘espaço lógico das razões’ e o ‘espaço lógico da natureza’¹⁶⁴. O que ele não quer é que essa distância permita inferir qualquer espécie de ‘anti-naturalidade’ das razões. Mas isso não significa que não exista uma distância ontológica entre os factos do domínio da lei e os factos racionais

¹⁶⁴ Cf. J. McDowell, *Mind and World*, Cambridge, MA: Harvard University Press, 1994, Introdução.

(se quisermos, podemos dizer que McDowell mantém a distância factó-valor, mas insere ambos os elementos da dicotomia num domínio natural alargado). McDowell vai, portanto, manter que o aparato conceptual das ciências naturais é distinto do que convocamos na esfera da razão, são inteligibilidades distintas; entender o mundo de um ponto de vista científico-natural é distinto do que obtemos quando nos importamos com as razões (revisitação da diferença entre explicar e justificar que vem já desde os alvares das ciências humanas). Será isto ainda um dogma? ¹⁶⁵

O que eu queria neste ponto mostrar é que uma das consequências da rejeição das duplas analítico-sintético e factó-valor, que Putnam considera dogmas injustificados, é precisamente a ideia de que a filosofia não pode mais olhar para o seu objecto de estudo de uma forma desnaturalizada, como se o homem não fizesse parte da natureza, não estivesse submetido às leis da evolução, etc. Mas essa não é uma ilação que muitos filósofos, sobretudo de inspiração racionalista, gostem de tirar. Provavelmente, nem Putnam estará interessado em chegar às últimas consequências do colapso das dicotomias que advoga. Para filósofos como McDowell, por exemplo, essas questões de ‘engenharia’ não importam para a filosofia. Todo o programa da psicologia cognitiva ou evolucionista, que dá conta da caracterização cognitiva de agentes reais, dos seus limites e potencialidades, assim como da nossa história evolutiva, não tem nada a dizer-nos se o que queremos é pensar na nossa condição de seres morais. E isto explica-se, precisamente, porque McDowell traça uma distinção entre o que são as tarefas da ciência – explicativas e descritivas – e as tarefas da filosofia – justificativas e normativas. E também porque traça uma fronteira muito nítida – e, diria eu, demasiado abrupta – entre conceptual e não-conceptual, agentes cognitivos e pensadores, desejar e querer. Defender a autonomia do domínio

¹⁶⁵ E poderá haver alguma contradição aqui? Se, por um lado, um dos desígnios de McDowell (nomeadamente em *Mind, Value and Reality*) é mostrar que não existe uma diferença abismal entre, por exemplo, juízos perceptivos e juízos morais ao nível do seu estatuto epistemológico (na medida em que ambos são susceptíveis de ser avaliados como verdadeiros ou falsos, ambos convocam capacidades conceptuais e resultam de uma sensibilidade adequada, ambos são objectivos), por outro lado quer manter que juízos descritivos e juízos normativos são distintos porque o aparato conceptual convocado num caso e noutra é distinto, e dão origem a inteligibilidades distintas. Como sempre, McDowell quer ficar com o melhor de dois mundos: que juízos morais não se distingam de outro tipo de juízos, para que possamos com legitimidade falar de coisas como verdade, correcção e objectividade a propósito também desses juízos. Ao mesmo tempo, porém, quer manter a distância entre o domínio normativo e o domínio descritivo para não cair num naturalismo galopante nem em nenhum tipo de falácia naturalista.

normativo passa precisamente por considerar que quando pensamos na nossa dimensão moral estamos em pleno território das exigências da razão, e tudo quanto diz respeito às nossas características mais básicas e instintivas como que emudece. McDowell, por exemplo, é peremptório: as exigências da razão não dependem de factos acerca dos seres humanos que sejam capturáveis independentemente da estrutura das razões. O ser sensível às exigências da razão (e, portanto, o ser sensível às razões morais) é uma característica natural do homem, mas não é uma característica que esteja relacionada com outras características naturais. Porquê? Não subsiste ainda aqui algum resquício do dogma/da dicotomia?

O que me parece é que pelo facto de sermos racionais não pertencemos a outro mundo e, portanto, o que explica a moralidade será o mesmo que explica demais aptidões humanas. McDowell – e outros filósofos na mesma linha – aproximam-se perigosamente de Kant (e do que, hoje em dia, se vê como indefensável no kantismo ético, como pudemos ler no artigo anterior, “Hilary Putnam e a ética – respostas a Susana Cadilha”) ao considerar que quando falamos do dever-ser nos libertamos de todos os determinismos, somos quase vontade numérica. Parece-me demasiado abrupta e forçada essa passagem entre imperativos biológicos e racionais. E parece-me também que, de facto, não é isso que se verifica.¹⁶⁶

O que aqui objectamos é a que o ‘ponto de vista racional’ seja incontaminado. Repare-se que a questão não é a de tentar alojar o normativo no natural. A essa questão McDowell responde com proficiência, por meio do conceito de segunda natureza. O ponto é que se resolve o dualismo razão-natureza, outros dualismos se mantêm (mesmo que a razão não seja em nada alheia ao que é natural).

¹⁶⁶ Há muitas características e aptidões que até há pouco se pensavam exclusivas dos seres racionais – seres que têm ‘mundo’ e não apenas se movimentam num ambiente – mas que estão afinal também presentes em seres não-racionais e não-linguísticos. Falo, por exemplo, do comportamento de logro e dissimulação, da capacidade de ter crenças acerca de crenças, da reciprocidade na atribuição de crenças. (Cf., por exemplo, M. Hauser, *Moral Minds – how nature designed our universal sense of right and wrong*, Harper Collins Publishers, 2006.) Como não pensar que há uma linha de continuidade (e de dependência) entre essas capacidades e o pensamento racional (e moral) plenamente constituído que é o nosso?

Neste ponto, é impossível não nos lembrarmos de Dennett e da ideia de que não existe um corte abrupto entre o que é da ordem do que ele chama os *free floating rationales* - o comportamento que pode ser lido como inteligente e intencional - e o que é da ordem das *full blown reasons* - a racionalidade em grau mais avançado, a consciência de si a si mesmo das razões para agir. Não é crível que o que se dá seja uma passagem de seres determinados biologicamente para seres movidos apenas a razões. Ou, nas palavras de Dennett, não podemos dizer que onde há razões não há causas – as “máquinas semânticas” são uma impossibilidade.

O problema é que pressupor a continuidade e não aquele salto abrupto parece representar, para estes autores, a perfeita anulação do domínio das razões. O problema para McDowell é, parece-me, que se pense na possibilidade de estabelecer uma relação de justificação entre factos naturais e factos morais (em que os primeiros justificariam os segundos). Eu não vejo a relação facto-valor da mesma forma, e talvez por isso ela não me pareça problemática. A meu ver, a continuidade facto-valor não anula o pensamento moral nos seus próprios termos, nem elimina o discurso moral (tornando-o num discurso científico). A linguagem moral continua a fazer pleno sentido e continua a ser a melhor forma de nos entendermos em conjunto. Mas acho que faz sentido des-sacralizar o pensamento moral, trazê-lo à terra. Isto é, entendê-lo como a manifestação de um ser que está sujeito aos mesmos imperativos biológicos e às mesmas regras de evolução que todos os outros, e que é por isso muito mais uma criatura sensível que por acaso pensa do que uma criatura racional que também sente. A ideia não é justificar os factos morais com os factos naturais, porque a dimensão propriamente moral não se anula, mas antes entender ou explicar os factos morais à luz dos factos naturais. Isto talvez faça com que o nosso pensamento moral pareça muito mais contingente e muito menos alicerçado/fundamentado do que gostaríamos que fosse. Isso é verdade, e é algo com que talvez tenhamos que lidar. Mas tal não implica eliminar a dimensão normativa, eliminar as razões, apenas repensar o seu estatuto. Continuamos a ser seres tais que vêem razões para repudiar a escravatura. Mas essa capacidade não aparece desfasada ou apartada de todas as nossas outras capacidades. Vemos essas razões porque temos uma determinada história colectiva natural e social e uma determinada constituição; elas só se entendem completamente se formos capazes de ver o quadro completo. O ponto é que não me parece que tenhamos que adoptar uma posição exclusivista, como se uma explicação biologista, de tipo evolucionista ou mais próxima das teorias sentimentalistas, pusesse em causa ou de algum modo diminuísse a realidade e validade do pensamento moral, ou fosse inconciliável com a ideia de que temos razões para fazer X. O mesmo para as questões entre ciência e filosofia: estar atento ao que a ciência nos diz não significa a irrelevância da filosofia.

III.

Um último ponto, apenas, para pegar em Wittgenstein e na noção de realismo (uma vez que o móbil desta discussão foi uma particular interpretação de Wittgenstein, por

parte de Cora Diamond, do que deve ser a tarefa dos filósofos morais). Tanto Wittgenstein, como Diamond, Putnam ou McDowell dizem-se realistas (tendo o cuidado de dar designações específicas a esse realismo). No que toca ao domínio moral, para estes autores o realismo entende-se desta forma: nós vemos os factos morais porque somos um certo tipo de pensadores, e isto envolve/significa várias coisas: que aprendemos a usar as palavras de certa maneira, estamos imbuídos em certas práticas, fomos iniciados em determinadas tradições e relações humanas. (De acordo com Wittgenstein II, é, aliás, a esse nível que a filosofia deve ficar. Não deve tentar procurar fundamento mais fundo do que esse, mas sim “deixar tudo como está”.) Aquilo que nos dá o direito de dizer ‘isto é vermelho’ ou ‘isto é errado’ é o facto de estarmos imersos numa certa prática, de termos aprendido o modo como se usa a palavra ‘vermelho’ ou ‘errado’ (Diamond diria algo como isto: não há forma mais funda de dizer o que é ser errado; a única coisa que podemos fazer é dizer: ‘olha, isto é errado’ – e é por isso que a literatura pode ser útil).

Mas porque devemos apenas olhar para o lado social ou comunitário – no fundo, para a nossa dimensão linguística – para entender o tipo de pensadores que somos? As capacidades paroquiais a que todos esses autores fazem menção – paroquiais no sentido em que estão disponíveis a um certo tipo de pensador, mas não a qualquer tipo de pensador ¹⁶⁷ – são características de uma determinada espécie, a nossa, e nesse sentido, mais uma vez, não se percebe a razão de manter a distância de segurança entre o que são as investigações cognitivas ou evolucionistas e o pensamento moral. As capacidades paroquiais, que fazem de nós os seres que somos – inclusive, seres capazes de enxergar factos morais – não são apenas as capacidades que dizem respeito à nossa dimensão racional e linguística. Nós vimos equipados com uma determinada constituição, e dessa constituição farão parte tanto as capacidades de categorização ou conceptualização, de trazer as coisas sobre certas generalidades, como a capacidade de distinguir seres animados de inanimados, de ter crenças acerca de crenças, ou a propensão para o altruísmo. Tudo isso explica o facto de sermos capazes de pensamento moral.

¹⁶⁷ Cf. Ch. Travis, “Morally alien thought”, in T. Marvan (ed.), *What determines content*, Newcastle: Cambridge Scholars Press, 2006, pp. 243-270.

Portanto, o meu problema não é com a ideia de que, rejeitando-se o platonismo, tenhamos que negar a realidade moral ou a validade do pensamento moral. Nós somos seres talhados de tal forma que acontece termos acesso a uma determinada área da realidade, e somos por isso capazes de ver que assassinar pessoas é errado, por exemplo. O meu ponto é que esta capacidade se entende à luz de uma história muito mais vasta do que aquela que certos autores racionalistas querem fazer crer. Fazer ancorar o pensamento moral em características menos superiores e distintas do ser humano – como a racionalidade –, e pensar que se tivéssemos outra constituição e outro tipo de necessidades, provavelmente teríamos outro tipo de crenças morais (ou sequer nenhuma) confere alguma fragilidade a esse pensamento e retira-lhe a objectividade que habitualmente associamos ao pensamento racionalmente sustentado. A minha solução é que simplesmente teremos de viver com isso. O problema dos realismos ‘adjectivados’ de que aqui falamos é que a noção de objectividade fica muito limitada também num outro sentido – os juízos morais não são objectivos porque verdadeiros independentemente do que qualquer pensador tem razões para acreditar, mas são verdadeiros apenas para uma classe muito restrita de pessoas, que seriam os virtuosos.

Ou seja, os realismos ‘mitigados’ parecem-me apostas muito mais sensatas do que outro tipo de realismos; só não sei se conseguem sustentar os anseios ou tendências objectivistas e universalistas da ética.

APÊNDICE

Referências bibliográficas do projecto “*Conversations on Practical Rationality and Human Action*” (2007-2011)¹⁶⁸

¹⁶⁸ Esta Bibliografia é uma selecção do material com que trabalhámos.

Alfred Mele – Referências bibliográficas selecionadas

Livros-autor:

Effective Intentions: The Power of Conscious Will. Oxford University Press, 2009.

Free Will and Luck. Oxford University Press, 2006.

Motivation and Agency. Oxford University Press, 2003.

Self-Deception Unmasked. Princeton University Press, 2001.

Autonomous Agents: From Self-Control to Autonomy. Oxford University Press, 1995.

Springs of Action: Understanding Intentional Behavior. Oxford University Press, 1992.

Irrationality: An Essay on Akrasia, Self-Deception, and Self-Control. Oxford University Press, 1987.

Livros–Coordenador:

Free Will and Consciousness: How Might They Work? (co-edited with R. Baumeister & K. Vohs). Oxford University Press, 2010.

Rationality and the Good (co-edited with M. Timmons and J. Greco). Oxford University Press, 2007.

The Oxford Handbook of Rationality (co-edited with J. P. Rawling). Oxford University Press, 2004.

The Philosophy of Action (Oxford Readings in Philosophy series), Oxford University Press, 1997.

Mental Causation (co-edited with J. Heil). Oxford: Clarendon Press, 1993.

Artigos:

"Vetoing and Consciousness" in T. Vierkant, J. Kiverstein & A. Clark (eds.), *Decomposing the Will*, Oxford University Press (forthcoming).

T. Stillman, R. Baumeister & A. Mele, "Free Will in Everyday Life: Autobiographical Accounts of Free and Unfree Actions," *Philosophical Psychology* (forthcoming).

"Libet on Free Will: Readiness Potentials, Decisions, and Awareness," in L. Nadel & W. Sinnott-Armstrong (eds.), *Libet, Free Will, and Responsibility*, Oxford University Press (forthcoming).

"Teleological Explanations of Actions: Anticausalism vs. Causalism," in J. Aguilar & A. Buckareff (eds.), *Causing Human Actions: New Perspectives on the Causal Theory of Action*. MIT Press, 2010, pp. 183-198.

"Scientific Skepticism about Free Will," in T. Nadelhoffer, E. Nahmias, & S. Nichols (eds.), *Moral Psychology: Classical and Contemporary Readings*, Blackwell, 2010, pp. 295-305.

"Conscious Intentions" in J. Campbell, M. O'Rourke, & H. Silverstein (eds.), *Action, Ethics, and Responsibility*. MIT Press, 2010, pp. 85-107.

"Weakness of Will and *Akrasia*" *Philosophical Studies* 150 (2010): 391-404.

"Moral Responsibility for Actions: Epistemic and Freedom Conditions," *Philosophical Explorations* 13 (2010): 101-111.

"Testing Free Will," *Neuroethics* 3 (2010): 161-172.

"Conscious Deciding and the Science of Free Will," in R. Baumeister, A. Mele, & K. Vohs (eds.), *Free Will and Consciousness: How Might They Work?* Oxford University Press, 2010, pp. 43-65.

"Moral Responsibility and History Revisited". *Ethical Theory and Moral Practice* (2009) 12: 463-475.

"Mental Action: A Case Study," in L. O'Brien & M. Soteriou (eds.), *Mental Actions and Agency* Oxford: Clarendon Press, 2009, pp. 17-37.

"Delusional Confabulations and Self-Deception," in W. Hirstein (ed.), *Confabulation: Views from Neuroscience, Psychiatry, Psychology, and Philosophy*. Oxford University Press, 2009, pp. 139-157.

"Have I Unmasked Self-Deception or Am I Self-Deceived?," in C. Martin (ed.), *The Philosophy of Deception*. Oxford University Press, 2009, pp. 260-276.

"Manipulation, Compatibilism, and Moral Responsibility". *Journal of Ethics* (2008) 12: 263-286.

F. Cushman and A. Mele, "Intentional Action: Two-and-a-half Folk Concepts," in J. Knobe & S. Nichols (eds.), *Experimental Philosophy*. Oxford University Press, 2008, pp. 171-188.

"Proximal Intentions, Intention-Reports, and Vetoing," *Philosophical Psychology* 21 (2008): 1-14.

"A Libertarian View of Akratic Action," *Studies in Philosophy and the History of Philosophy* 49 (2008): 252-275.

"Persisting Intentions," *Noûs* 41 (2007): 735-757.

"Self-Deception and Three Psychiatric Delusions: On Robert Audi's Transition from Self-Deception to Delusion," in M. Timmons, J. Greco, & A. Mele (eds.), *Rationality and the Good*, Oxford University Press, 2007, pp. 163-175.

"Reasonology and False Beliefs," *Philosophical Papers* 36 (2007): 91-118.

"Self-Deception and Hypothesis Testing," in M. Marraffa, M. De Caro, & F. Ferreti (eds.), *Cartographies of the Mind*, Kluwer, 2007, pp. 159-167.

"Decisions, Intentions, Urges, and Free Will: Why Libet Has Not Shown What He Says He Has," in J. Campbell, M. O'Rourke, & D. Shier (eds.), *Explanation and Causation: Topics in Contemporary Philosophy*, MIT Press, 2007, pp. 241-263.

"Free Will: Action Theory Meets Neuroscience," in C. Lumer (ed.), *Intentionality, Deliberation, and Autonomy: The Action-theoretic Basis of Practical Philosophy*, Ashgate, 2007, pp. 257-272.

"Practical Mistakes and Intentional Actions," *American Philosophical Quarterly* 43 (2006): 249-260.

"Free Will: Theories, Analysis, and Data," in S. Pockett, W. Banks, & S. Gallagher (eds.), *Does Consciousness Cause Behavior? An Investigation of the Nature of Volition*, MIT Press, 2006, pp. 187-205.

"Libertarianism, Luck, and Control," *Pacific Philosophical Quarterly* 86 (2005): 395-421.

"Agnostic Autonomism Revisited," in J. Taylor (ed.), *Personal Autonomy*, Cambridge University Press, 2005, pp. 109-123.

"Acção humana par excellence ("Human Agency Par Excellence") in F. Mão de Ferro (ed.), *A explicação da interpretação humana*, Lisboa: Edições Colibri, 2005.

J. Gert & A. Mele, "Lenman on Externalism and Amoralism: An Interplanetary Exploration." *Philosophia*, 32 (2005): 275-283.

"The Illusion of Conscious Will and the Causation of Intentional Actions," *Philosophical Topics* 32 (2004): 193-213.

"Can Libertarians Make Promises?" in J. Hyman & H. Steward (eds.), *Agency and Action*, Cambridge University Press, 2004, pp. 217-241.

“Action: Volitional Disorder and Addiction,” in J. Radden (ed.), *The Philosophy of Psychiatry*, Oxford University Press, 2004, pp. 78-88.

“Outcomes of Internal Conflicts in the Sphere of Akrasia and Self-Control,” in P. Baumann & M. Betzler (eds.), *Practical Conflicts*, Cambridge University Press, 2004, pp. 262-78.

“Agents’ Abilities,” *Nous* 37 (2003): 447-470.

“Emotion and Desire in Self-Deception,” in A. Hatzimoysis (ed.), *Philosophy and the Emotions*, Cambridge University Press, 2003, pp. 163-179.

“Soft Libertarianism and Flickers of Freedom,” in D. Widerker & M. McKenna (eds.), *Moral Responsibility and Alternative Possibilities*, Ashgate, 2003, pp. 251-264.

“Philosophy of Action,” in K. Ludwig (ed.), *Donald Davidson*, Cambridge: Cambridge University Press, 2003, pp. 64-84.

H. Beebe & A. Mele, “Humean Compatibilism.” *Mind*, 111 (2002): 201-223.

“Acting Intentionally: Probing Folk Notions,” in B. Malle, L. Moses, & D. Baldwin (eds.), *Intentions and Intentionality: Foundations of Social Cognition*, Cambridge: MIT Press, 2001, pp. 27-43.

“Goal-Directed Action: Teleological Explanations, Causal Theories, and Deviance,” *Philosophical Perspectives*, 14 (2000): 279-300.

“Deciding to Act,” *Philosophical Studies* 100 (2000): 81-108.

“Reactive Attitudes, Reactivity, and Omissions,” *Philosophy and Phenomenological Research* 61 (2000): 447-452.

“Responsibility and Freedom: The Challenge of Frankfurt-Style Cases,” in M. Betzler & B. Guckes (eds.), *Autonomes Handeln*, Berlin: Akademie Verlag, 2000, pp. 25-38.

“Twisted Self-Deception,” *Philosophical Psychology* 12 (1999): 117-137.

“Ultimate Responsibility and Dumb Luck,” *Social Philosophy & Policy* 16 (1999): 274-293. Reimpresso em E. Paul, F. Miller, & J. Paul (eds.), *Responsibility*, Cambridge: Cambridge University Press, 1999.

“Motivation, Self-Control, and the Agglomeration of Desires,” *Facta Philosophica* 1 (1999): 77-86.

“Is There a Place for Intention in an Analysis of Intentional Action?” *Philosophia* 27 (1999): 419-432.

“Autoinganno e Controllo Delle Ipotesi” (Self-Deception and Hypothesis Testing; traduzido por Massimo Marraffa), *Sistemi Intelligenti* 11 (1999): 503-519.

- “Motivational Strength,” *Noûs* 32 (1998): 23-36.
- “Noninstrumental Rationalizing,” *Pacific Philosophical Quarterly* 79 (1998): 236-250.
- “Motivated Belief and Agency,” *Philosophical Psychology* 11 (1998): 353-369.
- “Flickers of Freedom,” *Journal of Social Philosophy* 29 (1998): 144-156.
- “Two Paradoxes of Self-Deception,” in J. Dupuy (ed.), *Self-Deception and Paradoxes of Rationality*, Stanford: CSLI Publications, 1998, pp. 37-58.
- “Practical Irrationality: Two Kinds of Akratic Action,” *Cadernos de Filosofia* 4 (1998): 9-37.
- “Agency and Mental Action,” *Philosophical Perspectives* 11 (1997): 231-249.
- “Passive Action,” in G. Holmström-Hintikka & R. Tuomela (eds.), *Contemporary Action Theory*, vol. 1, Dordrecht: Kluwer, 1997, pp. 135-143.
- P. Livingston & A. Mele, “Evaluating Emotional Responses to Fiction,” in M. Hjort & S. Laver (eds.), *Emotion and the Arts*, New York: Oxford University Press, 1997, pp. 157-176.
- “Internalist Moral Cognitivism and Listlessness,” *Ethics* 106 (1996): 727-753.
- “Socratic Akratic Action,” *Philosophical Papers* 25 (1996): 149-159.
- “Rational Intentions and the Toxin Puzzle,” *Proto Sociology* 8/9 (1996): 39-52. Reimpresso em G. Preyer & G. Peter (eds.), *The Contextualization of Rationality*, Mentis: Paderborn, 2000.
- A. Mele & S. Sverdlik, “Intention, Intentional Action, and Moral Responsibility.” *Philosophical Studies* (1996) 82: 265-287.
- “Conceptualizing Self-Control,” *Behavioral and Brain Sciences* 18 (1995): 136-137.
- “Justifying Intentions,” *Mind* 102 (1993): 335-337.
- “History and Personal Autonomy,” *Canadian Journal of Philosophy* 23 (1993): 271-280.
- “Intending for Reasons,” *Mind* 101 (1992): 327-333.
- F. Adams & A. Mele, “The Intention/Volition Debate.” *Canadian Journal of Philosophy* 22 (1992): 323-338.
- J. Heil & A. Mele, “Mental Causes.” *American Philosophical Quarterly* 28 (1991): 49-59.

- “Irresistible Desires,” *Noûs* 24 (1990): 455-472.
- “Exciting Intentions,” *Philosophical Studies* 59 (1990): 289-312.
- “He Wants to Try,” *Analysis* 50 (1990): 251-253.
- “Errant Self-Control and the Self-Controlled Person,” *Pacific Philosophical Quarterly* 71 (1990): 47-59.
- “Intention, Belief, and Intentional Action,” *American Philosophical Quarterly* 26 (1989): 19-30.
- “She Intends to Try,” *Philosophical Studies* 55 (1989): 101-106.
- “Intentions by Default,” *Pacific Philosophical Quarterly* 70 (1989): 155-166.
- “Akratic Feelings,” *Philosophy and Phenomenological Research* 50 (1989): 277-288.
- “Motivational Internalism: The Powers and Limits of Practical Reasoning,” *Philosophia* 19 (1989): 417-436. Reimpresso em R. J. Wallace (ed.), *Reason, Emotion, and Will*, Ashgate, 1999.
- “Against a Belief/Desire Analysis of Intention,” *Philosophia* 18 (1988): 239-242.
- A. Mele & M. Smith, “The New Paradox of the Stone.” *Faith and Philosophy* 5 (1988): 283-290.
- “Are Intentions Self-Referential?” *Philosophical Studies* 52 (1987): 309-329.
- “Intentional Action and Wayward Causal Chains: The Problem of Tertiary Waywardness,” *Philosophical Studies* 51 (1987): 55-60.
- “Is Akratic Action Unfree?” *Philosophy and Phenomenological Research* 46 (1986): 673-679.
- “Aristotle on *Akrasia*, *Eudaimonia*, and the Psychology of Action,” *History of Philosophy Quarterly* 2 (1985): 375-393. Reimpresso em N. Sherman (ed.), *Aristotle’s Ethics: Critical Essays*, Rowman and Littlefield, 1999.
- “How to Represent Aristotelian Deliberation Syllogistically,” *The New Scholasticism* 59 (1985): 484-492.
- “Aristotle on the Roles of Reason in Motivation and Justification,” *Archiv für Geschichte der Philosophie* 66 (1984): 124-147.
- “Aristotle’s Wish,” *Journal of the History of Philosophy* 22 (1984): 139-156.
- “Aristotle on the Proximate Efficient Cause of Action,” *Canadian Journal of Philosophy*, Supplementary Vol. X (1984): 133-155.

“Choice and Virtue in the *Nicomachean Ethics*,” *Journal of the History of Philosophy* 19 (1981): 405-423.

“The Practical Syllogism and Deliberation in Aristotle’s Causal Theory of Action,” *The New Scholasticism* 55 (1981): 281-316.

“Aristotle on *Akrasia* and Knowledge,” *The Modern Schoolman* 58 (1981): 137-157.

“On ‘Happiness and the Good Life’,” *Southwestern Journal of Philosophy* 10 (1979): 181-187.

Algumas entradas em enciclopédias e dicionários:

"Free Will", *International Encyclopedia of Ethics*, Blackwell, forthcoming.

“Action Theory”, *Encyclopedia Americana*, forthcoming.

"Self-Deception", in E. Craig (ed.), *Routledge Encyclopedia of Philosophy*, Routledge, 2010. <http://www.rep.routledge.com/article/V033>

“Freedom of Will”, in M. Binder, N. Hirokawa & U. Windhorst (eds.), *Encyclopedia of Neuroscience*, Springer, 2009, pp. 1631-34.

“Free Will,” in W. Banks (ed.), *Encyclopedia of Consciousness*, Elsevier, 2009, pp. 265-277.

“Action”, in D. Borchert (ed.), *Encyclopedia of Philosophy: Second Edition*, New York: Macmillan, 2006, vol. 1, pp. 14-22.

“Weakness of the Will”, in D. Borchert (ed.), *Encyclopedia of Philosophy: Second Edition*, New York: Macmillan, 2006, vol. 9, pp. 728-732.

“Agent Causation” (p. 18); “Belief and Desire” (pp. 85-86); “Counterexample, Philosophy By” (p.181); “Moral Motivation” (p. 622); “Practical Syllogism,” (p. 747), in T. Honderich (ed.), *Oxford Companion to Philosophy, 2nd ed.*, Clarendon Press, 2005.

“Action, Philosophical Issues About”, in L. Nadel (ed.), *Encyclopedia of Cognitive Science*, Nature Publishing Group, 2003, vol. 1, pp. 20-23.

“Self-Control” (pp. 1548-51); “Temperance” (pp. 1693-96), in L. Becker & C. Becker (eds.), *Encyclopedia of Ethics*, 2nd. ed., New York: Routledge, 2001.

“Accidie” (p. 6), “Control” (p. 184), “Motivation” (pp. 591-92), “Socratic Paradoxes” (p. 861), “Toxin Puzzle” (pp. 924-25), in R. Audi (ed.), *Cambridge Dictionary of Philosophy*, 2nd ed., Cambridge: Cambridge University Press, 1999.

“Intentionality”, in P. Werhane & R. Freeman (eds.), *Encyclopedic Dictionary of Business Ethics*, Oxford: Blackwell, 1997, pp. 339-340.

“Basic Action”, in J. Kim & E. Sosa (eds.), *A Companion to Metaphysics*, Oxford: Blackwell, 1995, pp. 44-46.

Hugh McCann – Referências bibliográficas selecionadas

Livros-autor:

The Works of Agency: On Human Action, Will and Freedom. Ithaca, New York: Cornell University Press, 1998.

Creative and Critical Thinking, 2nd ed. (textbook). Em parceria com W. Edgar Moore e Janet McCann. Boston: Houghton Mifflin, 1985.

Artigos:

“Is Raising One’s Arm a Basic Action?” *The Journal of Philosophy* 69 (1972): 235-249.

“Volition and Basic Action.” *The Philosophical Review* 83 (1974): 451-473.

“Trying, Paralysis, and Volition.” *The Review of Metaphysics* 28 (1974-75): 423-442.

“Nominals, Facts, and Two Conceptions of Events.” *Philosophical Studies* 34 (1979): 129-149.

“On Mental Activity and Passivity: A Reply to Thalberg.” *Mind* 88 (1979): 592-596.

“The Trouble with Level-Generation.” *Mind* 91 (1982): 481-500.

“Individuating Actions: The Fine-Grained Approach.” *Canadian Journal of Philosophy* 13 (1983): 493-512.

“Rationality and the Range of Intention.” *Midwest Studies in Philosophy* 10 (1985): 191-211.

“Intrinsic Intentionality.” *Theory and Decision* 20 (1986): 247-273.

“Divine Conservation and the Persistence of the World.” (com Jonathan Kvanvig). In T. V. Morris, ed., *Divine and Human Action: Essays in the Metaphysics of Theism*. Ithaca, New York: Cornell University Press, 1988, pp. 13-49.

“Intending and Planning: A Reply to Mele.” *Philosophical Studies* 55 (1989): 107-110.

“Practical Rationality: Some Kantian Reflections.” *Journal of Philosophical*

Research 15 (1990): 57-77.

“Settled Objectives and Rational Constraints.” *American Philosophical Quarterly*, 28 (1991): 25-36. Reimpresso em A. R. Mele. ed., *Readings in the Philosophy of Action*. New York: Oxford University Press, 1997, pp. 204-222.

“The Occasionalist Proselytizer: A Modified Catechism.” (com Jonathan Kvanvig). In J. E. Tomberlin, ed., *Philosophical Perspectives* 5, *Philosophy of Religion*. Atascadero, California: Ridgeview Publishing Co., 1991, pp. 587-615.

“The God Beyond Time.” In L. P. Pojman, ed., *Philosophy of Religion*, second edition. Belmont, California: Wadsworth, 1993, pp. 231-245.

“Dretske on the Metaphysics of Freedom.” *Canadian Journal of Philosophy* 23 (1993): 617-628.

“Intenzione e Forza Motivazionale.” *Discipline Filosofiche* 2 (1993): 251-272.

“Paralysis and the Springs of Action.” *Philosophia* 23 (1993): 193-205.

“Divine Sovereignty and the Freedom of the Will.” *Faith and Philosophy* 12 (1995): 582-598.

“On When the Will Is Free.” In G. Holmström-Hintikka and R. Tuomela, eds., *Contemporary Action Theory*, vol. I. Dordrecht, Netherlands: Kluwer Academic Publishers, 1997, pp. 219-232.

“Sovereignty and Freedom: A Reply to Rowe.” *Faith and Philosophy* 18 (2001): 110-116.

“Edwards on Free Will.” In P. Helm and O. D. Crisp, eds., *Jonathan Edwards: Philosophical Theologian*. Aldershot, England: Ashgate Press, 2003, pp. 27-43.

“The Author of Sin?” *Faith and Philosophy* 22 (2005): 144-159.

“Intentional Action and Intending: Recent Empirical Studies.” *Philosophical Psychology* 18 (2005): 737-748.

“Resisting Naturalism: The Case of Free Will.” In A. Corradini, S. Galvan and E. J. Lowe, eds., *Analytic Philosophy Without Naturalism*. Boston: Routledge, 2006, pp. 225-240.

“Metaethical Reflections on Robert Audi’s Moral Intuitionism.” In J. Greco, A. Mele, and M. Timmons, eds., *Rationality and the Good*. New York: Oxford University Press, 2007.

“The Will and the Good.” In C. Lumer, ed., *Intentionality, Deliberation, and Autonomy: The Action-theoretic Basis of Practical Philosophy*, Ashgate, 2007.

“Pointless Suffering: How to Make the Problem of Evil Sufficiently Serious.” In J. L.

Kvanvig, ed., *Oxford Studies in the Philosophy of Religion*, vol. 1. New York: Oxford University Press, forthcoming.

Entradas em enciclopédias e dicionários:

“Action, Philosophy of,” “Practical Reason,” “Practical Reasoning,” and “Reasons for Action.” In R. Audi, ed., *The Cambridge Dictionary of Philosophy*. New York: Cambridge University Press, 1995.

“Creation and Conservation.” In C. Taliaferro and P. L. Quinn, eds., *A Companion to the Philosophy of Religion*. Oxford: Basil Blackwell, 1997, pp. 306-312.

“Creation.” In A. Hastings, ed., *Oxford Companion to Religious Thought*. New York: Oxford University Press, 2000, pp. 143-144.

“Divine Power and Action.” In W. Mann, ed., *The Blackwell Guide to Philosophy of Religion*. Oxford: Blackwell, 2005, pp. 26-47.

“Creation and Conservation.” In D. Borchert, ed., *Encyclopedia of Philosophy*, 2nd ed. Detroit: Macmillan Reference USA, 2006.

Michael Bratman – Referências bibliográficas selecionadas

Livros-autor:

Structures of Agency: Essays. Oxford University Press, 2007.

Faces of Intention: Selected Essays on Intention and Agency. Cambridge University Press, 1999.

Intention, Plans, and Practical Reason. Harvard University Press, 1987. (Reimpresso em 1999).

Livros-coordenador:

Introduction to Philosophy: Classical and Contemporary Readings. John Perry and Michael Bratman eds., Oxford University Press, 1986.

Artigos:

"Individuation and Action," *Philosophical Studies* 33(1978):367-375.

"Practical Reasoning and Weakness of the Will," *NOUS* 13(1979):153-171.

"Simple Intention," *Philosophical Studies* 36(1979): 245-259.

"Intention and Means-End Reasoning," *The Philosophical Review* 90(1981): 252-265.

"Castaneda's Theory of Thought and Action," in James Tomberlin (ed.), *Agent, Language and the Structure of the World: Essays Presented to Hector-Neri Castaneda with His Replies*, Indianapolis: Hackett, 1983, pp.149-169.

"Taking Plans Seriously," *Social Theory and Practice* 9 (1983):271-287. Reimpresso em R. Jay Wallace (ed.), *Reason, Emotion, and Will*, Ashgate, 1999 e em Elijah Millram, ed., *Varieties of Practical Reasoning*, Cambridge, Mass.: MIT Press, 2001.

"Two Faces of Intention," *The Philosophical Review* 93 (1984):375-405. Reimpresso em A. Mele (ed.), *Philosophy of Action* (the Oxford Readings in Philosophy Series), Oxford University Press, 1997.

"Davidson's Theory of Intention," in Bruce Vermazen and Merrill Hintikka (eds.), *Essays on Davidson: Actions and Events*, Oxford: Oxford University Press, 1985, pp. 13-26.

"What Is Intention?" in Philip R. Cohen, Jerry Morgan and Martha E. Pollack (eds.), *Intentions in Communication*, Cambridge: MIT Press, 1990, pp. 15-31.

(Com D. Israel e M. Pollack) "Toward an Architecture for Resource-Bounded Agents," CSLI Research Report #104 (1987), and SRI International Technical Note 425.

"Intention and Personal Policies," *Philosophical Perspectives III* (1989), 443-469.

(Com D. Israel e M. Pollack) "Plans and Resource-Bounded Practical Reasoning," *Computational Intelligence* 4 (1988): 349-355. Reimpresso em R. Cummins and J. Pollock (eds.), *Philosophy and AI: Essays at the Interface*, Cambridge: MIT Press, 1991.

"Cognitivism about Practical Reason" *Ethics* 102 (1991): 117-128.

"Shared Cooperative Activity," *The Philosophical Review* (1992): 327-341. Tradução para alemão em Hans Bernhard Schmid & David Schweikard, eds., *Collective Intentionality*, Suhrkamp, 2009, pp. 176-193. Tradução para castelhano em Jorge Fabra, ed., *Derecho y Convencionalidad*, Editorial Universidad Libre, 2010.

"Practical Reasoning and Acceptance in a Context," *Mind* 101 (1992): 1-14.

"Shared Intention," *Ethics* 104 (1993): 97-113.

"Intention partagee et obligation mutuelle," (tradução para francês por Joelle Proust de "Shared Intention and Mutual Obligation",) in *Les limites de la rationalite: I. Rationalite, ethique et cognition*, Paris: Decouverte, 1997, pp. 246-266.

"Pur une theorie modeste de l'action planifiee: response a Gauthier et Dupuy," (tradução para francês por Jean-Pierre Dupuy de "Toward a Modest Theory of Planning: Reply to Gauthier and Dupuy") in Jean-Pierre Dupuy and Pierre Livet, eds., *Les limites de la rationalite: I. Rationalite, ethique et cognition*, Paris: Decouverte, 1997, pp. 75-87.

"Moore on Intention and Volition," *University of Pennsylvania Law Review* 142 (1994): 1705-1718.

"Planning and Temptation," in Larry May, Marilyn Friedman, & Andy Clark (eds.) *Mind and Morals: Essays on Ethics and Cognitive Science* (Cambridge: Bradford/MIT, 1996): 293-310.

"Following Through with One's Plans: Reply to David Gauthier," in Peter Danielson, ed., *Modeling Rationality, Morality, and Evolution* (Oxford: Oxford University Press, 1998): 55-66.

"Toxin, Temptation, and the Stability of Intention," in Jules L. Coleman and Christopher W. Morris, eds., *Rational Commitment and Social Justice: Essays for Gregory Kavka* (Cambridge University Press, 1998): 59-83.

"I Intend that We J," in Raimo Tuomela and Ghita Holmstrom-Hintikka, eds., *Contemporary Action Theory* vol. II (Dordrecht: Kluwer -- Synthese Library Series, 1997): 49-63. Tradução alemã em Hans Bernhard Schmid & David Schweikard, eds., *Collective Intentionality* (Suhrkamp, 2009): 333-355.

"Identification, Decision, and Treating as a Reason," *Philosophical Topics* 24 (1996): 1-18. Reimpresso em Laura Waddell Ekstrom, ed., *Agency and Responsibility: Essays on the Metaphysics of Freedom* (Boulder, Colorado: Westview Press, 2001).

"Valuing and the Will" *Philosophical Perspectives: Action and Freedom* 14 (2000): 249-265.

"Hierarchy, Circularity, and Double Reduction," in S. Buss and L. Overton, eds., *Contours of Agency: Essays on Themes from Harry Frankfurt* (Cambridge, Mass.: MIT Press, 2002): 65-85.

"Nozick on Free Will," in David Schmidtz, ed., *Robert Nozick* (New York: Cambridge University Press, 2002): 155-174.

"Two Problems About Human Agency," *Proceedings of the Aristotelian Society* (editor: A. W. Price) 101 (2001): 309-326.

"Shapiro on Legal Positivism and Jointly Intentional Activity," *Legal Theory* 8 (2002): 511-517.

"Autonomy and Hierarchy," *Social Philosophy & Policy* 20 (2003): 156-176. Também em E. Paul, F. Miller, & J. Paul, eds., *Autonomy* (Cambridge University Press, 2003): 156-176.

"Shared Valuing and Frameworks for Practical Reasoning," in R. Jay Wallace, Philip Pettit, Samuel Scheffler, & Michael Smith eds., *Reason and Value: Themes from the Moral Philosophy of Joseph Raz* (Oxford University Press, 2004): 1-27.

"A Desire of One's Own," *Journal of Philosophy* 100 (2003): 221-242.

"Planning Agency, Autonomous Agency," in James Stacey Taylor, ed., *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy* (Cambridge University Press, 2005): 33-57.

"Personal Rules and Rational Willpower," *San Diego Law Review* 42: 1 (2005): 61-68.

"What is the Accordion Effect?" *The Journal of Ethics* 10: 1-2 (2006): 5-19.

"Three Theories of Self-Governance" in John Fischer, ed., *Philosophical Topics* 32: 1 and 2 (2004): 21-46.

"Dynamics of Sociality," *Midwest Studies in Philosophy: Shared Intentions and Collective Responsibility* XXX (2006): 1-15.

"Intention, Belief, Practical, Theoretical" in Simon Robertson, ed., *Spheres of Reason: New Essays on the Philosophy of Normativity* (Oxford University Press, 2009): 29-61.

"Anchors for Deliberation," in C. Lumer e S. Nannini, eds., *Intentionality, Deliberation and Autonomy* (Aldershot, etc.: Ashgate, 2007): 187-205.

"Intention, Belief and Instrumental Rationality," in David Sobel e Steven Wall, eds., *Reasons for Action* (Cambridge: Cambridge University Press, 2009): 13-36.

"Shared Agency," in Chris Mantzavinos, ed. *Philosophy of the Social Sciences: Philosophical Theory and Scientific Practice* (Cambridge: Cambridge University Press, 2009): 41-59.

"Reflections on the Philosophy of Action" in Jesus Aguilar & Andrei Buckareff, eds. *Philosophy of Action: 5 Questions* (Automatic Press/VIP, 2009): 17-23.

"Setiya on Intention, Rationality and Reasons," *Analysis* 69 (2009): 510-521.

"Intention Rationality," *Philosophical Explorations* 12 (2009): 227-41.

"Thoughts about the Philosophy of Action," in Sofia Miguens, Carlos Mauro, & Susana Cadilha, eds. *Conversations on Human Action and Practical Rationality*, (no prelo)

"Agency, Time, and Sociality," *Proceedings and Addresses of the American Philosophical Association* (forthcoming)

Entradas em enciclopédias e dicionários:

"Intention" in Robert Audi, ed., *The Cambridge Dictionary of Philosophy* (Cambridge: Cambridge University Press, 1995): 380-381.

"Intention" in Jaegwon Kim & Ernest Sosa, eds., *A Companion to Metaphysics* (Oxford: Basil Blackwell, 1995): 243-244.

"Intention" in Samuel Guttenplan, ed., *A Companion to Philosophy of Mind* (Oxford: Basil Blackwell, 1994): 375-379.

George Ainslie – Referências bibliográficas seleccionadas

Livros-autor:

Picoeconomics. Cambridge: Cambridge University Press, 1992.

Breakdown of will. Cambridge: Cambridge University Press, 2001.

Artigos:¹⁶⁹

"Impulse control in pigeons". *Journal of the Experimental Analysis of Behavior* 21: 485–489, 1974.

"Specious reward: A behavioral theory of impulsiveness and impulse control". *Psychological Bulletin*, 82: 463–496, 1975

(com Herrnstein, R. J.) "Preference reversal and delayed reinforcement". *Animal Learning and Behavior* 9: 476–482, 1981.

"A behavioral economic approach to the defense mechanisms: Freud's energy theory revisited", *Social Science Information* 21: 735-779, 1982.

"Beyond microeconomics: conflict among interest in a multiple self as a determinant of value". In J. Elster (ed.), *The Multiple Self*, Cambridge University Press, pp. 133-175, 1986.

"Freud and Picoeconomics", *Behaviorism* 17: 11-19, 1989.

"Derivation of "rational" economic behavior from hyperbolic discount curves". *American Economic Review* 81: 134–140, 1991.

¹⁶⁹ Muitos artigos podem ser encontrados no site <http://www.picoeconomics.org>

“Intertemporal choice. Derivation of "rational" economic behavior from hyperbolic discount curves”, *American Economic Review* 81: 334-340, 1991.

“A utility-maximizing mechanism for vicarious reward: Comments on Julian Simons "Interpersonal allocation continuous with intertemporal allocation”, *Rationality and Society* 7: 393-403, 1995

“Studying self-regulation the hard way”, *Psychological Inquiry* 7, 16-21, 1996

J. Monterosso, G. Ainslie, P. Toppi Mullen, B. Gault, “The fragility of cooperation: A false feedback study of a sequential iterated prisoner's dilemma”, *Journal of Economic Psychology* 23, 437-448, 2002

(com J.R. Monterosso), Building blocks of self-control: Increased tolerance for delay with bundled rewards, *Journal of the Experimental Analysis of Behavior* 79, 27-48, 2003

“Uncertainty as wealth”. *Behavioural Processes*, 64: 369-385, 2003.

(com J. Monterosso), Hyperbolic discounting as a factor in addiction: A critical analysis. In Rudy Vuchinich & Nick Heather, eds., *Choice, Behavioural Economics and Addiction*, Pergamon, 2003.

“The dangers of willpower: A piceconomic understanding of addiction and dissociation”. In J. Elster & O-J. Skog, (eds.), *Getting Hooked: Rationality and Addiction*, Cambridge University Press, pp. 65-92, 1999

J. Monterosso, G. Ainslie, The behavioral economics of will in recovery from addiction, *Drug Alcohol Depend.* 2007 September; 90 Suppl. 1, 2007

Artigos in progress:

(2009) A. Hofmeyr, G. Ainslie, R. Charlton, D. Ross, “The relationship between addiction and reward bundling: An experiment comparing smokers and non-smokers”

(2005) G. Ainslie, “Emotion as a motivated behavior”

Daniel Hausman – Referências bibliográficas selecionadas

Livros-autor:

Capital, Profits and Prices: An Essay in the Philosophy of Economics. New York: Columbia University Press, 1981.

The Philosophy of Economics: An Anthology. (ed.) New York: Cambridge University Press, 1984.

The Inexact and Separate Science of Economics. Cambridge: Cambridge University Press, 1992.

Essays on Philosophy and Economic Methodology. Cambridge: Cambridge University Press, 1992.

Economic Analysis and Moral Philosophy. Cambridge University Press, 1996 (com Michael S. McPherson).

Economic Methodology: Crossing Disciplinary Boundaries. London: Macmillan, 1998 (com Roger Backhouse, Uskali Mäki, e Andrea Salanti).

Causal Asymmetries. New York: Cambridge University Press, 1998.

Economic Analysis, Moral Philosophy, and Public Policy. Cambridge University Press, 2006 (com Michael S. McPherson).

Artigos:

"How to do Philosophy of Economics," in P. Asquith and R. Giere, eds., *PSA 1980.* East Lansing: Philosophy of Science Association, 1980, pp. 352-62.

"Are General Equilibrium Theories Explanatory?" in J. Pitt, ed., *Philosophy in Economics.* Dordrecht: Reidel, 1981, pp.17-32.

"John Stuart Mill's Philosophy of Economics," *Philosophy of Science*, 48 (1981): 363-85.

"Constructive Empiricism Contested," *Pacific Philosophical Quarterly*, 63 (1982): 21-28.

"Causal and Explanatory Asymmetry," in P. Asquith and T. Nickles, eds., *PSA 1982*, vol. 1. East Lansing: Philosophy of Science Association, 1982, pp. 43-54.

"The Limits of Economic Science," in N. Rescher, ed., *The Limits of Lawfulness.* Pittsburgh: Center for Philosophy of Science, University of Pittsburgh, 1983, pp. 93-100.

"Defending Microeconomic Theory," *Philosophical Forum*, 15 (1984): 392-404.

"Classical Wage Theory and the Causal Complications of Explaining Distribution," in J. Pitt, ed., *Change and Progress in Modern Science*, Dordrecht: Reidel, 1985, pp. 171-97.

"Is Falsificationism Unpracticed or Unpracticable?" *Philosophy of the Social Sciences* 15 (1985): 313-19.

"Liability, Responsibility and Harm." *Ethics* 97 (1986): 262-69.

"Health Care: Efficiency and Equity," in S. Spicker, S. Ingman and I. (eds.) *Ethical Dimensions of Geriatric Care: Value Conflicts for the 21st Century*. Dordrecht: Reidel, 1987, pp. 67-78.

"Economic Methodology and Philosophy of Science," in R. Teichgraeber e G. Winston, eds., *The Boundaries of Economics*. Cambridge: Cambridge University Press, 1987, pp. 88-116.

"An Appraisal of Popperian Economic Methodology," in N. de Marchi, ed., *The Popperian Legacy in Economics*. Cambridge: Cambridge University Press, 1988, pp. 65-86.

"Economic Methodology in a Nutshell," *Journal of Economic Perspectives* 3(1989): 115-27; rpt. in B. Caldwell, ed. *The Philosophy and Methodology of Economics*. Cheltenham: Edward Elgar, 1993, vol. 1, pp. 275-87

"The Insufficiency of Nomological Explanation," *Philosophical Quarterly* 39 (1989): 22-35.

"Explanatory Progress in Economics," *Social Research* 56 (1989): 361-81; rpt. in B. Caldwell, ed. *The Philosophy and Methodology of Economics*. Cheltenham: Edward Elgar, 1993, vol. 3, pp. 460-80.

"Ceteris Paribus Clauses and Causality in Economics," in A. Fine e J. Leplin, eds. *PSA 1988*, vol. 2. East Lansing: Philosophy of Science Association, 1989, pp. 308-17.

"Are Markets Morally Free Zones?" *Philosophy and Public Affairs* 18 (1989): 317-33.

"Decision Theory and the Deductive Method," *Recherche Economique* 43 (1989): 199-217.

"What Are General Equilibrium Theories?" in W. Sieg, ed., *Acting and Reflecting*. Dordrecht: Kluwer, 1990, pp. 107-114.

"Making Interpersonal Comparisons Coherently," *Economics and Philosophy* (com Martin Barrett) 6 (1990): 293-300.

"On Dogmatism in Economics: The Case of Preference Reversals," *Journal of Socio-Economics* 20 (1991): 205-25.

"Thresholds, Transitivity, Overdetermination, and Events," *Analysis* 52(1992): 159-63.

"On the Conceptual Structure of Neoclassical Economics - A Philosopher's View," in *Essays on Philosophy and Economic Methodology* (1992), pp. 25-32.

"Why Don't Effects Explain Their Causes?" *Synthese* 94 (1993), pp. 227-44.

"Liberalism, Welfare Economics, and Freedom," *Social Philosophy and Policy* 10(1993): 172-97; também em E. Paul, F. Miller, and J. Paul, eds. *Liberalism and the Economic Order*. Cambridge: Cambridge University Press, 1992, pp. 172-97.

"Taking Ethics Seriously: Economics and Contemporary Moral Philosophy" (com Michael McPherson) in *The Journal of Economic Literature*. 31 (June, 1993), pp. 670-731; reimpresso em A. Hamlin, ed. *Ethics and Economics*. Cheltenham: Edward Elgar, 1996.

"Kuhn, Lakatos and the Structure of Economics," in R. Backhouse, ed. *Contemporary Issues in Economic Methodology*. London: Routledge, 1994, pp. 195-215.

"The Impossibility of Interpersonal Utility Comparisons." *Mind* 104 (1995): 473-90.

"The Composition of Economic Causes," *The Monist* 78 (1995): 295-307.

"Why Does Evidence Matter So Little to Economic Theory?" in *Structures and Norms in Science*. ed. M.L. Dalla Chiara, K. Doets, D. Mundici, e J. van Benthem. Dordrecht: Kluwer, 1997, pp. 395-407.

"Problems with Supply-Side Egalitarianism," *Politics and Society* 24 (1996): 343-51; Rpt. in Erik Wright, ed. *Recasting Egalitarianism*. London: Verso, 1998, pp. 75-85.

"Economics as Separate and Inexact," *Economics and Philosophy* 12 (1996): 207-20.

"Rationalité, Bien-être, et Economie Normative," tr. J.S. Lenfant, in Hubert Brochier, Roger Frydman, Bernard Gazier, & Jérôme Lallement, eds. *L'Economie Normative*. Paris: Economica, 1997, pp. 219-29.

"Rationality and Knavery," in Werner Leinfellner and Eckehart Köhler, eds. *Game Theory, Experience, Rationality; Foundations of Social Sciences; Economics and Ethics: In Honor of John C. Harsanyi*. Dordrecht: Kluwer, 1998, pp. 67-79.

"Liberty and its Value," in J. F. Laslier, N. Gravel, A Trannoy & M. Fleurbaey, eds. *Liberty*. Routledge, 1997, pp. 45-48.

"Economists' Responses to Anomalies: Full Cost Pricing versus Preference Reversals," in John Davis, ed. *New Economics and its History*, NC: Duke University Press, 1988, pp. 255-72 (com Philippe Mongin).

"Problems with Realism in Economics," *Economics and Philosophy* 14 (1998): 185-213.

"Ontology and Methodology in Economics," *Economics and Philosophy* 15 (1999): 283-88.

"Realist Philosophy and Methodology of Economics: What Is It?" *Journal of Economic Methodology* 7 (2000): 127-33.

"Revealed Preference, Belief, and Game Theory." *Economics and Philosophy* 16(2000): 99-115.

"The Basic Problems of Economic Methodology," *Économies et Sociétés* 7 (2000): 147-55.

"In Defense of Two Policy Platitudes," In *Macroeconomics and the Real World*, vol. 2, *Keynesian Economics, Unemployment, and Policy*. Ed. Roger Backhouse e Andrea Salanti. Oxford: Oxford University Press, 2000, pp. 269-76.

"Does Economics Need Laws?" publicado como "¿Necesita Leyes la Economía?" *Argumentos de Razón Técnica* 3 (2000): 115-37.

"Critical Realism and Theories of Open Systems" publicado como "El Realismo Crítico y las Teorías de Sistemas Abiertos." *Argumentos de Razón Técnica* 3 (2000): 61-92.

"A New Era for Economic Methodology," *Journal of Economic Methodology* 8 (2001): 65-68.

"Tendencies, Laws and the Composition of Economic Causes," in Uskali Mäki, ed. *The Economic World View. Studies in the Ontology of Economics*. Cambridge: Cambridge University Press, 2001, pp. 293-307.

"Explanation and Diagnosis in Economics," *Revue Internationale De Philosophie* 55 (2001): 311-26.

"The Limits to Empirical Ethics," in Christopher Murray, Joshua Salomon, Colin Mathers, e Alan Lopez, eds. *Summary Measures of Population Health: Concepts, Ethics, Measurements and Applications*. Geneva: World Health Organization, 2002, pp. 663-68.

"Rational Belief and Social Interaction." *Behavior and Brain Sciences* 26 (2003): 163-64.

"Philosophical Foundations of Normative Economics," in Melvin Ayogu e Don Ross, eds. *Development Dilemmas: The Methods and Political Ethics of Growth Policy*. New York: Routledge, 2005, pp. 40-61.

"Filosofia dell'economia," in *Filosofia delle scienze*, ed. Nicla Vassallo. Turin: Giulio Einaudi, 2003, pp. 71-97.

"Testing Game Theory." *Journal of Economic Methodology* 12(2005): 211-23.

"Valuing Health," *Philosophy and Public Affairs* 34 (2006): 246-74.

"The Philosophical Foundations of Mainstream Normative Economics," *The Philosophy of Economics: An Anthology* (3^a ed.) (com Michael McPherson). Cambridge: Cambridge University Press, 2007.

"Protecting Groups from Genetic Research." *Bioethics* 22 (2008): 157-65.

"Philosophy of the Social Sciences" in Diego Rios and Christoph Schmidt-Petri, eds. *Philosophy of the Social Sciences: Five Questions*. Roskilde, Denmark: Automatic Press, 2008, pp. 57-67.

"Market Failure, Government Failure, and the Hard Problems of Cooperation." *Éthique et économique/Ethics and Economics*, 6 (1), 2008. <http://ethique-economique.net/>

"Experimenting on Markets and the World," *Journal of Economic Methodology* 15 (2009): 209-16.

"Preference Satisfaction and Welfare Economics," *Economics and Philosophy* 25 (2009): 1-25 (com Michael McPherson).

"Equality of Autonomy," *Ethics* 119 (2009): 742-56.

"To Nudge or Not to Nudge," *Journal of Political Philosophy* 18 (2010): 123-36 (com Brynn Welch)

"The Philosophy of Economics," in Fritz Allhoff, ed. *Philosophies of the Sciences*. West Sussex: John Wiley & Sons, 2010, pp. 324-55.

Entradas em enciclopédias e dicionários:

"Economics and Ethics," *Routledge Encyclopedia of Philosophy*. ed. Edward Craig. London: Routledge, 1998, vol. 3, pp. 205-11 (com Michael S. McPherson).

"Economics, Philosophy of," *Routledge Encyclopedia of Philosophy*. ed. Edward Craig. London: Routledge, 1998, vol. 3, pp. 211-22.

"Philosophy of Economics," *Stanford Encyclopedia of Philosophy*; on line <http://plato.stanford.edu/>

"Falsification," *New Palgrave Dictionary of Economics*.

"Theory Appraisal," *New Palgrave Dictionary of Economics* (com Ellery Eells).

"Mindless or Mindful Economics: A Methodological Evaluation," in Andrew Caplin and Andrew Schotter, eds. *Handbook of Economic Methodology*. Oxford University Press, 2008, pp. 125-51.

"Laws, Causation, and Economic Methodology," in Harold Kincaid e Don Ross, eds., *The Oxford Handbook of Philosophy of Economics*. Oxford: Oxford University Press, 2009, pp. 35-54.

"Philosophy of Economics," *International Encyclopedia of the Social and Behavioral Sciences*. Elsevier, 2001, pp. 4159-4165.

No prelo:

"Why Not Just Ask? Preferences, "Empirical Ethics" and the Role of Ethical Reflection," in *From Fairness to Goodness*, ed. Dan Wikler, World Health Organization.

"Probabilistic Causation and Practical Causal Generalizations," In Ellery Eells and James Fetzer, eds. *The Place of Probability in Science*.

"Ethics and Economics." *Encyclopedia of Philosophy*, 2nd ed. Routledge.

"Social Scientific Naturalism and Experimentation in Economics," In Uskali Mäki, ed. *The Handbook of the Philosophy of Economics*. North-Holland, Elsevier, 2006.

"Morgenbesser's Schemata," in Sanford Goldberg, ed. *Sidney Morgenbesser's Contributions*.

"Egalitarianism Reconsidered," *Journal of Moral Philosophy* (co-authored with Matt Waldren).

Joshua Knobe – Referências bibliográficas selecionadas

Livros-coordenador:

J. Knobe & S. Nichols, *Experimental Philosophy*, New York: Oxford University Press, 2008.

Artigos:

J. Knobe, "Answers to Five Questions". In Aguilar, J & Buckareff, A (eds.) *Philosophy of Action: 5 Questions*, London: Automatic Press, 2009.

J. Knobe & S. D. Kelly, "Can one act for a reason without acting intentionally?" In C. Sandis (ed.), *New Essays on the Explanation of Action*, Basingstoke: Palgrave Macmillan, 2009, pp. 169-183.

S. Nichols & J. Knobe, "Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions". *Nous* 41 (2007), pp. 663-685.

J. Knobe, "The Concept of Intentional Action: A Case Study in the Uses of Folk Psychology", *Philosophical Studies* 130 (2006), pp. 203-231.

J. Knobe, "Intentional Action and Side Effects in Ordinary Language", *Analysis* 63 (2003), pp. 190-193.

J. Knobe, "Intentional Action in Folk Psychology: An Experimental Investigation". *Philosophical Psychology* 16 (2003), pp. 309-324.

C. Hitchcock & J. Knobe, (no prelo), "Cause and Norm". *Journal of Philosophy*.

John Broome – Referências bibliográficas selecionadas

Livros-autor:

The Microeconomics of Capitalism, Academic Press, 1983.

Weighing Goods: Equality Uncertainty and Time. Blackwell, 1991. Paperback 1995.

Counting the Cost of Global Warming, White Horse Press, 1992.

Ethics Out of Economics, Cambridge, Cambridge University Press, 1999.

Weighing Lives, Oxford University Press, 2004.

Artigos selecionados acerca de normatividade e raciocínio prático:

“Normative requirements” *Ratio*, 12 (1999), pp. 398–419. Reimpresso em Jonathan Dancy, ed., *Normativity*, Blackwell, 2000, pp. 78–99.

“Normative practical reasoning”, *Proceedings of the Aristotelian Society*, Supplementary Volume 75 (2001), pp. 175–93.

“Are intentions reasons? And how should we cope with incommensurable values?”, in Christopher Morris e Arthur Ripstein, eds., *Practical Rationality and Preference: Essays for David Gauthier*, Cambridge University Press, 2001, pp. 98–120.

“Practical reasoning” in by José Bermúdez and Alan Millar, eds., *Reason and Nature: Essays in the Theory of Rationality*, Oxford University Press, 2002, pp. 85–111.

‘Reasons’ in Jay Wallace, Michael Smith, Samuel Scheffler and Philip Pettit, eds., *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, Oxford University Press, 2004. pp. 28–55.

‘Does rationality give us reasons?’, *Philosophical Issues*, 15 (2005), pp. 321–37.

‘Have we reason to do as rationality requires? A comment on Raz’ *Journal of Ethics and Social Philosophy*, 1 (2005).

‘Reasoning with preferences’ in Serena Olsaretti (ed.), *Preferences and Well-Being*, Cambridge University Press, 2006, pp. 183–208.

‘Wide or narrow scope?’, *Mind*, 116 (2007), pp. 359–70.

‘Does rationality consist in responding correctly to reasons?’, *Journal of Moral Philosophy*, 4 (2007), pp. 349–74.

‘Is rationality normative?’, *Disputatio*, 11 (2008), pp. 153–71.

‘The unity of reasoning?’ in S. Robertson (ed.), *Spheres of Reason*, Oxford University Press, 2009, pp. 62–92.

‘Motivation’, *Theoria*, 75 (2009), pp. 79–99.

‘Requirements’, in Toni Rønnow-Rasmussen, Björn Petersson, Jonas Josefsson and Dan Egonsson, eds., *Homage à Wlodek: Philosophical Papers Dedicated to Wlodek Rabinowicz*.

Alguns artigos sobre igualdade:

‘What’s the good of equality?’, in John Hey, ed., *Current Issues in Microeconomics*, Macmillan, 1989, pp. 236–62.

‘Respects and levelling down’ presented at the European Congress of Analytical Philosophy, Lund, 2002.

‘Equality versus priority: a useful distinction’, in Daniel Wikler et al., eds., *Fairness and Goodness in Health*, World Health Organization, forthcoming.

Escritos sobre alterações climáticas:

[*Counting the Cost of Global Warming*](#), White Horse Press, 1992.

‘[The ethics of climate change](#)’, *Scientific American*, June 2008, reimpresso em Tim Folger e Elizabeth Kolbert, eds., *The Best American Science and Nature Writing*, 2008, Houghton Mifflin, 2009, pp. 11–18.

‘[The most important thing about climate change](#)’, in Jonathan Boston, Andrew

Bradstock e David Eng, eds., *Public Policy: Why Ethics Matters*, ANU E Press, 2010, pp. 101–116.

Outros:

‘[Representing an ordering when the population varies](#)’, *Social Choice and Welfare*, 20 (2003), pp. 243–6.

‘[Should we value population?](#)’, *Journal of Political Philosophy*, 13 (2005), pp. 399–413.

‘[Can there be a preference-based utilitarianism?](#)’, in Marc Fleurbaey, Maurice Salles & John Weymark, eds., *Justice, Political Liberalism and Utilitarianism: Themes from Harsanyi and Rawls*, Cambridge University Press, 2007.

‘[What is your life worth?](#)’, *Daedalus*, 137 (2008), pp. 49–56.

‘[Why economics needs ethical theory](#)’, in Kaushik Basu e Ravi Kanbur, eds., *Welfare, Development, Philosophy and Social Science: Essays for Amartya Sen’s 75th Birthday. Volume 3*, Oxford University Press, forthcoming.

Artigos para a Organização Mundial de Saúde:

‘[Measuring the burden of disease by aggregating wellbeing](#)’ e ‘[Fairness, goodness and levelling down](#)’, ‘[All goods are relevant](#)’, ambos in Christopher J.L. Murray, Joshua A. Salomon, Colin D. Mathers & Alan D. Lopez, eds., *Summary Measures of Population Health: Concepts, Ethics, Measurement and Applications*, World Health Organization, 2002, pp. 91–113, 135–7, e 727–9.

‘[Measuring the burden of disease](#)’, ‘[Equality versus priority: a useful distinction](#)’ e ‘[A comment on Temkin’s trade-offs](#)’, in Daniel Wikler et al., eds., *Fairness and Goodness in Health*, World Health Organization, forthcoming.

Michael Smith – Referências bibliográficas selecionadas

Livros – autor:

The Moral Problem, Oxford: Blackwell, 1994.

Mind, Morality and Explanation: Selected Collaborations, Oxford: Clarendon Press, 2004 (com Frank Jackson e Philip Pettit).

Ethics and the A Priori: Selected Essays on Moral Psychology and Meta-Ethics, New York: Cambridge University Press, 2004.

Livros - coordenador:

Meta-Ethics, Aldershot: Dartmouth Publishing, 1995.

Reason and Value: Themes from the Moral Philosophy of Joseph Raz, Oxford: Oxford University Press, 2004 (com R.Jay Wallace, Philip Pettit, e Samuel Scheffler).

Oxford Handbook of Contemporary Philosophy, Oxford: Oxford University Press, 2005 (com Frank Jackson).

Common Minds: Themes from the Philosophy of Philip Pettit, Oxford: Oxford University Press, 2007 (com Geoffrey Brennan, Robert Goodin, e Frank Jackson).

Artigos:

"Actions, Attempts and Internal Events", *Analysis*, 43, 1983, pp.142-146.

"Should We Believe in Emotivism?" in Graham Macdonald e Crispin Wright (eds.), *Fact, Science and Morality: Essays on A.J.Ayer's Language, Truth and Logic*, Oxford: Blackwell, 1986, pp.289-310.

"The Humean Theory of Motivation", *Mind*, 96, 1987, pp.36-61.

"Reason and Desire", *Proceedings of the Aristotelian Society*, 88, 1987-8, pp.243-256.

"On Humeans, Anti-Humeans and Motivation: A Reply to Pettit", *Mind*, 97, 1988, pp.589-595.

"Dispositional Theories of Value", *Supplement to the Proceedings of the Aristotelian Society*, 63, 1989, pp.89-111.

"Valuing: Desiring or Believing?" in David Charles e Kathleen Lennon (eds.), *Reduction, Explanation and Realism*, Oxford: Oxford University Press, 1992, pp.323-360.

"Practical Unreason", *Mind*, 102, 1993, pp.53-79 (com Philip Pettit).

"Objectivity and Moral Realism: On the Significance of the Phenomenology of Moral Experience" in John Haldane e Crispin Wright (eds.), *Reality, Representation and Projection*, Oxford: Oxford University Press, 1993, pp.235-255.

"Colour, Transparency, Mind-Independence" in John Haldane e Crispin Wright (eds.), *Reality, Representation and Projection*, Oxford: Oxford University Press, 1993, pp.269-277.

"Brandt on Self-Control" in Brad Hooker (ed.), *Rationality, Rules and Utility: New Essays on the Moral Philosophy of Richard B.Brandt*, Boulder: Westview Press, 1993, pp.33-50 (com Philip Pettit).

"Why Expressivists About Value Should Love Minimalism About Truth", *Analysis*, 54, 1994, pp.1-12.

"Minimalism and Truth-Aptness", *Mind*, 103, 1994, pp.287-302 (com Frank Jackson e Graham Oppy).

"Philosophy and Commonsense: The Case of Weakness of Will" in Michaelis Michael e John O'Leary-Hawthorne (eds.), *Philosophy in Mind: The Place of Philosophy in the Study of Mind*, Dordrecht: Kluwer Academic Publishers, 1994, pp.141-157 (com Jeanette Kennett).

"Internal Reasons", *Philosophy and Phenomenological Research*, 55, 1995, pp.109-131.

"Internalism's Wheel", *Ratio*, 8, 1995, pp.277-302.

"Freedom in Belief and Desire", *Journal of Philosophy*, 93, 1996, pp.429-449 (com Philip Pettit).

"In Defence of *The Moral Problem*: A Reply to Brink, Copp and Sayre-McCord", *Ethics*, Symposium on Michael Smith's *The Moral Problem*, 108, 1997, pp.84-119.

"Parfit's P" in J. Dancy (ed.), *Reading Parfit*, Oxford: Blackwell, 1997, pp.71-95 (com Philip Pettit).

"A Theory of Freedom and Responsibility" in Garrett Cullity e Berys Gaut (eds.), *Ethics and Practical Reason*, Oxford: Oxford University Press, 1997, pp.293-319.

"Global Response-Dependence and Noumenal Realism", *The Monist*:_Special Issue on Secondary Qualities Generalized, Peter Menzies ed., 81, 1998, pp. 85-111 (com Daniel Stoljar).

"Response-Dependence Without Reduction", *European Review of Philosophy*, Special Issue on Response-Dependence, Roberto Casati e Christine Tappolet eds., 3, 1998, pp.85-108.

"Ethics and the A Priori: A Modern Parable", *Philosophical Studies*, Special Issue on the A Priori John Hawthorne ed., 92, 1998, pp. 149-174.

"The Possibility of Philosophy of Action" in Jan Bransen e Stefaan Cuypers (eds.), *Human Action, Deliberation and Causation*, Dordrecht: Kluwer Academic Publishers, 1998, pp.17-41.

"The Definition of 'Moral'" in Dale Jamieson (eds.), *Singer and His Critics*, Oxford: Blackwell, 1999, pp.38-63.

"Does the Evaluative Supervene on the Natural?" in Roger Crisp e Brad Hooker (eds.), *Well-Being and Morality: Essays in Honour of James Griffin*, Oxford: Oxford University Press, 2000, pp.91-114.

"Global Consequentialism" in Brad Hooker, Elinor Mason & Dale E. Miller (eds.), *Morality, Rules, and Consequences: A Critical Reader*, Edinburgh: Edinburgh University Press, 2000, pp. 121-133 (com Philip Pettit).

"Ethical Particularism and Patterns" in Brad Hooker e Margaret Little (eds.), *Moral Particularism*, Oxford: Oxford University Press 2000, pp.79-99 (com Frank Jackson and Philip Pettit).

"Quelques énigmes concernant le contrôle de soi", *Philosophiques* 27, 2000, pp 287-304.

"Normative Skepticism", *Dialogue*, 20, 2001 pp.25-30.

"Some Not-Much-Discussed Problems for Non-Cognitivism in Ethics", *Ratio*, 14, 2001, pp.93-115.

"Irresistible Impulse" in Ngaire Naffine, Rosemary Owens e John Williams (eds.), *Intention in Law and Philosophy*, Aldershot: Ashgate, 2001, pp.37-56.

"Immodest Consequentialism and Character", *Utilitas*, Special Issue on Consequentialism and Character, Julia Driver ed., 13, 2001 pp. 173-194.

"The Resentment Argument" in Dan Egonsson, Jonas Josefsson, Björn Petersson, Toni Rønnow-Rasmussen (eds.), *Exploring Practical Philosophy*, Aldershot: Ashgate, 2001, pp.109-122.

"Which Passions Rule?", *Philosophy and Phenomenological Research*, 65, 2002, pp.157-63.

"Exploring the Implications of the Dispositional Theory of Value," *Philosophical Issues: Realism and Relativism*, 12, 2002, pp.329-347.

"Bernard Gert's Complex Hybrid Conception of Rationality" in Walter Sinnott-Armstrong e Robert Audi (eds.), *Rationality, Rules, and Ideals: Critical Essays on Bernard Gert's Moral Theory*, Lanham, Maryland: Rowman and Littlefield, 2002, pp.109-123.

"Evaluation, Uncertainty and Motivation", *Ethical Theory and Moral Practice*, 5, 2002, pp.305-320.

"Is There a Lockean Argument Against Expressivism?", *Analysis*, 63, 2003, pp.76-86 (com Daniel Stoljar).

"Rational Capacities" in Stroud e Christine Tappolet (eds.), *Weakness of Will and Varieties of Practical Irrationality*, Oxford: Oxford University Press, 2003, pp.17-38.

"Neutral and Relative Value after Moore", *Ethics*, Centenary Symposium on G.E.Moore's *Principia Ethica*, 113, 2003, pp.576-598.

"Humeanism, Psychologism, and the Normative Story", *Philosophy and Phenomenological Research*, 67, 2003, pp.460-467.

"The Truth in Deontology" in R.Jay Wallace, Philip Pettit, Samuel Scheffler & Michael Smith (eds.), *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, Oxford: Oxford University Press, 2004, pp.153-175 (com Philip Pettit).

"Instrumental Desires, Instrumental Rationality", *Supplement to the Proceedings of the Aristotelian Society*, 78, 2004, pp.93-109.

"The Structure of Orthonomy" in John Hyman e Helen Steward (eds.), *Action and Agency* (Royal Institute of Philosophy Supplement: 55), Cambridge: Cambridge University Press, 2004, pp.165-193.

"Is That All There Is?", *The Journal of Ethics*, 10, 2006, Special Issue on Joel Feinberg, pp.75-106.

"Moore on the Right, the Good, and Uncertainty" in Terence Horgan e Mark Timmons (eds.), *Metaethics After Moore*, Oxford: Oxford University Press, 2006, pp.133-148.

"External Reasons" in Cynthia Macdonald e Graham Macdonald (eds.), *McDowell and His Critics*, Oxford: Blackwell, 2006 (com Philip Pettit) pp.140-168.

"Is there a Nexus between Reasons and Rationality?" in Sergio Tenenbaum (ed.), *Moral Psychology*, Amsterdam: Rodopi, 2007, pp.279-298.

"The Truth About Internalism" in Walter Sinnott-Armstrong (ed.), *Moral Psychology Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, New York: Oxford University Press, 2008, pp.207-215.

"Consequentialism and the Nearest and Dearest Objection" in Ian Ravenscroft (ed.), *Minds, Ethics, and Conditionals: Themes from the Philosophy of Frank Jackson*, Oxford: Oxford University Press, 2009, pp.237-266.

"The Explanatory Role of Being Rational" in David Sobel e Steven Wall (eds.), *Reasons for Action*, New York: Cambridge University Press, 2009, pp.58-80.

"Desires, Values, Reasons, and the Dualism of Practical Reason", *Ratio: Special Issue: Parfit's On What Matters*, John Cottingham e Jussi Suikkanen eds., 22, 2009, pp.98-125.

"Reasons With Rationalism After All", *Analysis Reviews*, 69, 2009, pp.1-10.

"Two Kinds of Consequentialism", *Philosophical Issues*, 19, 2009, pp. 257-272.

"Beyond the Error Theory" in Richard Joyce e Simon Kirchin (eds.), *A World Without Values: Essays on John Mackie's Moral Error Theory*, New York: Springer, 2010, pp.119-139.

Entradas em enciclopédias e dicionários:

"Realism" in P. Singer (ed.), *Companion to Ethics*, Oxford: Blackwell, 1991, pp. 399-410.

"Le Descriptivisme et le Naturalisme Moral" in Monique Canto Sperber (ed.), *Dictionnaire de Philosophie Morale*, Paris: Presses Universitaires de France, 1997, pp.390-397.

"Reasons and Causes" in E. Craig (ed.), *Routledge Encyclopedia of Philosophy: Volume 8*, London: Routledge, 1998, pp.291-293.

"Emotivism" in E. Craig (ed.), *Routledge Encyclopedia of Philosophy: Volume 3*, London: Routledge, 1998, pp.124-127.

"Direction of Fit" in R. Audi (ed.), *Cambridge Dictionary of Philosophy*, Cambridge: Cambridge University Press, 1999, p.237.

"Moral Rationalism" in R. Audi (ed.), *Cambridge Dictionary of Philosophy*, Cambridge: Cambridge University Press, 1999, pp.587-8.

"Morality and Law" in Christopher B.Gray (ed.), *The Philosophy of Law: An Encyclopedia*, New York: Garland, 1999, pp.567a-570b.

"Moral Realism" in Hugh LaFollette (eds.), *Blackwell Guide to Ethical Theory*, Oxford: Blackwell, 2000, pp.15-37.

"Free Will and Action" in Neil J. Smelser e Paul B. Baltes (eds.), *International Encyclopedia of the Social and Behavioural Sciences*, Amsterdam: Elsevier Science, 2001.

"Metaethics" in Frank Jackson e Michael Smith (eds.), *Oxford Handbook of Contemporary Philosophy*, Oxford: Oxford University Press, 2005, pp.3-30.

"Links Between Philosophy at Princeton University and Australasian Philosophy" forthcoming in *Companion to Philosophy in Australia and New Zealand*, Graham Oppy, Nick Trakakis, Lynda Burns, Steve Gardner, & Fiona Leigh eds., Monash ePress.

"Humeanism about Motivation", in Constantine Sandis e Tim O'Connor (eds.), *Blackwell Companion to the Philosophy of Action* (no prelo).

No prelo:

"Moral Obligation, Accountability, and Second-Personal Reasons", *Philosophy and Phenomenological Research* (com Jada Twedt Strabbing).

"Schiffer's Unhappy Face Solution to a Puzzle about Moral Judgement", in Gary Ostertag (ed.), *Meanings and Other Things: Essays on Stephen Schiffer*, Cambridge: MIT Press.

"The Motivation Argument for Non-Cognitivism" in Charles Pigden (ed.), *Hume, Motivation and Virtue*, Rochester: Rochester University Press.

"The Standard Story of Action: An Exchange", in Jesús H. Aguilar e Andrei A. Buckareff (eds.), *Causing Human Action: New Perspectives on the Causal Theory of Action*, Cambridge: Bradford Book, The MIT Press.

"The Value of Making and Keeping Promises", in Hanoeh Sheinman (ed.), *Promises and Agreements: Philosophical Essays*, New York: Oxford University Press.

"Dworkin on External Skepticism", *Boston Law Review*.

"Four Objections to the Standard Story of Action (and Four Replies)", *Philosophical Issues*.

"On Normativity", *Analysis Reviews*.

Jon Elster – Referências bibliográficas selecionadas¹⁷⁰

Livros-autor:

Leibniz et la Formation de l'Esprit Capitaliste, Paris: Aubier-Montaigne, 1975.

Logic and Society, Chichester: Wiley, 1978.

Ulysses and the Sirens, Cambridge University Press, 1979.

Explaining Technical Change, Cambridge University Press, 1983.

Sour Grapes, Cambridge University Press, 1983.

Making Sense of Marx, Cambridge University Press, 1985.

An Introduction to Karl Marx, Cambridge University Press, 1986.

The Cement of Society, Cambridge University Press, 1989.

¹⁷⁰ O trabalho de Elster está traduzido em muitas línguas, desde o português ao coreano e japonês: aqui referimos apenas a publicação original.

Solomonic Judgments, Cambridge University Press, 1989.

Nuts and Bolts for the Social Sciences, Cambridge University Press, 1989.

Psychologie politique, Editions de Minuit 1990.

Local Justice, New York: Russell Sage Foundation (em co-publicação com Cambridge University Press), 1992.

Alchemies of the Mind : Rationality and the Emotions, Cambridge University Press, 1999.

Strong Feelings: Emotion, Addiction and Human Behavior, MIT Press, 1999.

Ulysses Unbound, Cambridge University Press, 2000.

Closing the Books: Transitional Justice in Historical Perspective, Cambridge University Press, 2004.

Explaining Social Behavior, Cambridge University Press, 2007.

Agir contre soi, Paris: Editions Odile Jacob, 2007.

Le désintéressement, Paris: Seuil, 2009.

The First Social Scientist. A study of Alexis de Tocqueville, Cambridge University Press, 2009.

Livros-coordenador:

J. Elster e A.Hylland, eds., *Foundations of Social Choice Theory*, Cambridge University Press, 1986.

J. Elster, ed., *The Multiple Self*, Cambridge University Press, 1986.

J. Elster, ed., *Rational Choice*, Oxford : Blackwell. 1986.

J. Elster, ed., *Karl Marx: A Reader*, Cambridge University Press, 1986.

J. Elster e R.Slagstad, eds., *Constitutionalism and Democracy*, Cambridge University Press, 1988

J. Elster e K.O.Moene, eds., *Alternatives to Capitalism*, Cambridge University Press, 1989.

J.Elster e J.Roemer, eds., *Interpersonal Comparisons of Well-Being*, Cambridge University Press, 1991.

G.Loewenstein e J.Elster, eds., *Choice over Time*, New York: Russell Sage Foundation, 1992.

J.Elster e N.Herpin, eds., *Ethique des choix médicaux*, Poitiers: Actes Sud, 1992.

J. Elster, ed., *Local Justice in America*, New York: Russell Sage, 1995.

J. Elster, ed., *The Round Table Talks in Eastern Europe*, University of Chicago Press, 1996.

J. Elster, ed., *Deliberative Democracy*, Cambridge University Press, 1998.

J. Elster e O.J. Skog, eds., *Getting Hooked: Rationality and Addiction*, Cambridge University Press, 1999.

J. Elster, ed., *Addiction: Entries and Exits*, New York: Russell Sage, 1999.

J. Elster, ed., *Retribution and Restitution in the Transition to Democracy*, Cambridge University Press, 2006.

J. Elster et al., eds., *Understanding Choice, Explaining Behavior: Essays in Honour of Ole-Jørgen Skog*, Oslo Academic Press, 2006.

Artigos:

"Some conceptual problems in political theory", in B.Barry, ed., *Power and Political Theory*. Chichester: Wiley, pp.245-70.

"The labor theory of value", *Marxist Perspectives* 1(1978), 70-101.

"Risk, uncertainty and nuclear power", *Social Science Information* 18(1978), 371-400.

"Anomalies of rationality", in L.Lévy-Garboua, ed., *Sociological Economics*. London: Sage, 1979, pp.65-85.

"Sour grapes: Utilitarianism and the genesis of wants", in A.Sen e B.Williams, eds., *Utilitarianism and Beyond*, Cambridge University Press, 1982, pp.219-38.

"Marxism, functionalism and game theory", *Theory and Society* 11(1982), 452- 82.

"Rationality", in G.Fløistad, ed., *Contemporary Philosophy: A New Survey*, vol.2, The Hague: Martinus Nijhoff, pp.111-31.

"Exploitation, freedom and justice", *Nomos* 26(1983), 277-304.

"One hundred years of Marxist social science", *London Review of Books* 16 (1983).

"Marx et Leibniz", *Revue Philosophique* 108(1983), 167-77.

"Historical materialism and economic backwardness", in T.Ball e J.Farr, eds., *After Marx*, Cambridge University Press, 1984, pp.36-58.

"Further thoughts on Marxism, functionalism and game theory", in B.Chavance, ed., *Marx en Perspective*, Paris: Editions de l'Ecole des Hautes Etudes en Sciences Sociales, 1985, pp.627-48.

"Sadder but wiser: Rationality and the emotions", *Social Science Information* 24 (1985), 375-406.

"Weakness of will and the free-rider problem", *Economics and Philosophy* 1 (1985), 331-65.

"Rationality, morality and collective action", *Ethics* 96(1985), 136-55.

"The nature and scope of rational-choice explanation" in E.Lepore e B.Maclaughlin, eds., *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell 1985, pp.60-72.

"The possibility of rational politics", *Archives Européennes de Sociologie* 28(1987), 67-103.

"Is there (or should there be) a right to work?", in A.Guttman, ed., *Democracy and the Welfare State*, Princeton University Press 1988, pp.53-78.

"Marx, revolution and rational choice", in M.Taylor, ed., *Rationality and Revolution*, Cambridge University Press 1988, pp.206-28.

"Jon Elster goes to China", *London Review of Books* 27.10.1988.

"Taming chance", in S.McMurrin, ed., *The Tanner Lectures on Human Values*, vol.IX, Utah University Press 1988, 105-80.

"Marxism and individualism", in M.Dascal e O.Gruengard, eds., *Knowledge and Politics*, Boulder: Westview Press, 1989, pp. 189-206.

"Social norms and economic theory", *Journal of Economic Perspectives* 3 (1989), 99-118.

"When rationality fails", in K.Cook e M. Levi (eds.), *Limits of Rationality*, University of Chicago Press 1989.

"Utility, duty and fairness", in H.Carlsson e B.Larsson, eds., *Problems of the Mixed Economy: Cooperation, Efficiency, and Stability*. Amsterdam: North-Holland 1990, pp.23-52.

"Self-interest and altruism", in J.Mansbridge (ed.), *Against Self-interest*, University of

Chicago Press, 1990.

"Merton's functionalism and the unintended consequences of action", in J. Clark, C.Modgil e S.Modgil (eds.), *Robert Merton: Consensus and Controversy* Falmer Press 1990, pp.129-35.

"Envy in social life", in R.Zeckhauser (ed.), *Strategy and Choice*, Cambridge, Mass.: M.I.T.Press 1991, pp. 49-82.

"Patterns of causal analysis in Tocqueville's Democracy in America", *Rationality and Society* 3 (1991), 277-97.

"Conventions, creativity, originality", in M.Hjort (ed.), *Rules and Conventions* , Baltimore: Johns Hopkins University Press 1992.

"Fullness and parsimony", in S.Kemal e I. Gaskell (eds.), *Explanation and Value in the Arts*, Cambridge University Press 1993, pp.146-72.

"Rebuilding the boat in the open sea: Constitution-making in Eastern Europe", *Public Administration* 71 (1993), 169-217.

"Majority rule and individual rights", in S.Hurley e S.Shute (eds.), *On Human Rights*, Basic Books 1993, pp.175-216 + 249-56.

"Egonomics", in *The Great Ideas Today*, Chicago: Encyclopedia Britannica 1993, pp.87-135.

"Rationality, emotions, and social norms", *Synthese*. 98 (1994), 21-49.

"Some unresolved problems in the theory of rational choice", *Acta Sociologica*. 36 (1993), 179-190.

"Justice and the allocation of scarce resources", in B.Mellers e J.Baron (eds.), *Psychological Perspectives on Justice*, Cambridge University Press 1993, pp. 259-78.

"Why things don't happen as planned", in N.Åkerman (ed.), *The Necessity of Friction*, Heidelberg: Physica Verlag 1993, 248-56.

"Strategic uses of argument", in K.Arrow et al. (eds.), *Barriers to the Negotiated Resolution of Conflict*, New York: Norton 1995, pp.236-57.

"The idea of equality revisited", in *World, Mind, and Ethics*, a *Festschrift* for Bernard Williams edited by J.Altham and R.Harrison, Cambridge University Press 1995, pp.4-18.

"The psychology of *Democracy in America* ", in *The Great Ideas Today*, Chicago: Encyclopedia Britannica 1994, pp.87-126.

"The empirical study of justice", in D. Miller e M. Walzer (eds.), *Pluralism, Justice and Equality*, Oxford University Press, 1995.

“The impact of rights on economic performance”, in A. Sajó (ed.), *Western Rights? Post-Communist Applications*, The Hague: Kluwer 1996, pp.347-60.

“Rationality and the emotions”, *Economic Journal*, 106 (1996), 1386-97.

“Montaigne’s psychology”, *The Great Ideas Today* 1996, pp. 108-55.

“Ways of constitution-making”, in A. Hadenius (ed.), *Triumph and Crisis of Democracy*, Cambridge University Press 1997, 123-42.

“Afterword: The making of postcommunist presidencies”, in R. Taras (ed.), *The Post-Communist Presidents*, Cambridge University Press 1997, pp.225-37.

“A plea for mechanisms”, in P. Hedström e R. Swedberg (eds.), *Social Mechanisms* , Cambridge University Press 1998, pp.45-73.

“Aristotle on the emotions”, *The Great Ideas Today* Chicago: Encyclopedia Britannica 1998, pp.230-70.

“Rationality and addiction”, in P. De Greiff (ed.), *Drugs and the Limits of Liberalism*, Cornell University Press 1999, pp.25-45.

“Accountability in Athenian politics”, in B. Manin, A. Przeworski e S. Stokes (eds.), *Democracy, Accountability and Representation*, Cambridge University Press 1999, 253-78.

“Davidson on weakness of will and self-deception”, in L. Hahn (ed.), *The Philosophy of Donald Davidson*, Chicago: Open Court 1999, pp. 425-42.

“Knowledge and the politics of transition”, in R. Viale (ed.), *Knowledge and Politics*, Physica Verlag 2003.

“Moral dilemmas in transitional justice”, in Monika Betzler (ed.), *Practical Conflict*, Cambridge University Press 2004.

“Emotion and action”, in R. Solomon (ed.), *Thinking about Feeling: Contemporary Philosophers on Emotion*, Oxford University Press, 2004.

“Emotions and rationality”, in A.S.R. Manstead, N.H. Frijda, & A.H. Fischer (Eds.), *Feelings and Emotions: The Amsterdam Symposium*. Cambridge University Press, 2004.

“A case study of transitional justice: Athens in the fifth century B.C“, in L. Meyer (ed.), *Historical Justice*, Baden-Baden: Nomos 2004.

“Costs and constraints in the economy of the mind”, in I. Brocas e J. Carrillo (eds.), *The Psychology of Economic Decisions*. Vol. 2: *Reasons and Consequences*, Oxford University Press 2004.

“Mimicking impartiality”, in Keith Dowding et al. (eds.), *Justice and Democracy*, Cambridge University Press 2004.

“Motivations and beliefs in suicide missions”, in D. Gambetta (ed.), *Making Sense of Suicide Missions*, Oxford University Press 2005.

“Droit et causalité”, in M. Troper et al. (eds.), *Théorie des contraintes juridiques*, Paris Bruylant 2005.

“Drawing a veil over equality: Equality and hypocrisy in the Revolutionary era”, in C. Sypnowich (ed.), *The Egalitarian Conscience*, Oxford University Press 2006.

“Constituent legislatures”, in R. Bauman e T. Kahana (eds), *The Least Examined Branch*, Cambridge University Press 2006.

“Beyond rational self-interest”, in I. Shapiro et al. (eds.), *The Art of the State: Rethinking Political Institutions*, NYU Press 2006.

“Redemption for wrongdoing: The fate of collaborators after 1945”, *Journal of Conflict Resolution* 2006.

“The night of August 4 1789: A study in collective decision making”, *Revue Européenne des sciences sociales* 2007.

“Interpretation and rational choice”, *Rationality and Society* 2009.

“Reciprocity and social norms”, in Mark Fleurbaey, Maurice Salles e John Weymark (eds.), *Essays in Honour of Serge-Christophe Kolm*, forthcoming from Springer.

“The Valmont effect: The warm-glow theory of philanthropy”, in Thomas Pogge, Patricia Illingworth e Leif Wenar (eds.), *The Ethics of Philanthropy*, forthcoming from Oxford University Press.

Entradas em enciclopédias e dicionários:

“Rationality, economy and society”, in S. Turner (ed.), *The Cambridge Companion to Weber*, Cambridge University Press 2000.

“Pascal and decision theory”, *The Cambridge Companion to Pascal*, Cambridge University Press 2003.

“Tocqueville on 1789: preconditions, precipitants, and triggers”, *The Cambridge Companion to Tocqueville* 2006.

“Altruistic behavior and altruistic motivations”, In S.-C. Kolm e J.Mercier-Ythier (eds.), *Handbook on the Economics of Giving, Reciprocity and Altruism*, Elsevier 2006.

“Emotions”, in *The Oxford Handbook of Analytical Sociology*, Oxford University Press 2009.

“Norms”, in *The Oxford Handbook of Analytical Sociology*, Oxford University Press 2009.

No prelo:

“Bad timing”, in Crisoula Andreou e Mark White (eds.) *The Thief of Time*, Oxford University Press.

“Reciprocity in Seneca and Adam Smith”, *The Adam Smith Review*.

“Emotional choice and rational choice”, in Peter Goldie (ed.), *The Oxford Handbook of the Philosophy of Emotion*, Oxford University Press.