

A questão da racionalidade, da avaliação do pensamento pelo pensamento, é uma questão antiga e um terreno tradicional da filosofia: é difícil, ou mesmo impossível, nomear um filósofo que não tenha pensado sobre os princípios que guiam a forma como pensamos e agimos. O que é novo é a forma como essa avaliação pode e deve ser feita considerando dados da ciência cognitiva. O que se procura, de qualquer forma, com uma teoria da racionalidade é: (i) uma descrição ou caracterização dos factores em jogo nas ocasiões em que agentes passam de determinadas crenças para outras crenças, adicionam ou eliminam crenças do seu corpo de crenças, ou optam, a partir de um conjunto de crenças e desejos, por um curso de acção por entre várias alternativas; (ii) um conjunto de hipóteses acerca da forma como decidimos entre critérios de correcção quando falamos da justificação ou racionalidade de crenças e acções; (iii) um conjunto de hipóteses acerca das razões por que queremos saber (se de facto queremos) se as nossas crenças são verdadeiras e os nossos raciocínios e acções racionais.

(Do projecto de investigação *Rationality, Belief, Desire*
- *Motivation to action from the viewpoint of the theory of mind*)



RACIONALIDADE

Sofia Miguens

Sofia Miguens

RACIONALIDADE

Campo da Filosofia

Sofia Miguens

Racionalidade

RACIONALIDADE

Autora: Sofia Miguens

Capa: Margarida Baldaia

© CAMPO DAS LETRAS – Editores, S.A., 2004
Rua D. Manuel II, 33 – 5.º 4050-345 Porto
Telef.: 22 6080870 Fax: 22 6080880
E-mail: campo.lettras@mail.telepac.pt
Site: www.campo-lettras.pt

Impressão: Tipografia do Carvalhido – Porto

1.ª edição: Dezembro de 2004

Depósito legal n.º 219340/04

ISBN: 972-610-858-6

Código de barras: 9789726108580

Colecção: Campo da Filosofia – 22

Este livro foi patrocinado por



GFMC
– Gabinete de Filosofia Moderna e Contemporânea,
da Faculdade de Letras da Universidade do Porto



*We might think of the pure theory of rationality as the theory of what standards are to be adhered to by any kind of being in any kind of world (...) But it seems unlikely that if there are such standards, they will be very contentful or get us very far.

Robert Nozick, *The Nature of Rationality*

Prólogo

Projecto

Este livro teve origem no Projecto de Investigação *Racionalidade, Desejo, Crença – a motivação para a acção do ponto de vista da teoria da mente* (Rationality, Belief, Desire – Motivation to action from the viewpoint of the theory of mind) levado a cabo pelo Gabinete de Filosofia Moderna e Contemporânea do Instituto de Filosofia da FLUP (Unidade 502 da FCT) a partir de 2003. Pareceu-me útil para o bom funcionamento das actividades do Projecto disponibilizar uma descrição acessível do campo de estudos da racionalidade no qual ele se enquadra.

Há uma explicação que se impõe. Desde a sua criação, em 1996, o Gabinete de Filosofia Moderna e Contemporânea da FLUP propõe-se manter como horizonte a comparação entre tradições distintas da filosofia contemporânea, nomeadamente a tradição analítica e a tradição fenomenológica e hermenêutica. Este livro e o Projecto que lhe deu origem estão mais directamente ligados à tradição analítica e à ciência cognitiva. Apesar disso, o Projecto não pretende encerrar-se, e por essa razão tem como subtítulo ‘a motivação para a acção do ponto de vista da teoria da mente’ e não da ‘filosofia da mente’, o que em princípio o limitaria à tradição analítica, e foi desenhado de modo a correr em paralelo a um outro projecto do Gabinete de Filosofia Moderna e Contemporânea, o Projecto intitulado *Subjectividade e Racionalidade – Fenomenologia e Hermenêutica*, dirigido pela Professora Maria

José Cantista. Mas a verdade é que para levar a cabo uma comparação entre o tratamento de problemas no seio de tradições de pensamento em grande medida incomunicantes como o são a filosofia analítica e a tradição fenomenológica e hermenêutica são no mínimo necessários mapas conceptuais, que sirvam como guias. Este livro pretende precisamente constituir um tal guia no que diz respeito aos estudos da racionalidade feitos no âmbito da filosofia analítica e da ciência cognitiva. Os dois Projectos referidos (*Racionalidade, Desejo, Crença – a motivação para a acção do ponto de vista da teoria da mente e Subjectividade e Racionalidade – Fenomenologia e Hermenêutica*) darão proxima-mente origem a dois volumes paralelos de *Ensaio sobre a Racionalidade*, que reunirão escritos de investigadores e colaboradores do Gabinete de Filosofia Moderna e Contemporânea.

Método

Os investigadores envolvidos no Projecto (quer os pertencentes à equipa básica, quer os convidados, quer os consultores) trabalham a partir de diferentes pressupostos e no âmbito de diferentes abordagens disciplinares da cognição e do pensamento. Interessam-nos análises da racionalidade, da crença e do desejo como aspectos do mental a partir de abordagens diversas, científicas e filosóficas. Pretende-se em contrapartida que as actividades do Projecto constituam uma contribuição da filosofia às várias disciplinas da cognição, situando a análise conceptual e normativa da filosofia na continuidade das ciências cognitivas e de forma alguma como uma última palavra – estamos conscientes de que a literatura relevante para os problemas que nos propusemos tratar se estende desde a epistemologia à filosofia moral, passando por investigações psicológicas (psicologia cognitiva, psicologia evolutiva) e questões de teoria da decisão, e está em crescimento constante. Este guia não pretende portanto, obviamente, ser exaustivo, mas apenas fornecer orientações.

Âmbito

Em termos de filosofia – a filosofia não é de forma alguma a única disciplina que importa nos estudos da racionalidade mas pode ajudar-nos a desenhar um terreno – os estudos sobre a racionalidade distribuem-se em geral por tratamentos da **racionalidade teórica** (racionalidade nas crenças e nos raciocínios) e da **racionalidade prática** (racionalidade nas decisões e na acção). Isso significa desde logo que no seio da filosofia os problemas da racionalidade teórica tendem a surgir em áreas epistemológicas e a pôr em jogo as relações entre lógica, filosofia da mente, filosofia do conhecimento e filosofia da ciência, enquanto que os problemas da racionalidade prática tendem a surgir nas áreas da teoria da acção e da filosofia moral e política. Temos assim que para além da diferença de teor entre as investigações filosóficas e as investigações em ciência cognitiva que é necessário considerar, são muitos e muito diversos entre si os domínios da filosofia envolvidos nos estudos da racionalidade.

Bibliografia

Este livro termina com uma compilação bibliográfica, reunindo nomeadamente a literatura de que partimos no Projecto e alguma que fomos encontrando. Não pretendemos mais do que fornecer referências de partida para quem venha a interessar-se pelos assuntos tratados.

Agradecimentos

Agradeço a todas as pessoas envolvidas no Projecto, nomeadamente Paulo Tunhas, Luísa Couto Soares, João Alberto Pinto, Sara Bizarro, Pedro Madeira, Júlio Campos, Pedro Galvão. À Sara Bizarro agradeço especialmente a disponibilização da

revista electrónica *Intelectu* (www.intelectu.com) para a publicação de resultados – o n.º 9 da Revista *Intelectu* foi a primeira publicação do Projecto. Ao Pedro Galvão agradeço a disponibilização da Revista electrónica de filosofia moral e política *Trólei* – o n.º 4 da Revista *Trólei* foi dedicado ao Projecto. Ao Agostinho Frias, e depois ao Bruno Pinheiro, agradeço a possibilidade de manter as páginas e o Arquivo On-Line do Projecto (http://www.letras.up.pt/df/if/gfmc/fil_mente.html). Pela oportunidade de participar nas suas actividades de investigação devo também agradecer a todo o grupo da Universidade de Santiago de Compostela, nomeadamente Juan Vazquez, Luis Villegas, Jose Luis Falguera e Concha Martinez. A possibilidade de cruzar experiências e de colaborar com o Departamento de Lógica e Filosofia Moral da Universidade de Santiago de Compostela foi muito importante para o próprio desenho do presente Projecto (e aqui devo agradecer especialmente a Juan Vazquez). Além do mais, o trabalho agora apresentado e que é afinal um estudo metodológico introdutório às abordagens da racionalidade feitas na intersecção da filosofia e da ciência cognitiva, constitui parte da minha participação no Projecto de Investigação *El cambio conceptual en la ciencia: de los enfoques ontosemánticos a los cognitivistas* (BFF 2003 – 01962), dirigido por Juan Vazquez. Este Projecto de investigação visa questões de alteração conceptual na ciência e respectivas implicações epistemológicas, e tem como um dos seus sub-objectivos o esclarecimento da possível relevância do ‘enfoque cognitivista’, nomeadamente o enfoque cognitivista da racionalidade, para os problemas da filosofia da ciência¹. Agradeço também o apoio dos filósofos do Centro de

¹ O presente livro visa a investigação acerca da natureza da racionalidade, investigação que é, desde os primórdios da disciplina, levada a cabo pela filosofia e na qual hoje várias ciências cognitivas estão também envolvidas. Mas porque é que a investigação acerca da natureza da racionalidade interessaria a filosofia da ciência? Antes de mais, interessa como (i) exemplo de ‘enfoque cognitivista’, gerando esclarecimentos epistemológicos acerca do que constitui tal enfoque cognitivista, depois

Ciência Cognitiva da Universidade de Rutgers, New Jersey, Alvin Goldman, Stephen Stich e Ernie Lepore, e em geral dos filósofos que se disponibilizaram para ser consultores do Projecto, Stephen Stich, Alvin Goldman, Simon Blackburn e Pascal Engel.

Agradeço finalmente, e de forma muito especial, à Professora Doutora Maria José Cantista, Coordenadora do Gabinete de Filosofia Moderna e Contemporânea da FLUP, por todo o apoio e estímulo às actividades deste Projecto.

Porto, Fevereiro de 2004

SOFIA MIGUENS

porque (ii) os processos da racionalidade interessam em filosofia da ciência tanto quanto a filosofia da ciência se preocupa em saber o que torna a ciência possível, i. e., que traços da mente humana e do desenvolvimento cognitivo permitem a investigação científica, o que subjaz ao raciocínio científico, o que o caracteriza. Ora, o que cognitivamente interessa para a ciência são questões tais como a observação, a geração e falsificação de hipóteses, os processos de inferência em geral, o uso de instrumentos para apoiar a percepção, a teoria da mente. Será portanto de esperar que perspectivas desenvolvimentais e evolutivas sobre as perícias cognitivas que virão a estar envolvidas na constituição e desenvolvimento de conceitos científicos interessem o filósofo da ciência. São pelo menos estas as razões pelas quais é razoável esperar que um enfoque cognitivista contribua para a compreensão da mudança conceptual em ciência.

Origens

Versões anteriores do Capítulo 3, Filosofia e racionalidade teórica: em que devemos acreditar? foram publicadas sob o título Em que devemos acreditar? Questões epistemológicas e investigações cognitivas em: *Intelectu* n.º 9, www.intelectu.com, 2003 – *Revista da Faculdade de Letras – Série de Filosofia*, n.º 19, 2004) e *Cadernos de Filosofia*, n.º 15, 2004.

Material do Capítulo 2, Filosofia e racionalidade prática: o que devemos fazer? e do Capítulo 4, Duas teorias filosóficas da racionalidade apareceu inicialmente sob o título Hume e Blackburn, Razão e Paixões – acerca da estrutura da motivação humana em *Intelectu* n.º 7, www.intelectu.com, 2002, e também na *Revista da Faculdade de Letras*, Série de Filosofia, n.º 18, 2002, (sob a forma de recensão a S. Blackburn 1998, *Ruling Passions*).

O artigo A Identidade Pessoal e a Posição Original Rawlsiana, *Revista Portuguesa de Filosofia*, LIX, 1, 2003, pp.139-170, constituiu um primeiro ensaio acerca da questão pessoal da racionalidade e do confronto entre teorias da motivação para a acção.

Introdução
Estudos da racionalidade
e a ideia de uma teoria filosófica da racionalidade.
Questões para uma teoria filosófica
da racionalidade.

– Escrevi! – exclamou Aquiles alegremente, enquanto colocava o lápis no estojo. – E, por fim, chegamos ao final desta corrida ideal! Agora, que aceita A, B, C e D, EVIDENTEMENTE aceita Z.

– Sim? – retorquiu a Tartaruga inocentemente. – Vamos deixar isso bem claro. Aceito A, B, C e D. Suponha que, ainda assim, me recuso a aceitar Z?

– Então a lógica pegá-lo-ia pelo pescoço e forçá-lo-ia a aceitar! – respondeu Aquiles triunfante. – A lógica dir-lhe-ia: «Não tem saída. Agora que aceitou A, B, C e D, TEM que aceitar Z!»; portanto, não tem escolha, não é?

– Qualquer lógica suficientemente boa para me convencer merece ser escrita – disse a Tartaruga. – Portanto, escreva aí no seu caderno, por favor. Chamar-lhe-emos (E). Se A, B, C e D são verdadeiros, Z tem de ser verdadeiro.

Enquanto não tiver aceite isso, evidentemente não terei de aceitar Z.

Lewis Carroll, What the Tortoise said to Achilles, *Mind*, 1895

O objectivo final do Projecto *Racionalidade, Desejo, Crença* – a motivação para a acção do ponto de vista da teoria da mente foi desde o início formulado como consistindo na obtenção

de uma **teoria filosófica da racionalidade**. Esta foi entendida como envolvendo várias coisas: (i) uma *descrição ou caracterização dos factores em jogo nas ocasiões em que agentes passam de determinadas crenças para outras crenças, adicionam ou eliminam crenças do seu corpo de crenças, ou optam, a partir de um conjunto de crenças e desejos, por um curso de acção por entre várias alternativas*, (ii) *um conjunto de hipóteses acerca da forma como decidimos entre critérios de correcção quando falamos da justificação ou racionalidade de crenças e acções*, (iii) *um conjunto de hipóteses acerca das razões por que queremos saber (se de facto queremos) se as nossas crenças são verdadeiras e os nossos raciocínios e acções racionais*.

Nenhum dos termos da formulação deste objectivo é neutro ou pode ser tomado como garantido – pelo contrário, termos e expressões tais como ‘teoria filosófica da racionalidade’, ‘agentes’, ‘critérios de correcção’ ou o próprio facto de se considerar a ‘motivação para a acção’ como referência para o problema da racionalidade (quando de facto interessam também as questões da racionalidade teórica), indicam só por si caminhos que nos interessou explorar, assim como opções de percurso.

Explico, antes de mais, a ideia de uma **teoria filosófica da racionalidade**. Pretendo que ela seja entendida por contraste com investigações empíricas sobre racionalidade, mesmo se em última análise pretende ser contínua com estas investigações e tomá-las sempre como referência. Isto significa que parto do seguinte princípio. Estamos hoje perante aportações importantes e incontornáveis nos estudos da racionalidade, por exemplo, investigações empíricas relativas a raciocínio, decisão e emoções, feitas no âmbito da psicologia cognitiva, da Inteligência Artificial e da neurociência. O facto de essas investigações deverem ser tomadas como referência numa teoria da racionalidade não significa no entanto que elas nos ofereçam, já pronta, uma ‘teoria filosófica da racionalidade’. Antes de mais porque aquilo que fazemos quando falamos de racionalidade é susceptível de várias interpretações (nomeadamente uma interpretação norma-

tiva e uma interpretação descritiva²). É em parte por isso que os resultados das investigações referidas não são suficientes para constituir imediatamente uma teoria filosófica da racionalidade. Além disso, várias tarefas são necessárias a uma teoria da racionalidade. Ao longo deste livro chamar-lhes-ei ‘questões para uma teoria filosófica da racionalidade’. Passo a enumerá-las.

1. *Cânones de racionalidade e critérios de correcção*

De certa forma aquém das investigações empíricas sobre raciocínio, decisão e emoções, feitas no âmbito da psicologia, da Inteligência Artificial e da neurociência, existem teorias que são naturalmente candidatas ao estatuto de Teorias da Racionalidade, no sentido em que poderíamos pensar que essas teorias nos oferecem os próprios princípios da racionalidade. É o caso da lógica, da teoria da decisão e da teoria das probabilidades (pelo menos concebida como ‘lógica das crenças parciais’, na expressão de F.P. Ramsey). Decidi chamar-lhes **cânones de racionalidade**³. Essas teorias seriam Teorias da Racionalidade

² De acordo com uma **interpretação normativa**, um agente racional, se é racional, obedecerá necessariamente a certos princípios de raciocínio e decisão. A pedra de toque de uma **interpretação descritiva** é a observação empírica: a referência dos estudos da racionalidade não serão então princípios mas os comportamentos de raciocínio e decisão que os agentes têm de facto. Esta segunda interpretação desvia a atenção de princípios de correcção declarados a priori para factores que influenciam o comportamento dos agentes, por exemplo, factores psicológicos. É de qualquer forma conveniente estar preparado para o facto de a evidência experimental não estar de acordo com os princípios abstractos da racionalidade – e saber o que isso significa é um problema em aberto.

³ Como é óbvio, é problemático dizer o que é a **Lógica**, ou em que consiste o nosso conhecimento da lógica, mas é possível tentar: a lógica seria uma ciência geral das inferências, que procura explicitar as regras pelas quais estas podem realizar-se, pelas quais coisas pensadas / afirmadas, se seguem de outras coisas pensadas / afirmadas. Poderíamos dizer

na medida em que explicitariam os princípios correctos de inferência e decisão. De acordo com esta perspectiva, os princípios correctos de inferência e decisão são aquilo que são, i. e., são por princípio as regras da correcção, mesmo se nem sempre regem raciocínios e acções de agentes reais. Mesmo que nós,

também, pensando na lógica dedutiva, que a lógica é uma teoria da validade dedutiva de argumentos (diz-se que um argumento é dedutivamente válido quando não é possível que as suas premissas sejam verdadeiras e a conclusão falsa). Um lógico como Jon Barwise não hesita em afirmar que a lógica é a *ciência do raciocínio*, e que aquilo que constitui um bom raciocínio é universal, ao contrário, por exemplo, do que constitui uma frase bem formada numa língua natural (BARWISE 1999: 483). Cf. também PRIEST 2000. Quanto à **Teoria das Probabilidades**, o problema de dizer o que ela é não tem a ver com a matemática das probabilidades mas com a aplicação destas ao mundo. Por 'probabilidade' pode-se entender nomeadamente frequência relativa, grau de crença, regra para acontecimentos aleatórios. A teoria das probabilidades interessa alguém preocupado com racionalidade na medida em que pode ser considerada como uma 'lógica de crenças parciais', medindo expectativas em relação à ocorrência de acontecimentos que não temos possibilidade de saber antecipadamente se ocorrerão de facto. A interpretação da teoria em causa é a interpretação subjectivista (de acordo com uma interpretação frequencista, a probabilidade de um acontecimento é identificável com a frequência da ocorrência deste, de acordo com uma interpretação subjectivista a probabilidade é uma medida da expectativa de um sujeito em relação à ocorrência de um acontecimento). A teoria das probabilidades permite ainda a formulação daquilo que torna determinadas inferências indutivas (inferências nas quais as premissas nos dão algum fundamento, apenas, para sustentar a conclusão) boas. Quanto à **Teoria da Decisão**, ela é uma teoria abstracta da escolha racional, em cujo núcleo está a seguinte ideia: há agentes, estes necessitam deliberar, i. e., a partir de crenças e preferências necessitam de avaliar cursos de acção em termos de consequências, nomeadamente em circunstâncias que não podem ser totalmente controladas e previstas. Um agente racional será o agente que, tendo em consideração a utilidade (a utilidade é uma unidade de 'desejabilidade') dos estados do mundo que podem ser consequência das suas acções e a probabilidade da obtenção desses estados, age de modo a maximizar a utilidade esperada (utilidade esperada = utilidade do resultado \times probabilidade da ocorrência deste).

enquanto agentes, nos afastemos dos cânones nos nossos raciocínios e acções, esses seriam meros erros de performance, teoricamente desprezáveis se o que nos interessa é a natureza da racionalidade, consistindo esta numa 'correcção ideal'.

Considerar que princípios normativos específicos, como os que estão em causa nas teorias referidas, são *os princípios da racionalidade aos quais de algum modo se conformam agentes reais*, é considerar que os princípios da correcção estão dados. Ora, como é que nós sabemos tal coisa? Apontar as teorias referidas como cânones de racionalidade constitui por si uma decisão, uma decisão acerca de ideais epistémicos, que tem que ser ela própria justificada. E o que é justificar? **Justificar** é dar razões a favor de pretensões avançadas, apresentar provas ou evidência, alegar que algo está conforme com determinados princípios, regras, normas. Uma forma de justificar um raciocínio, por exemplo, é mostrar que ele está 'conforme com determinadas regras', lógicas. Ora, o problema aqui é, evidentemente, que as teorias acima mencionadas poderiam ser consideradas como fornecendo elas próprias as 'regras gerais da justificação' para os nossos pensamentos e acções, não havendo mais nenhuma instância a que recorrer. Mas por que razão as consideramos assim, por que razão consideramos determinadas teorias como cânones de racionalidade? O que justifica a nossa decisão de as considerarmos assim? Teremos razão para fazê-lo? Procurar responder a essa questão envolve a explicitação daquilo a que chamei, utilizando um termo do filósofo americano Alvin Goldman⁴, um **critério de correcção**. Um critério de correcção explicita as razões por que consideramos determinados conjuntos de regras, normas, princípios, como cânones da racionalidade. Formular um critério de correcção envolve dizer exactamente, por exemplo, por que razões consideramos que teorias como a lógica, a teoria das probabilidades, a teoria da decisão nos dão os princípios da racionalidade, da nossa racionalidade. Temos assim que uma

⁴ GOLDMAN 1986.

teoria filosófica da racionalidade não pode limitar-se a apontar cânones de racionalidade, para cujos princípios seria remetida qualquer justificação de raciocínios e acções: uma teoria filosófica da racionalidade envolve a formulação dos critérios de correcção que nos fazem considerar determinados princípios como princípios da racionalidade. Este livro na sua totalidade pretende tratar o problema dos critérios de correcção, embora o assunto seja visado mais especificamente no Capítulo 4 *Filosofia e racionalidade teórica – em que devemos acreditar?*, sob a forma de uma análise da natureza da justificação.

2. Teorias normativas e agentes reais

Os estudos da racionalidade, quer teórica, quer prática, põem em causa normas e justificação, e os conceitos de norma e justificação tornam-se complicados de tratar quando estamos a falar de agentes reais, tais como humanos, animais ou mesmo máquinas (por oposição a situações em que estão em causa apenas princípios formais, que à partida podem ser considerados de forma puramente abstracta). Porque é que isto é assim? Porque falar de agentes reais nos obriga a encarar a relação entre os princípios abstractamente formuláveis e a realidade (física, cognitiva) dos agentes (por exemplo, a relação entre princípios lógicos e arquitecturas e processos cognitivos de agentes concretos, submetidos a constrangimentos e limitações definidas⁵). Por outras palavras, os problemas da racionalidade,

⁵ C. Cherniak (CHERNIAK 1986) diz isto muito claramente: ao levarmos em conta questões relativas a cognição, computação, informação, facilmente verificaremos que os sistemas cognitivos a que queremos chamar agentes racionais não são nem poderiam ser 'perfeitamente racionais' (lógicos perfeitos, por exemplo). Duas aportações de que Cherniak se serve em *Minimal Rationality* para sustentar esta sua tese são a teoria da complexidade (um domínio da teoria da computação no qual se avalia a exequibilidade de algoritmos) e os modelos de memória em psicologia cognitiva. Em agentes

que parecem à partida problemas formais de correcção, justificação e validade, dizem também respeito a mentes que lidam com o mundo (pessoas, animais ou outro tipo de agentes), e a

racionais reais, questões como a exequibilidade de algoritmos, a estrutura da memória e a dimensão dos espaços de busca pesam na reordenação da teia das crenças. À luz de tais considerações, por exemplo, a imagem quineana de *Two Dogmas of Empiricism*, de acordo com a qual evidência contraditória poderia levar a um reordenamento de grande dimensão da teia das crenças, revela-se totalmente irrealista. No entanto, a verdade é que queremos continuar a chamar racionais a vidas mentais que não correspondem à imagem ideal do reordenamento da teia das crenças. Desta forma, segundo Cherniak, se consideramos que caracterizar a racionalidade de um agente nos obriga a explicitar a forma como este 'gere' as suas crenças, nem uma teoria do assentimento (de acordo com a qual o agente tem as crenças que assente ter, sem quaisquer obrigações de inferências), nem uma teoria da racionalidade ideal (de acordo com a qual o agente pratica todas as inferências válidas, todas as acções aparentemente apropriadas a partir do seu corpo de crenças/desejos) constituem uma boa descrição da forma como as crenças estão para um agente. Ao contrário de um agente ideal, que navegaria sem constrangimento nos seu corpo de crenças, e o manteria incontaminado, por exemplo, de inconsistências, agentes racionais como os humanos têm problemas de memória e estão constrangidos por organizações específicas desta, de uma forma que Cherniak considera filosoficamente significativa. Por exemplo, a memória humana está estruturada em termos de activação / não activação, armazenamento de curto prazo / armazenamento de longo prazo. Ora, os requisitos de consistência entre conteúdos de memória com estatutos diversos não são os mesmos. Aliás, e porque se trata de espaços de busca muito grandes (pense-se, por exemplo, na totalidade das crenças que cada um de nós tem neste momento), a memória humana tem uma estrutura compartimentalizada, o que isola umas crenças relativamente às outras (um exemplo de Cherniak é o exemplo do homem que usa um fósforo para iluminar o seu tanque de gasolina: ele sabe que a gasolina é inflamável, e sabe que o fósforo tem uma chama, mas sabe essas coisas de forma compartimentalizada, e portanto não se coíbe de usar um fósforo para iluminar o tanque da gasolina, ver se está vazio, vindo a causar uma explosão, na qual morre, mesmo não sendo suicida). A estrutura da memória é assim uma das fontes de explicação da presença de inconsistências no corpo das crenças de agentes, presença essa que não é suficiente para desqualificar agentes como irracionais.

partir desse momento deixamos de poder considerá-los como 'meramente formais'. Assim, colocar a questão da racionalidade relativamente a agentes reais faz recair a questão da correcção e das normas de um âmbito abstracto para um terreno cognitivo. E a partir do momento em que entramos num terreno cognitivo entram inevitavelmente em cena – o que não é necessariamente o caso ao nível da lógica, teoria das probabilidades, teoria da decisão e dos jogos – ciência natural e concepções de epistemologia. Coloca-se nomeadamente a seguinte questão: a quem cabe o estudo da racionalidade de agentes reais? Aos filósofos? Aos cientistas cognitivos? Os filósofos tratam tradicionalmente questões epistemológicas relativas a 'conhecimento', 'verdade', 'justificação', e fazem-no, também tradicionalmente, de forma apriorista. Os estudos científicos avançam para a recolha de dados acerca de cognição. Mas quando falamos de racionalidade o que acontece é que nem os factos do cientista cognitivo nem as normas do lógico ou do epistemólogo falam por si. Factos sem normas são 'cegos', normas sem factos são 'vazias'. O problema torna-se, obviamente, saber a quem devemos perguntar pela natureza da racionalidade. Formulada a partir do terreno da epistemologia, toda esta discussão pode ser considerada como rodando em torno da noção quineana de 'epistemologia naturalizada' e da interpretação desta (será que 'epistemologia

Aliás, de acordo com Cherniak, uma descrição psicologicamente realista da estrutura e dimensão da memória humana é suficiente só por si para impedir idealizações filosóficas do agente racional. Considerar a dimensão do corpo de crenças quando estamos perante um agente que crê que *p* e não-*p* faz toda a diferença quando se trata de avaliar quão irracional o agente é de facto. O ponto de Cherniak com a proposta da **racionalidade mínima** que resulta desta sua perspectiva é que os sistemas cognitivos a que queremos continuar a chamar agentes racionais – e de acordo com Cherniak precisamos da noção de racionalidade para fazer teoria da cognição – apenas têm que ser minimamente, e não perfeitamente, racionais (por exemplo, 'lógicos suficientemente bons', não lógicos perfeitos). Resta, é claro, caracterizar essa racionalidade mínima.

naturalizada' significa que a tarefa epistemológica passa para as mãos da ciência, ou será que significa qualquer outra coisa?). Também esta questão é tratada no Capítulo 4, *Filosofia e racionalidade teórica – em que devemos acreditar?*.

3. Modelos de racionalidade e sua suposta refutabilidade empírica

Se admitirmos que há candidatos naturais a cânones de racionalidade e observarmos que os raciocínios e decisões de agentes reais não seguem necessariamente os princípios que tais teorias nos oferecem, poderemos saber algo de mais preciso acerca da natureza, frequência, condições, destes desvios dos raciocínios e decisões de agentes reais relativamente aos princípios? Sim, podemos. De facto, existe, como referi atrás, todo um conjunto de investigações empíricas, levadas a cabo, nomeadamente, no âmbito da psicologia cognitiva, sobre raciocínio e tomada de decisão. O problema é que essas investigações aparentemente 'refutam ou desmentem' princípios normativos da racionalidade. Refiro-me, especificamente e antes de mais, ao programa de investigação ligado aos nomes de Amos Tversky e Daniel Kahneman, um psicólogo cognitivo e um economista⁶, então baseados na Universidade Hebraica de Jerusalém, e ao trabalho que eles iniciaram nos anos 70 do século XX acerca de *heuristics and biases* (heurísticas e distorções ou enviesamentos gerados pelo uso dessas heurísticas) no comportamento de juízo e decisão das pessoas⁷. A investigação psicológica sobre juízos visa os pro-

⁶ Daniel Kahneman recebeu o Prémio Nobel da Economia em 2002.

⁷ Uma série de artigos (cf. por exemplo, TVERSKY & KAHNEMAN 1971, KAHNEMAN & TVERSKY 1972, KAHNEMAN & TVERSKY 1979) despoletou uma grande quantidade de literatura e de investigação. Uma das razões para tanto interesse terá sido o facto de esses escritos indicarem caminhos para estudar experimentalmente aquilo a que Herbert Simon (1916-2001) – um dos fundadores da Inteligência Artificial – tinha chamado *bounded*

cessos utilizados para tirar conclusões a partir do conhecimento e evidência disponíveis para os sujeitos, a investigação sobre tomada de decisão visa a escolha entre opções. A conclusão principal destas investigações é que de facto a maior parte das pessoas não tem, ou pelo menos parece não utilizar, princípios correctos de raciocínio e decisão. Pelo contrário os juízos e as decisões dos agentes reais processam-se com base em princípios mais simples, heurísticas (por contraste com um **algoritmo**, que é um conjunto de regras ou instruções específicas para resolver um problema, uma **heurística** é uma regra imperfeita, que propicia uma avaliação natural e intuitiva de uma situação e constitui uma forma prática de resolver um problema na ausência de um algoritmo). No domínio dos juízos foram estudadas, por exemplo, heurísticas da representatividade (é o caso do famoso exemplo de Linda, descrito no Capítulo 1), disponibilidade, ancoragem e ajustamento.

Autores envolvidos em investigações sobre **tomada de decisão**⁸ defenderam que o seu trabalho mostrava que a teoria da escolha racional se aproxima mais das intuições normativas das pessoas (daquilo que elas pensam que deve ser feito quando se

rationality, **racionalidade limitada**. Esta racionalidade limitada (cf. DOYLE 1999) seria, segundo Simon, característica de agentes com recursos finitos (de tempo, informação, etc.), como por exemplo, os humanos, agentes cujos raciocínios e decisões não são nem podem ser, devido a essa condição, os melhores possíveis, sendo apenas suficientemente bons. A ideia de racionalidade limitada constitui uma referência na crítica às idealizações acerca de racionalidade (a ideia de *satisficing* seria assim uma caracterização mais realista do que a ideia de maximização, de que se falará mais à frente, quando se trata de decisões reais de agentes com racionalidade limitada). Na medida em que a proposta de Simon acerca de racionalidade limitada constituiu mais propriamente uma crítica filosófica do que de um paradigma de investigação, as investigações psicológicas que referi viriam a concretizá-la e a testá-la em terreno prático. A questão será retomada no Capítulo 1. KAHNEMAN, SLOVIC & TVERSKY 1982 é uma antologia de referência das investigações sobre heurísticas e 'enviesamentos' (*heuristics and biases*).

⁸ Cf. SHAFIR 1999.

escolhe) do que do seu comportamento real de escolha, e que em geral princípios psicológicos podiam ser formulados para explicar os desvios atestados no comportamento (por exemplo, em comportamento de decisão, aversão ao risco no domínio dos ganhos, procura do risco no domínio das perdas, alteração de preferências em função do *framing* – enquadramento ou descrição – da situação de escolha)⁹. Princípios semelhantes não constituem obviamente uma teoria com um grau de elaboração comparável com a teoria da decisão. No entanto parecem capturar melhor a realidade.

Como devemos interpretar os resultados dos estudos psicológicos do juízo e da decisão? Concluindo que somos irracionais, porque o nosso comportamento não se conforma aos princípios de racionalidade, ou concluindo que a a racionalidade real não se rege pelos cânones de racionalidade? Proponho que se olhe para a situação da seguinte forma: quando um investigador parte para investigações empíricas da racionalidade, especificamente investigações acerca de juízos e de tomada de decisão, ele toma determinados princípios das teorias a que acima chamei cânones de racionalidade como **modelos de racionalidade**, i. e., utiliza-os como quadros conceptuais (ideais, simplificadores) de interpretação e previsão, relativamente aos quais vai verificar se o comportamento de agentes de alguma forma se adapta. Por exemplo, a lógica diz tal e tal acerca de condicionais (por exemplo, caracteriza de determinada forma as regras do condicional material), a teoria das probabilidades diz tal e tal acerca de probabilidades (por exemplo, que a probabilidade de um evento composto não pode ser maior do que a probabilidade dos eventos simples que o compõem, ou que probabilidades prévias devem ser tomadas em consideração) e a teoria da decisão diz tal e tal acerca das preferências do agente (por exemplo, que a estruturação destas se mantém ao longo do tempo e não se modifica perante diferentes formulações da situação de escolha).

⁹ Cf. Capítulo 1.

À partida, o comportamento de raciocínio e decisão de agentes deveria conformar-se a esses princípios. Ora, constata-se que o comportamento das pessoas pura e simplesmente não se lhes conforma, que as pessoas, nas suas práticas comuns não pensam e decidem de acordo com princípios comumente considerados da lógica, da teoria das probabilidades e da teoria da decisão.

A verdade é que sabemos à partida, ao utilizar como modelo do comportamento de agentes princípios da lógica, da teoria das probabilidades ou da teoria da decisão, que estamos a utilizar idealizações simplificadoras, que de alguma forma distorcem a realidade. Mas não poderemos ainda assim considerar que são precisamente essas idealizações que nos dão a verdadeira natureza da racionalidade e que os 'desvios relativamente aos princípios' são meros erros de performance, e não pensar mais nisso? O problema é se os desvios são sistemáticos e previsíveis, o que parece ser o caso. É perante esta situação que se torna importante saber se faz sentido defender que as investigações empíricas desmentem os princípios de racionalidade tal como os encontramos formulados nas teorias mencionadas, o que constituiria uma demonstração empírica da irracionalidade de agentes. O filósofo inglês Jonathan Cohen publicou em 1981, na revista *Behavioral em Brain Sciences*, um muito citado artigo, intitulado *Can Human Irrationality Be Empirically Demonstrated?*, em que argumenta que não, que a irracionalidade não pode ser empiricamente demonstrada, que não faz sentido pensarmos que esse poderia ser o caso. Porquê? Basicamente porque critérios normativos, princípios da racionalidade, como aqueles que podemos considerar que encontramos na lógica, teoria das probabilidades e teoria da decisão, não podem ser encarados como parte da ciência natural, i. e., como hipóteses a testar, a confirmar ou refutar. A racionalidade tem que estar sempre já lá primeiro como 'quadro de referência' para podermos sequer chegar a falar de irracionalidade de agentes e os critérios normativos explícitos derivam as suas credenciais de uma sistematização das intuições. Cohen evoca uma distinção competência/perfor-

mance à maneira de Chomsky para sustentar o seu ponto (i. e., defende que devemos atribuir aos humanos uma competência para raciocinar validamente e que apenas sobre este fundo poderemos estudar os defeitos de performance atestados pelas investigações empíricas). É também possível entrar aqui com considerações de filosofia da mente e psicologia, evocando nomeadamente a necessidade de uma atribuição global de racionalidade para haver sequer atribuição de mente a um sistema, sugerindo que esta atribuição está ligada a um 'módulo da mente' e portanto a algo que nós estamos desenhados para fazer. De qualquer forma, uma teoria filosófica da racionalidade deverá pronunciar-se perante a questão dos modelos de racionalidade e sua suposta refutabilidade empírica. O Capítulo 1, *O que é ser racional?*, e o Capítulo 3, *Filosofia e racionalidade teórica – em que devemos acreditar?*, deste livro procuram dar alguns passos nesse sentido.

4. Suposição de racionalidade na teoria da mente

A ideia segundo a qual a racionalidade tem que estar sempre já presente para podermos sequer chegar a falar de agentes e da racionalidade ou irracionalidade destes é bastante comum na filosofia da mente, nomeadamente nas teorias da mente de raiz quineana (nomeadamente na ideia de tradução radical do próprio Quine¹⁰, na ideia de interpretação radical de Donald Davidson¹¹ e na ideia de estratégia intencional de Daniel Dennett¹²). Nos três autores, a teoria da mente assume a forma de uma teoria da interpretação, regida por um princípio geral 'holista' a aplicar ao comportamento de sistemas no seu ambiente, segundo o qual nada pode ser interpretado como intencional ou mental

¹⁰ QUINE 1960: 58-59.

¹¹ DAVIDSON 1984.

¹² DENNETT 1987.

se não for suficientemente racional e pelo chamado princípio de caridade, de acordo com o qual agentes têm crenças na sua maioria verdadeiras e fazem inferências na sua maioria racionais. De acordo com estas abordagens filosóficas da natureza do mental, não é sequer possível falar da mente, por exemplo, investigar empiricamente a cognição, sem supor a racionalidade de determinados sistemas físicos globalmente considerados, na sua interacção com o ambiente. Nestas condições, em que não temos como contornar a racionalidade mal começamos a falar da mente, não faz propriamente sentido pensar que se pode refutar ou desmentir a racionalidade – ela seria algo como uma condição de possibilidade de qualquer teoria do ‘mental’. A suposição de racionalidade é uma proposta de filósofos, mas é claro que toda a discussão acerca de *theory of mind*¹³ que se passa na psicologia cognitiva se torna aqui relevante: subjacente a estas discussões

¹³ **Teoria da mente** (cf. GOPNIK 1999) é o nome que se dá à perícia psicológica de atribuir mente a si próprio e aos outros. A presença da teoria da mente costuma ser testada nomeadamente com tarefas de crenças falsas. Estas testam a capacidade de uma criança reconhecer em alguém a existência de representações do mundo que são diferentes das suas e que são falsas. Há uma determinada idade aquém da qual isto não é possível. A possibilidade adquirida faz parte da teoria da mente de um adulto humano normal. Uma conhecida interpretação do autismo (BARON COHEN 1995) vê este como um défice dessa perícia (o autismo constitui uma razão forte para os teóricos da mente acreditarem que existe um módulo inato para *mindreading*: crianças autistas falham sistematicamente na tarefa de crenças falsas). Os elementos de que se necessita para compreender o que está em causa são os seguintes: crianças observam, numa peça de marionetas, a situação seguinte: Maxi, o protagonista, põe um chocolate numa caixa e sai da sala. Enquanto Maxi está fora, a sua mãe remove o chocolate para um armário. A pergunta feita às crianças-observadoras é a seguinte: onde é que Maxi vai procurar o chocolate quando entrar outra vez na sala? De forma a responder acertadamente, a criança tem que ser capaz de distinguir entre o que ela própria sabe ser verdadeiro acerca do actual estado do mundo e aquilo que ela sabe sobre os estados mentais de Maxi. Crianças com três anos falham muito, com quatro são normalmente bem sucedidas. Para uma descrição e interpretação deste teste clássico das crenças falsas, cf. WIMMER & PERNER 1983.

está em última análise a forma como um sistema atribui mente a outro sistema, por exemplo, uma pessoa a outra pessoa, ou a si própria e a forma como a racionalidade está envolvida nessa atribuição. Ora, o que investigações psicológicas sugerem é que o comportamento interpretativo de agentes, que os filósofos capturam usualmente de forma exclusivamente conceptual e descritiva, pode bem-estar relacionado com dispositivos cognitivos internos aos agentes, resultantes da evolução, dispositivos para *mindreading*, que constituiriam então a ancoragem cognitiva da linguagem da racionalidade considerada pelos filósofos referidos como condição para a descrição de um sistema cognitivo físico como mental. A ser assim, a suposição de racionalidade evocada nas teorias da mente de raiz quineana teria uma realidade cognitiva sólida, independente de, e anterior a, qualquer interpretação consciente feita pelos agentes. Constitui mais uma tarefa incontornável de uma teoria filosófica da racionalidade (talvez uma tarefa prévia) esclarecer o estatuto das suposições de racionalidade, o que envolverá a explicitação da natureza da mente e da forma como temos acesso à consideração de algo como mental, a nossa própria realidade mental e a de (outros) agentes. O entroncamento destas questões com os estudos da racionalidade é delineado no final do Capítulo 1¹⁴.

5. Estatuto do mental na realidade e estatuto dos folkconcepts

Uma conclusão global possível das investigações empíricas sobre raciocínio e decisão é o pessimismo quanto à racionalidade humana: os humanos não seriam racionais. No entanto, um outro enfoque nas investigações psicológicas da racionalidade conduz a interpretações menos pessimistas. Refiro-me à psicologia evo-

¹⁴ Um desenvolvimento do tema a partir da filosofia encontra-se em NICHOLS & STICH 2003.

lutiva. Mas de que forma pode uma perspectiva evolucionista sobre a mente e a racionalidade pesar sobre a interpretação de estudos empíricos? De acordo com uma perspectiva darwinista, ligada a uma concepção modular da arquitectura mental, os desvios sistemáticos revelados nas investigações sobre racionalidade não põem definitivamente em causa a nossa intrínseca racionalidade, antes mostram qualquer coisa acerca da natureza das nossas mentes. A ideia é que os desvios decorrem do facto de as nossas mentes serem compostas por um grande número de módulos especializados em processamento da informação, módulos moldados pela selecção natural de forma a lidarem com o tipo de problemas com que os nossos antepassados frequentemente lidavam (problemas em ‘formatos’ muito diferentes daqueles que hoje usualmente exemplificarão tarefas de racionalidade). Se isso é assim, existe razão para esperar que a performance em tarefas formuladas de modo a corresponderem àquilo que teria sido mais comum no ambiente dos nossos antepassados melhoraria. Ora, isto parece verificar-se. Do ponto de vista da teoria da mente, o resultado seria que as categorias e distinções utilizadas por filósofos e cientistas cognitivos para analisar conhecimento e raciocínio – nomeadamente a própria noção de ‘racionalidade’ – teriam que ser reformuladas (ou, pelo menos, o seu estatuto teria que ser repensado). Uma discussão semelhante põe em causa não apenas o estatuto do mental na realidade como o estatuto dos *folkconcepts* que usamos para falar e pensar sobre a mente. Podemos olhar para a situação da seguinte forma: a ciência cognitiva toma a psicologia de senso comum (a forma usual de falarmos do mental, de nos interpretarmos e aos outros, como mentais, em termos de crenças, desejos, finalidades, motivações) como ponto de partida – nada a obriga no entanto a mantê-la como ponto de chegada, após investigações científicas sobre a mente e a cognição. Mesmo se a forma intuitiva e comum de falarmos e pensarmos sobre o mental for em última análise inabdicável na nossa vida comum, na medida em que ela pode bem ser produto de dispositivos cognitivos específicos em nós e não um reflexo directo

da realidade, ela poderá ser explicada a partir de resultados científicos sobre a mente e a cognição que escapam ao nosso auto-apercebimento. É por essa razão que é necessária uma análise filosófica – epistemológica e metafísica – do que fazemos quando falamos e pensamos sobre a mente, do que fazemos quando procuramos elaborar as relações entre a psicologia de senso comum e a ciência cognitiva. Pode bem dar-se o caso de essa análise estar destinada a desfazer o privilégio das nossas intuições, mesmo se estas dizem respeito à nossa própria mentalidade. Também essa tarefa cabe a uma teoria filosófica da racionalidade: não basta esclarecer o que fazemos quando transpomos cânones abstractos e os usamos como modelos de racionalidade na investigação empírica acerca de raciocínio e decisão de agentes – é preciso esclarecer o estatuto da forma usual de tomarmos o mental, forma que pode estar instalada em nós profundamente, ser nomeadamente resultante da evolução de uma ‘teoria da mente’. Esta por sua vez pode ser bem real no nosso funcionamento cognitivo – tudo nos leva a pensar que o é – sem necessariamente reflectir a realidade das coisas. Uma teoria filosófica da racionalidade deverá assim incluir uma posição perante a questão da relação entre a linguagem mentalista, a ciência cognitiva e a metafísica. Esta questão é referida nos Capítulos 1 e 3.

6. *Desejos e estrutura motivacional dos agentes. Acção e racionalidade na acção*

O facto de determinados princípios normativos do raciocínio e da acção serem estipulados por lógicos, estatísticos e teóricos da decisão, não significa de forma alguma que estes sejam necessariamente seguidos por agentes reais nos seus afazeres cognitivos. Mas porque é que os agentes não seguiriam os princípios normativos? Uma primeira hipótese é que esses princípios são idealizações que não capturam o funcionamento da racionalidade

real (a arquitectura cognitiva dos agentes, o processamento da informação, etc.). Uma outra hipótese tem a ver com a **estrutura motivacional** do agente, com a forma como crenças, desejos, intenções, emoções, razões se relacionam para conduzir o agente à acção. De facto, porque é que havemos de pensar e agir racionalmente, mesmo vendo claramente o que isso significaria? Qualquer um de nós pode saber o que deveria pensar, o que deveria fazer, à luz de determinadas considerações de racionalidade em que inclusivamente pode assentir conscientemente, e no entanto não o pensar, não o fazer. O que isto mostra é que outros aspectos da mente ou espírito além de representação e cálculo estão (pelo menos em nós, humanos) envolvidos naquilo a que chamamos racionalidade. Se queremos saber o que nos move à acção temos que considerar uma dimensão de desejo. A questão da **motivação** pertence certamente à psicologia e à neurociência¹⁵, vou no entanto tratá-la neste livro sobretudo como uma questão do domínio da filosofia moral, a questão de saber o que nos faz prosseguir objectivos simbólicos complexos e também a questão de saber como podemos moldar a nossa vontade, a nossa forma de avançar para a acção. Noutras palavras, penso que falar da motivação para a acção, de iniciação da acção, pode e deve envolver questões relativas a neurónios motores e funções executivas e iniciadoras do córtex frontal, questões relativas à base neurobiológica das **emoções**¹⁶, questões relativas

¹⁵ Não é simples definir **motivação**. Em contextos de psicologia e neurociência, a motivação é ligada à avaliação e selecção de finalidades (*goals*) do comportamento e aos mecanismos cognitivos/ neuroquímicos/ hormonais subjacentes a esses processos e pode ser definida como «uma influência moduladora e coordenadora sobre a direcção, vigor e composição do comportamento», com fontes internas, ambientais e sociais. Mudanças de estado motivacional manifestam-se em vários níveis da organização comportamental e neuronal de agentes (SHIZGAL 1999: 566).

¹⁶ Podemos definir **emoções** (cf. BROTHERS 1999 e OATLEY 1999) como estado ou processos psicológicos que funcionam na gestão de fins de agentes. Emoções são tipicamente provocadas por avaliações de eventos como relevantes para dados fins: elas serão positivas quando se avança

ao 'amor-próprio' que se esconde por baixo de todos os nossos fins, questões relativas aquilo que o nosso amor-próprio', o nosso egoísmo, 'quer' para nós, mas deve também poder ir até ao ponto em que procuramos compreender a forma como as ideias de Deus, liberdade ou verdade, nos podem motivar. Entretanto, entre a iniciação cortical e o ser movido à acção por ideias como Deus ou a verdade, ou o interesse próprio, está em causa a passagem conceptual de consideração acerca do movimento corporal de um corpo humano e daquilo que o controla, à 'acção', com todas as conotações mentais e éticas do termo 'acção', relativas nomeadamente a consciência e voluntariedade, questões estas que estão ausentes do discurso acerca de funções motoras e executivas. Este é um problema fundamental para a filosofia da

no caminho em ordem ao fim, negativas quando se é de alguma forma impedido de avançar; os agentes de que aqui se fala são animais, humanos e não-humanos. As emoções funcionam ainda como forma de 'colocar como prioritário' algum fim ou plano do agente relativamente à acção ou vida mental em geral. No seu sentido comum, que importa reter, 'emoção' cobre quer o sentimento interno quer a expressão comportamental. As emoções têm assim uma dupla face exprimir / sentir. Encontram-se teorias da natureza das emoções em autores que vão desde Aristóteles, Descartes e Espinosa a Charles Darwin e Sigmund Freud (estes dois últimos autores são especialmente importantes, historicamente, por terem considerado as emoções fundamentais naquilo a que hoje se chamaria a *mente naturalizada*). O trabalho neurológico e filosófico de António Damásio tem dado grande importância à questão das emoções, e precisamente, ao seu peso naquilo a que chamamos 'racionalidade'. Damásio distingue emoções primárias (medo/ansiedade, fúria/raiva, tristeza, felicidade, nojo – é esta de resto a lista comum nos manuais de psicologia) de emoções sociais tais como simpatia, compaixão, embaraço, vergonha, culpa, orgulho, ciúme, inveja, gratidão, admiração, espanto, indignação, desprezo. Em *O Erro de Descartes* (DAMÁSIO 1994) Damásio analisa o papel das emoções no comportamento de decisão, em *O Sentimento de Si* (DAMÁSIO 1999) analisa o papel das emoções (e sentimentos, que considera virem depois) na constituição da identidade própria (self, eu, identidade pessoal). Em *Ao Encontro de Espinosa* (DAMÁSIO 2003) analisa directamente a natureza e significado de emoções e sentimentos.

mente e a filosofia da acção. Mas é também fundamental notar, partir do ponto de vista da filosofia moral, e depois de se ter aceite alguma concepção de acção e de intenção, que não existe uma versão única daquilo a que vou chamar estrutura motivacional de agentes (estou aqui a falar apenas de agentes humanos). Se a história da filosofia pode aqui ensinar-nos alguma coisa, o que ela nos ensina é que é possível desenvolver e defender de forma razoável **concepções mais ou menos racionalistas da nossa estrutura motivacional**. Neste livro a oposição entre mais e menos racionalismo na forma de conceber a motivação para a acção será simbolizada por I. Kant e D. Hume. Tanto Kant como Hume se preocuparam em saber o que move um agente à acção. No entanto, as suas posições acerca da natureza da mente levaram-nos a defender propostas totalmente diferentes. De acordo com D. Hume, a razão é uma escrava das paixões, não tendo por si própria qualquer força motivacional. O que Hume quer dizer é que aquilo que cabe à razão (representar que tal e tal é o caso, calcular as possíveis consequências de acções) não é suficiente para nos mover à acção. É preciso que estejam lá os desejos. Só desejos nos movem à acção. Pelo contrário de acordo com Kant os desejos constituem em última análise uma interferência (indesejada...) na passagem do agente à acção, uma espécie de escudo psicológico incómodo que além do mais 'esconde' do próprio agente a natureza da sua (possivelmente racional) motivação para agir. A Razão é para Kant o único motor apropriado, correcto, da acção e é sua convicção que por princípio os seres humanos podem agir de forma totalmente racional (ou pelo menos espera-se que sim, embora a questão – saber se estão ou não a agir de forma totalmente racional – possa ser, em última análise, para os sujeitos humanos eles próprios, obscura). É muito importante compreender que isto não significa que Hume defende que nós somos irracionais e Kant defende que somos racionais: de facto, ambos desenvolvem uma teoria do que é sermos racionais. Acontece que as teorias propostas apresentam ideias profundamente diferentes acerca daquilo que motiva os

humanos à acção. Hume e Kant constituem uma referência, na literatura filosófica contemporânea – mais exactamente nos estudos da racionalidade feitos no âmbito da filosofia, e dentro da filosofia, na filosofia moral e política. Podemos mesmo falar de duas linhagens, uma linhagem sentimentalista e uma linhagem racionalista, exponenciadas respectivamente por Hume e Kant, na forma de conceber a motivação para a acção, linhagens estas que são perceptíveis na filosofia moral e política.

Este não é um problema indiferente: a forma como a estrutura motivacional do agente é concebida, a forma como crenças e desejos se relacionam, é sempre tácita ou explicitamente considerada em qualquer tratamento da racionalidade. Evidentemente, como afirmei, estas questões de filosofia moral quanto ao que move o agente à acção pressupõem, têm aquém, questões de teoria da acção (antes de mais, saber o que constitui uma acção, o que é para um agente agir, por contraste com o que lhe acontece, e o que é uma acção intencional). Só perante eventos considerados como acções intencionais poderemos vir a falar de racionalidade na acção, motivação para a acção, razões para agir. Todas estas questões são tratadas no Capítulo 2, *Filosofia e racionalidade prática – o que devemos fazer?*, e em grande medida também através da análise da teoria sentimentalista da racionalidade prática de Simon Blackburn no Capítulo 5, *Algumas teorias da racionalidade*.

7. A questão alargada da racionalidade

Se a racionalidade começa por ser uma característica de agentes cognitivos – pelo menos é como tal que ela é mais directamente analisável, nomeadamente em termos de ciência cognitiva – à filosofia ela interessou desde sempre também em termos de 'produtos', i. e., teorias, construções de pensamento, abordagens não apenas científicas mas também morais, estéticas, políticas e religiosas da realidade e organizações da própria realidade (por

exemplo, organizações sociais). Em última análise, apenas uma resposta ambiciosa, que cobrisse um campo que fosse desde os agentes, a sua arquitectura cognitiva e a evolução desta por selecção natural, às produções científicas, morais, estéticas, religiosas do pensamento humano e às organizações sociais e políticas dos humanos, constituiria uma teoria filosófica da racionalidade. Quando falo de produções falo da racionalidade não já apenas enquanto **cognição** mas enquanto **pensamento**. Quando falo de pensamento por contraste com cognição estou a propor que se considere a 'cognição' como os vários processos de informação, representação, computação e decisão envolvidos na adaptação de um agente ao seu ambiente, processos esses que podem ser sub-pessoais e inconscientes e que se considere o 'pensamento' como envolvendo as dimensões pragmática, qualitativa, voluntária e explicitamente normativa que esses processos podem atingir num ser humano consciente. Se é essa dimensão que visamos, o foco de análise de uma teoria da racionalidade não pode ser apenas o agente cognitivo, tem que ser o sujeito, e o pensamento enquanto acção (de assentir, negar, opor, subscrever e defender teses acerca da forma como as coisas são) e 'sentimento' (compreensão, *qualia* do conhecimento¹⁷). É a isso que podemos chamar entendimento. É em função desta distinção que eu falaria, por exemplo, em 'pensamento filosófico' e 'pensamento científico' e não 'cognição filosófica' ou 'cognição científica', muito embora, como é óbvio esse pensamento só seja possível a partir da cognição. A distinção é importante quando se fala de racionalidade, e pode ser esclarecedora porque alguns estudos da racionalidade são feitos do ponto de vista da cognição (por exemplo, o estudo dos processos de decisão, raciocínio e emoções na psicologia cognitiva e na neurociência) mas os problemas da racionalidade não se restringem a problemas cognitivos (por exemplo, eu não consideraria que os problemas acerca da natureza da justificação mencionados no ponto 1 são problemas cognitivos).

¹⁷ A expressão é de Fernando Gil.

É relativamente a esta questão alargada da racionalidade que a comparação entre a tradição analítica e a tradição fenomenológica e hermenêutica no que respeito aos estudos da racionalidade se revelaria possivelmente pertinente. De facto, o que frequentemente acontece é o seguinte: as abordagens da racionalidade feitas no âmbito da tradição fenomenológica e hermenêutica e mesmo das ciências sociais frequentemente passam por cima das questões cognitivas da racionalidade, da forma como estas pesam no comportamento dos indivíduos, avançando directamente para as macro-questões da 'Razão na sociedade, na história e na civilização' relativas a produções teóricas e organização social. Por sua vez, os estudos da racionalidade feitos no âmbito da tradição analítica e da ciência cognitiva demasiadas vezes ficam aquém destas questões¹⁸. Que ideias acerca de racionalidade se encontra então no 'outro campo' da tradição fenomenológica e hermenêutica e das ciências sociais? Antes de mais, verifica-se que o tema da racionalidade faz frequentemente surgir os nomes de M. Weber (1864-1920) e S. Freud (1856-1939), e as referências a **racionalização no sentido weberiano** e **racionalização no sentido freudiano**. A racionalização weberiana – a racionalização como realização social da razão no sentido burocrático, administrativo, ligada à substituição social da religião por uma razão instrumental, identificada com a técnica, o cálculo, a organização – é usualmente considerada como um traço característico da Modernidade enquanto época e a questão que se coloca diz respeito à capacidade de a razão ligar os indivíduos em alguma totalidade ética, como em tempos a religião o teria feito. Este é o uso de 'racionalidade' e 'racionalização' mais comum nas ciências sociais e é muito importante, nessa literatura, para pensar sobre as sociedades contemporâneas. Por sua vez, a noção freudiana de racionalização (processo pelo qual um indivíduo

¹⁸ Uma intersecção possível dá-se na área da filosofia moral e política. Pense-se, por exemplo, na teoria rawlsiana da justiça, que pretende incorporar o kantismo e a teoria da decisão racional, e prolongá-los numa teoria da racionalidade social.

apresenta uma explicação ou descrição aceitável e coerente para uma acção ou pensamento (seu cujos motivos verdadeiros não conhece) é importante para pensar sobre o indivíduo e a forma como este representa as acções próprias (no contexto filosófico tratado neste livro, estes temas surgirão relacionados com a intenção na acção e as razões para agir). Weber e Freud são dois dos nomes mais importantes do pensamento social e psicológico sobre a racionalidade, marcando a forma como a discussão em torno da racionalidade se trava em contextos académicos e mais gerais. Convém notar que as teorias da racionalidade a encontrar aqui não se auto-denominam, necessariamente, teorias da racionalidade, ou com certa humildade metodológica estudos da racionalidade, mas frequentemente 'críticas da racionalidade'. A racionalidade é pejorativamente apelidada de 'ocidental' ou 'cartesiana', e os discursos tomam frequentemente a forma de um coro de vozes contra a razão – acusada de dominação, opressão, patriarcado, totalitarismo, terror. Em termos de produções teóricas, esta crítica é frequentemente dirigida aos aspectos teleológicos e finalistas das narrativas filosóficas, que seriam narrativas da realização da Razão (a ideia seria que 'finalmente em Descartes', 'finalmente em Husserl', etc., o pensamento veio a chegar à compreensão de sua própria natureza) e em termos de organização social e civilizacional à ideia optimista, iluminista, progressista de um desenvolvimento da Razão na história humana. Se estas abordagens críticas da racionalidade não visam usualmente aspectos cognitivos (agentes, arquitectura de agentes, raciocínio, decisão), elas têm a vantagem de enfatizar questões históricas (relativas ao 'debate' entre a Razão Iluminista – a razão pensada como liberdade, emancipação e progresso, em última análise como fonte do projecto civilizacional iluminista da Modernidade – e a face oculta dessa Razão, concebida como opressão, extensão global de uma concepção – ocidental – de civilização, concebida como a única possível). É difícil estabelecer a passagem entre estas formas muito diferentes de abordar a questão da racionalidade, uma centrada nos indi-

víduos, na cognição, em raciocínios e decisões e nos princípios que justificam estes, e outra uma macro-perspectiva visando a sociedade, a história e a civilização. No entanto, o interesse da racionalidade enquanto tema filosófico tem precisamente a ver com essa possível passagem dupla da filosofia para fora da si: da filosofia para as ciências cognitivas e em última análise para as ciências naturais quando consideramos agentes, arquitecturas, etc., por um lado, e por outro lado a passagem da filosofia para as ciências sociais, políticas e jurídicas quando consideramos questões de sociedade, história, civilização. Esta última passagem é aquilo a que chamo no contexto do presente livro 'questão alargada da racionalidade'. Embora a questão não vá aqui ser tratada directamente, ela é muito importante para o propósito comparativo de tradições filosóficas que deu origem ao presente livro. De qualquer forma, estou convencida de que começar pelos estudos da racionalidade na filosofia analítica e ciência cognitiva não é uma má escolha metodológica, até porque, como pergunta o filósofo americano Robert Nozick (1938-2002) em *The Nature of Rationality* (1993), reconhecendo precisamente este coro comum de crítica à racionalidade na filosofia e cultura: qual é a alternativa à racionalidade se queremos discutir a natureza e a qualidade do pensamento e da acção?

8. A questão fenomenológica da racionalidade

Um outro aspecto a considerar na questão da racionalidade, e que também não é directamente visado pelas investigações empíricas, relaciona-se com a dimensão fenomenológica do mental. Usualmente distingue-se, ao falar da mente, a intencionalidade ou representação por um lado, e o aspecto qualitativo da mentalidade, o sentir-se ser, a que normalmente se chama *qualia*, por outro. Ora, essa distinção pode ser alargada da mente ao pensamento, e o alargamento passa pelas seguintes considerações. No conhecimento, no pensamento, não se trata apenas

do que é pensado/ afirmado, justificadamente ou não, cerca da realidade, em qualquer domínio de inquérito, mas também de algo que se passa 'a partir do interior', a partir da subjectividade, nomeadamente a aquiescência e rejeição pela mente quanto àquilo que é tomado como verdadeiro, bom, belo, justo ou falso, mau, feio, injusto. À forma como o conhecimento e o pensamento estão para o sujeito podemos chamar 'compreensão'. A compreensão é o *quale* do pensamento (F. Gil), e aquilo a que estou a chamar aqui questão fenomenológica da racionalidade visa descrever a partir de dentro como é ter entendimento, compreensão, pensamento. Trata-se de compreender 'como é ser racional', para o próprio sujeito. Todo o programa investigação do filósofo português Fernando Gil tem sido dedicado a esta questão, e penso que é assim que devem ser entendidas as suas teses acerca de provas, evidência, convicção (Gil 1993, Gil 2000). Esse programa assume a necessidade de uma fenomenologia de termos epistémicos tais como evidência, prova e convicção, a qual revelará a importância da explicitação da dimensão volitiva / afectiva do espírito na relação com os conteúdos representados, afirmados e negados. Apenas a partir de uma análise desses aspectos será possível falar da racionalidade como 'entendimento'. Também este ponto, como o anterior, opera a transição entre a tradição analítica e ciência cognitiva e a tradição fenomenológica e hermenêutica e não será directamente focado neste livro¹⁹.

¹⁹ Faz, no entanto, parte dos objectivos do Projecto *Racionalidade, Desejo, Crença – a motivação para a acção do ponto de vista da teoria da mente* abordar esse problema para uma teoria da racionalidade a partir da obra de F. Gil (GIL 1986, GIL 1993, 2000). Ao contrário da maioria dos autores directamente analisados nesse projecto, esta obra não se situa no âmbito da filosofia analítica. Ela faz aliás confluír no estudo das questões visadas o pensamento de autores como Husserl e Freud, habitualmente estranhos à tradição analítica. Agradeço a Paulo Tunhas e a André Barata os textos sobre a obra de F. Gil com que contribuíram para o projecto e que se encontram em www.lettras.up.pt/df/if/gfmc/fil_mente.html.

9. A questão pessoal da racionalidade

Ao definir²⁰ o homem como o animal racional Aristóteles ligou definitivamente a questão da racionalidade ao que pensamos (e valorizamos) acerca de nós próprios enquanto espécie. Mas a questão da racionalidade preocupa-nos e diz-nos respeito também, talvez sobretudo, como questão pessoal – queremos saber, interessa-nos saber, a cada um de nós, se somos e como somos racionais; a nossa própria natureza interessa-nos. Tanto quanto queremos saber porque pensamos como pensamos, por que agimos como agimos, por que acreditamos no que não temos razões para acreditar, por que não acreditamos no que temos razões para acreditar, por que não fazemos o que queremos fazer, acreditamos que devemos fazer e temos razões para fazer, por que fazemos aquilo que acreditamos que não devemos fazer, o assunto não nos é alheio. Chamarei a esta questão a questão pessoal da racionalidade: ela liga o interesse que cada um de nós tem pela sua própria natureza, por aquilo que em nós é talvez distintamente humano, ao nosso interesse na racionalidade. Ora, aquilo que é novo na situação anti-cartesiana instaurada pela ciência cognitiva, e que se reflecte no tratamento da racionalidade, é que não basta olhar para dentro do espírito para saber o que é o espírito, não basta ser suficientemente capaz de pensar e agir racionalmente para saber o que é a racionalidade – é necessário levar em conta a investigação empírica. Aquilo que se mantém inalterado é a importância, no tipo de ser que somos, do compromisso com nossa própria vida mental, nomeadamente em termos de identidade pessoal, que em grande medida se exerce como avaliação da raciona-

²⁰ Embora quase sempre atribuída a Aristóteles, e certamente tornada corrente no comentário do seu pensamento, uma definição do próprio do homem como sendo a racionalidade não se encontra exactamente nesses termos no corpus aristotélico (definição, para Aristóteles, envolveria predicação essencial, dizer o que é distintivo, o que é próprio). Cf. SARDO 2000.

lidade própria²¹. Penso que o nosso interesse pela questão da racionalidade se relaciona em grande medida com esta questão pessoal da racionalidade. Ela será tratada sobretudo no Capítulo 4, a partir de Robert Nozick.

Pretendi, na presente Introdução, enumerar, sob a forma de questões que constituem tarefas para a filosofia, várias razões pelas quais me parece não existir coincidência entre os estudos científicos da racionalidade e uma teoria filosófica da racionalidade. A situação presente aparece então com as seguintes características: temos um terreno tradicional da filosofia – parece-me que é difícil ou impossível nomear um filósofo que não tenha pensado sobre os processos do pensamento e sobre a motivação para a acção – no qual as ciências da cognição entraram, através dos estudos do raciocínio, da decisão, das emoções. Mas isso não significa que a filosofia tenha sido substituída ou afastada – pelo contrário parece-me que as abordagens científicas de aspectos da racionalidade requerem de novo a filosofia, pelas várias razões enumeradas. De forma a preparar o caminho para enfrentar esse desafio, nos capítulos que se seguem vou procurar mapear os estudos científicos e filosóficos da racionalidade e referenciar autores, dados e programas de investigação. Decidi apontar ainda alguns exemplos de teorias filosóficas da racionalidade, ou passos nesse sentido, dados por filósofos analíticos contemporâneos. Alguns dos filósofos cuja obra pode ser considerada como uma resposta ao desafio que referi são Robert Nozick, Simon Blackburn, Alvin Goldman e Stephen Stich (Nozick 1993, Blackburn 19998, Stich 1993 Goldman 1986). Este livro inclui a análise de algumas das suas propostas. Serão, nomeadamente,

²¹ Para um filósofo como D. Dennett a avaliação da racionalidade própria é um factor essencial na constituição da *personalidade* da nossa mente. Esta é uma questão que se coloca depois das questões da representação, da consciência e da auto-consciência (MIGUENS 2002: 420-421).

consideradas e discutidas as teses acerca da natureza da justificação avançadas por A. Goldman (Goldman 1986), a teoria pragmatista da racionalidade defendida por S. Stich, bem como os *papeis relativos da filosofia e da ciência cognitiva* nessa tarefa (Stich 1993) a teoria humeana, sentimentalista, da racionalidade prática avançada por S. Blackburn (Blackburn 1998) e a teoria da racionalidade simbólica de R. Nozick (Nozick 1993). Idealmente será possível a partir destas propostas avaliar concepções mais ou menos racionalistas daquilo que em geral se entende por racionalidade, no seio de uma teoria geral da mente e da acção que não se pretende apriorista mas que parte das aporções das várias ciências da cognição.

Apesar de este livro pretender ser sobretudo um guia, espero que ele seja suficiente para dar uma noção da variedade de tratamentos da racionalidade em curso na filosofia analítica e na ciência cognitiva. O problema da natureza da racionalidade é, de certa forma, tão antigo como a própria filosofia, no entanto é um desafio específico procurar mapear e explorar os meios conceptuais com que hoje o enfrentamos e explicitar as mudanças para o conceito que essas investigações podem trazer. Filósofos como Aristóteles, Hume, Kant, Ramsey, Anscombe, Davidson, Stich, Goldman, Blackburn e Nozick servem-me neste livro de apoio e pretexto.

Capítulo 1

O que é ser racional? Investigações sobre racionalidade: história e organização da literatura.

«Ora, quem exerce e cultiva a sua razão parece desfrutar ao mesmo tempo a melhor disposição de espírito e ser extremamente caro aos deuses»

Aristóteles, *Ética a Nicómaco* (Livro X, 8)

«Vou procurar provar, primeiro, que a razão só por si não pode ser motivo para qualquer acto da vontade, e, segundo, que ela jamais se pode opor à paixão na direcção da vontade. (...) A razão é, e deve ser apenas, a serva das paixões; não pode aspirar a outro papel senão servi-las e obedecer-lhes»

David Hume, *Tratado da Natureza Humana*
(Livro II, Parte III, Secção III)

Racionalidade: racionalidade instrumental, racionalidade teórica, racionalidade prática

O que é ser racional? De quê ou de quem afirmamos a racionalidade? De pessoas, acções, pensamentos, crenças? Onde podemos ir buscar uma primeira definição operacional de racionalidade, um ponto de partida para uma investigação?

Antes de olharmos um pouco para a história do problema da racionalidade e para o seu tratamento na literatura, procuremos nas nossas intuições: veremos que começar pela **definição instrumental de ‘racionalidade’** é a forma natural de começar. 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).

A definição de ‘racionalidade’ que acabei de avançar é a mais consensual e comum. Corresponde à chamada definição instrumental da racionalidade, que nos diz o que é **racionalidade na acção**. A definição instrumental pode ainda fazer referência a processos mentais (crenças, desejos) *envolvidos num processo de controlo 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 ela com as teorias a que na Introdução chamei 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 e 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 acção¹.

¹ Por exemplo, de acordo com o princípio da maximização da utilidade esperada, o agente vai pela opção que apresente a maior utilidade esperada, de acordo com o princípio maximin o agente age de modo a que a pior consequência possível da sua acção seja a melhor possível. As escolhas não têm que coincidir: ao contrário do primeiro princípio, o segundo visa ‘maximizar’ o ‘resultado mínimo’, i. e., escolher de forma a que o pior resultado possível seja ainda assim o melhor possível.

A definição instrumental da racionalidade tem como referência o comportamento de agentes. No entanto, a verdade é que frequentemente evocamos a racionalidade para falar do decurso dos nossos raciocínios explícitos e portanto passamos do campo da acção e do comportamento para o campo das inferências, crenças, teste de hipóteses, teorias. Não se trata aqui apenas de um comportamento global de um agente 'ordenado' em direcção a uma finalidade e para tal mobilizando meios, mas de movimentos de pensamento, passagens de crença em crença, movimentos mediante os quais acontecem coisas tais como adquirir novas crenças, eliminar crenças, sustentar crenças com outras crenças, justificar crenças. Como se liga uma coisa e outra, a racionalidade comportamental dos agentes, instrumentalmente avaliada, e a racionalidade nos movimentos de pensamento? Será que alguma delas pode ser 'reduzida' à outra? Esta é uma questão cuja resposta convoca tudo o que vai ser dito neste livro. Para já, introduzirei classificações. Vou chamar a umas questões da racionalidade questões da **racionalidade prática** e a outras questões de **racionalidade teórica**.

Olhando para a literatura contemporânea dedicada à questão da racionalidade, vemos que ela se distribui em geral por tratamentos da racionalidade teórica e da racionalidade prática. Frequentemente – sobretudo na ciência cognitiva e especialmente na psicologia cognitiva – as questões são discutidas sob a forma de análise (ou estudo experimental) do raciocínio (*reasoning*) e da decisão. De que se fala então quando se fala de racionalidade teórica e raciocínio teórico? Basicamente, e como já sabemos, a **racionalidade teórica é racionalidade nas crenças**. Em princípio – embora compreender o que isso envolve seja o fulcro da discussão filosófica – as nossas crenças são racionais se temos boas razões para as sustentar e se são fiáveis (*reliable*) na forma como representam o mundo. Como disse o filósofo americano Robert Nozick, de quem falarei bastante neste livro, numa clara alusão a Kant, e a propósito da racionalidade nas crenças, «razões sem fiabilidade são vazias,

fiabilidade sem razões é cega»². Quanto à **racionalidade prática**, ela é **racionalidade na acção**, e o raciocínio prático é o raciocínio que eventualmente afecta planos e intenções do agente e resulta numa determinada decisão que conduz à acção.

Na literatura dedicada às questões da racionalidade temos portanto de um lado um foco sobre o raciocínio teórico, aquele que afecta crenças. Como afirmei, a racionalidade destas é avaliada em termos de razões para acreditar e fiabilidade. Sob análise temos então um agente, que tem determinadas crenças, sobre o que é e não é o caso no mundo e que as altera ou mantém face a determinada evidência, utilizando processos de inferência para partir das crenças e de nova evidência, e em princípio não chegando a conclusões injustificadas a partir da evidência disponível. Aquilo que se passa nesta literatura é uma análise ou investigação do que são crenças justificadas, do que pode e deve contar como razão para o agente passar a ter, ou deixar de ter, uma crença, uma explicitação e proposta dos princípios que determinam que crenças devem seguir-se de outras crenças, etc.³. A questão básica da racionalidade teórica é saber *em que devemos acreditar*.

² NOZICK 1993: 64.

³ O cruzamento das questões da racionalidade teórica com a literatura filosófica da área da epistemologia é óbvio. Cabe por isso recordar que as principais teorias da justificação epistémica discutidas na epistemologia são o fundacionalismo, o coerentismo e o fiabilismo. O **fundacionalismo** é a ideia segundo a qual uma crença conta como justificada se e só se é ou uma crença básica, caso em que é 'auto-justificada' (por exemplo, uma crença acerca de experiências sensoriais, no caso do empirismo, ou uma intuição intelectual tal como o cogito, no caso do racionalismo cartesiano), ou uma crença mediada que em última análise se reporta a uma crença básica. O **coerentismo** é a ideia segundo a qual uma crença está justificada se a nossa teoria acerca do mundo é mais coerente com ela do que sem ela. O **fiabilismo** é a ideia segundo a qual uma crença é justificada se e só se é produzida por processos fiáveis do sistema cognitivo (a tese é usualmente acompanhada com a referência à *ratio* de crenças verdadeiras produzidas pelo sistema).

Por outro lado temos um foco sobre o raciocínio prático, que afecta planos e intenções (no seu núcleo está a questão da desejabilidade de acções: um agente tem determinadas crenças, planos e intenções, delibera sobre qual acção deve realizar, por qual curso de acção deve optar por entre várias alternativas, e finalmente age). A questão básica da racionalidade prática é saber o *que devemos fazer*.

Em ambos os domínios, racionalidade nas crenças e racionalidade na acção, há espaço para a **irracionalidade**: são casos de irracionalidade, acreditar no que não temos razões para acreditar, não acreditar no que temos razões para acreditar, não fazer o que queremos fazer, acreditamos que devemos fazer e temos boas razões para fazer, fazer aquilo que acreditamos que não devemos fazer (os exemplos abundam e incluem não apenas os casos – que chamarão a atenção em contexto de investigação científica ou policial, em que não fazemos pesar a evidência ou provas de que dispomos de forma a reajustar o nosso corpo de crenças, ou acreditamos em conclusões injustificáveis a partir dos dados de partida, como também casos importantes para compreender a nossa própria natureza (na medida em que constituem patologias correntes da nossa forma de sermos racionais), tais como o **auto-engano**, caso em que escondemos de nós próprios que possuímos razões para acreditar que algo é o caso e não acreditamos e a **fraqueza da vontade** (*akrasia*), caso em que queremos fazer algo e não o fazemos – i. e., agimos e agimos intencionalmente de forma contrária ao nosso melhor juízo, etc.)⁴.

⁴ Parece haver certas diferenças entre raciocínio teórico e raciocínio prático no que respeita à irracionalidade ou não irracionalidade de procedimentos: como nota G. Harman (HARMAN 1995), escolher arbitrariamente o que fazer pode ser racional num campo prático (não é irracional, por exemplo, decidir arbitrariamente por que estrada ir, quando três estradas me parecem igualmente boas e vão ter ao mesmo destino) mas o mesmo não acontece no campo teórico (se não disponho de evidência que me permita saber se quem ganhou a lotaria foi o João, o António ou o Manuel

A distinção entre racionalidade teórica e racionalidade prática tem o seu quê de artificial e rígido, mas ela convém-nos para começar, na medida em que nos permite organizar contributos quer da história da filosofia quer de várias disciplinas da cognição.

Aristóteles, a racionalidade teórica e a racionalidade prática

Voltemos agora atrás na história do pensamento, a um autor, Aristóteles, que é inevitavelmente convocado, e acertadamente, quando se fala de racionalidade. Aristóteles teve certamente muito a dizer quer acerca do raciocínio teórico quer acerca do raciocínio prático, nomeadamente enquanto fundador da lógica, teorizador da natureza da acção, da natureza da ética e daquilo que se entende por 'ciência'. Pensando no que Aristóteles fez enquanto lógico e teórico da ciência por um lado e enquanto teórico da acção e da ética por outro, temos de resto um primeiro desenho da partição do campo de estudos da racionalidade em racionalidade teórica e racionalidade prática.

Como é sabido, Aristóteles interessou-se pela análise da forma como pensamos naquilo que faz uma coisa ser aquilo que é, e esse seria o contexto apropriado para entendermos a afirmação que lhe atribuída segundo a qual o homem é aquele de entre os animais cuja diferença específica é a racionalidade, um ser no qual a 'parte racional' da alma é superior à 'parte irracional'. A racionalidade seria assim algo que caracteriza os humanos

é irracional decidir arbitrariamente acreditar que foi o João). Assim, escolher agir à sorte parece poder ser racional, enquanto o mesmo não se passa com escolher acreditar à sorte. De forma algo análoga, *wishful thinking* pode ter bons efeitos práticos (Maria deseja muito ter sucesso numa competição desportiva, acredita que vai ter sucesso, e isso leva-a a treinar muito para a provar), mas não teóricos (Maria deseja muito passar de ano, fez um exame, e, antes de saber o resultado, conclui que passou de ano).

por contraste, por exemplo, com a percepção sensorial, que é comum a várias espécies animais, e portanto seria uma marca dos humanos por entre os animais. Se quisermos encontrar a definição do homem como animal racional no próprio Aristóteles, teremos dificuldades: o mais próximo seria talvez '*zoon legon echon*' (o animal cujo próprio é falar). De qualquer forma, quando Aristóteles (supostamente) definiu o homem como o animal racional não pensava certamente na racionalidade como mais uma adaptação evolutiva da nossa espécie – como hoje temos boas razões para fazer⁵ – mas como algo que distinguia os homens por entre os outros seres, especificamente por entre os animais, ligado nomeadamente à capacidade de 'ver os universais', inserir pensamentos em demonstrações, fazer ciência e agir moralmente, por 'desejo deliberado'. O animal racional é o animal com *logos* (razão, linguagem), com *nous* (intelecto), capaz de abstrair das realidades universais ou ideias, e de articular ideias em pensamentos e os pensamentos em ciência, ciência essa que capta a essência da realidade. São essas capacidades que não se encontram nos animais, mesmo se estes são capazes de cognição no sentido mais elementar de percepção sensorial e acção por instinto. Sendo o homem o animal capaz de tais coisas – abstrair da realidade universais ou ideias, articular ideias em pensamentos e pensamentos em ciência – Aristóteles preocupou-se também em explicitar os princípios que determinam a forma como pensamentos devem seguir-se de outros pensamentos: foi o que fez enquanto fundador da lógica.

⁵ Cf., por exemplo, SOBER 1994 a, 1994b, 1994c, 1994d, 1994e, em que são colocadas, de um ponto de vista evolucionista, questões relativas a 'elementos da noção de racionalidade' tais como o egoísmo psicológico (o *egoísmo psicológico* é a ideia segundo a qual as pessoas agem em função do interesse próprio – a questão será retomada mais à frente), o solipsismo, as crenças inatas, dizer a verdade, mentir. Elliot Sober aborda desde estas características de um ponto de vista evolucionista, perguntando pelas razões para elas terem sido seleccionadas.

Assim, e no que diz respeito às questões da racionalidade teórica, Aristóteles pensou, de uma forma que podemos considerar inaugural, no que se entende por *ciência*, e que considerou ser o 'entendimento explicativo de alguma coisa', sua natureza e causas, envolvendo processos tais como definições, demonstrações e primeiros princípios. Aristóteles explicitou que falamos de 'definição' quando alguma coisa é predicável, dita de, alguma outra coisa essencialmente. Analisou a forma como demonstrações constituem provas científicas (este é o principal assunto dos *Analíticos Anteriores e Posteriores*). Foi ainda reconhecidamente o criador dessa ciência dos princípios explícitos do pensamento que é a lógica, que visa a forma como, certas afirmações sendo postas, outras afirmações validamente se seguem (a teoria do raciocínio silogístico aparece nomeadamente nos *Analíticos Anteriores*). Formulou as condições das várias modalidades de prova em campos diferentes de discurso, sugerindo, por exemplo (nos *Tópicos*), que argumentos dialécticos não são provas científicas no sentido estrito de 'demonstrações', as quais envolvem deduções a partir de primeiros princípios indemonstráveis (*Analíticos Posteriores*), mas são ainda assim um tipo de 'condução do espírito', visando a persuasão. Aristóteles desenvolveu também uma hipótese fundacionalista acerca da forma como alcançamos os referidos primeiros princípios de cada ciência, primeiros princípios estes que seriam auto-justificados e intuitivamente capturados. Podemos considerar que é no campo definido por Aristóteles que se situam ainda hoje as investigações sobre racionalidade teórica, investigações que, se seguirmos o espírito da abordagem de Aristóteles, devem ter uma palavra a dizer acerca de questões tais como saber a que chamamos 'ciência', de que é que a ciência é ciência, como é que a alcançamos, o que constitui prova em diversos domínios de pensamento⁶, quais são os bons métodos de pensamento e como podemos controlá-los.

⁶ Para uma análise das provas em Aristóteles, cf. GIL 1986: 31.

O que Aristóteles pensou sobre racionalidade prática pode ser encontrado sobretudo na *Ética a Nicómaco*, uma das mais importantes obras de ética de toda a história da filosofia. Nela são investigadas as finalidades da acção, a natureza do bem, da felicidade, da virtude e de virtudes (tais como a coragem, a generosidade, a temperança, a justiça), do prazer, da amizade, do amor e basicamente aquilo em que consiste a sabedoria prática, a sabedoria acerca da forma de sermos e agirmos, e aquilo em que consiste uma vida boa para seres humanos. Nesta obra Aristóteles preocupou-se também em esclarecer questões de teoria da acção, i. e., em analisar o âmbito em que, para seres como nós, existe espaço de manobra para uma escolha que pode ser racional, e em que faz sentido falar de **deliberação** e **decisão**: e esse é espaço que se situa entre o desejo e a acção, e em que está em causa uma concretização do possível. Aristóteles fez notar que não se delibera sobre 'coisas eternas' ou sobre aquilo que não depende de nós⁷ (por exemplo, sobre o resultado de uma demonstração matemática, sobre as letras do alfabeto ou sobre se vamos voar ou não) mas apenas sobre aquilo que está ao nosso alcance, que pode ser realizado e que depende de nós. Mais precisamente Aristóteles pensa que deliberamos sobre o futuro, sobre o que no futuro é indeterminado e depende de nós (*Ética a Nicómaco*, Livro III, 3). Neste contexto, Aristóteles define **decisão** como desejo deliberado de fazer alguma coisa que pertence ao leque de alternativas para o agente (por oposição a mero desejo, intenção ou opinião). Uma decisão ou desejo deliberado é um desejo formado após um processo de deliberação.

É frequente vermos Aristóteles apontado como um defensor da teoria instrumental da racionalidade na acção. A questão da teoria instrumental da racionalidade será retomada mais à frente mas pode desde já ficar claro que isso significa que para Aristóteles os fins da acção não são objecto de deliberação. A deliberação acontece quando os fins já estão afirmados pelo desejo: «o desejo

⁷ Para uma panorâmica da importância de Aristóteles no pensamento sobre a deliberação, cf. TUNHAS, *no prelo*, Deliberação.

relaciona-se com o fim, a escolha com os meios» (*Ética a Nicómaco*, Livro III, 2)⁸. Aristóteles é apontado como um defensor da teoria instrumental da racionalidade na acção tanto quanto defende que nós não deliberamos acerca de fins, mas a respeito de meios (*Ética a Nicómaco*, Livro III, 3): dando a finalidade por estabelecida, considera-se os meios de alcançá-la e se há vários meios, procura-se o mais simples e eficaz. Assim, «A escolha é um desejo deliberado de coisas que estão ao nosso alcance (...) após decidir em resultado de uma deliberação, desejamos de acordo com o que deliberamos» (*Ética a Nicómaco*, Livro III, 3)⁹.

Teoria da decisão racional

«O tema das nossa investigação é a lógica das crenças parciais, e não penso que possamos levá-la longe a não ser que tenhamos, pelo menos, uma noção aproximada do que é uma crença parcial, e de que maneira pode ser medida (se é que de todo o pode)»

F. Ramsey¹⁰

«Para fazer um uso prático destes conceitos temos que os pôr em contacto com a conversa comum acerca de necessidades, desejos, preferências, etc., um pouco como temos que estar preparados para reconhecer as cores, odores e sabores das coisas de forma a pormos a uso a química teórica»

R. Jeffrey¹¹

Aristóteles é um exemplo natural do interesse dos filósofos pela racionalidade, especialmente apropriado como referência para o tratamento do problema filosófico da racionalidade na me-

⁸ ARISTÓTELES, *Ética a Nicómaco*, Livro III, 2.

⁹ O filósofo americano Thomas Nagel (cf. NAGEL 1970, *The Possibility of Altruism*) retoma estes 'desejos deliberados' de Aristóteles na caracterização que faz do papel dos desejos na motivação racional.

¹⁰ RAMSEY 1926: 62.

¹¹ JEFFREY 1965: 215-216.

dida em que não apenas identifica e caracteriza os problemas da racionalidade teórica e da racionalidade prática como considera ainda a racionalidade peculiar ao homem, no sentido em que é o exercício da racionalidade que faz dos humanos humanos e não outra coisa qualquer. Mas a verdade é que os filósofos sempre procuraram, ao longo da história, caracterizar a mente (chamando-lhe alma, espírito, intelecto, etc.), os processos de pensamento, e a forma como a acção pode ser racional. Se quiséssemos restringir a nossa definição de racionalidade às inferências válidas – deixando por um momento de lado considerações comportamentais, ambientais e evidenciais – veríamos que os lógicos, por seu lado, têm vindo a desenvolver, desde há mais de um século, sistemas formais para capturar a natureza dessas inferências válidas, e veríamos a história da lógica como uma componente importante da história das investigações recentes sobre racionalidade. No entanto, mais do que a filosofia ou a lógica – e já vimos na Introdução o quanto a ligação directa entre lógica e racionalidade pode ser problemática – foi o desenvolvimento da **teoria da decisão racional** no século XX que instituiu a **visão normativa dominante**, ou pelo menos a **referência central dos estudos da racionalidade**. Como foi já dito, a teoria da decisão é uma teoria abstracta e normativa da acção racional e não uma teoria descritiva do comportamento. A teoria da decisão é uma teoria de agentes que deliberam, i. e., de agentes que a partir das suas crenças e preferências, avaliam cursos de acção em termos das consequências, nomeadamente em circunstâncias que não podem prever nem controlar totalmente, de forma a escolher racionalmente um curso de acção. Um agente racional é um agente que, tendo em consideração a **utilidade** ('desejabilidade', no termo de R. Jeffrey¹²) dos estados do mundo que podem ser consequência das suas acções e a probabilidade (um número entre 0 e 1) da obtenção desses estados, age de modo a **maximizar a utilidade esperada** (utilidade esperada = utilida-

¹² JEFFREY 1965.

de do resultado X probabilidade da ocorrência deste). O que a 'moldura teórica' da teoria da decisão nos dá são então agentes, estruturas (*rankings*) de preferências, situações ou contextos de escolha, em que existem alternativas, sendo que esses agentes devem, em função daquilo que os caracteriza, optar pelo melhor, em termos de resultados da acção, sendo estes pensados em termos de desejabilidade e probabilidades. À partida nada impede a teoria da decisão racional de pretender ser de alguma forma neutra relativamente a quaisquer suposições psicológicas e filosóficas acerca da mente de agentes particulares: aquilo que ela visa são agentes racionais em geral.

A teoria da decisão, ou teoria da escolha racional, foi axiomatizada há mais de 50 anos no âmbito da economia. Constituem referência as obras de J. Von Neumann e O. Morgenstern, *Theory of Games and Economic Behavior* (1947) e de Leonard Savage, *Foundations of Statistics* (1954)¹³. No entanto, Frank P. Ramsey (1903-1930) matemático e filósofo de Cambridge, tinha já anteriormente elaborado uma teoria das crenças parciais e do raciocínio inconclusivo, baseada numa teoria subjectivista das probabilidades (Ramsey 1926, *Truth and Probability*), e os matemáticos B. Pascal (1623-1662) e D. Bernoulli tinham igualmente e muito antes procurado conceber a relação de agentes com bens ou valores em situações de incerteza. Foi de resto, em termos históricos, a partir da teoria das probabilidades que foi abordada e analisada a 'lógica' da escolha em situações em que os agentes não estão, nem podem estar, absolutamente certos daquilo que decorrerá das suas opções de acção (é por isso esclarecedor ter em mente a forma como a teoria das probabilidades se ligou com a avaliação de apostas por um apostador racional). Mas voltemos um pouco atrás para pensar em termos e formulações. Como Aristóteles notou, só faz sentido falar da

¹³ Parte do que descrevo a seguir baseia-se em JEFFREY 1965, *The Logic of Decision*. A obra constitui uma referência clássica quanto aos fundamentos e implicações filosóficas da teoria da decisão, e deve ser consultada directamente como iniciação à teoria da decisão.

escolha de um agente por entre opções quando se considera coisas que de facto podem ser escolhidas: não é o caso, para os humanos, que se possa escolher voar, ou ser uma alma que salta de corpo em corpo a cada cinco minutos. Uma **opção** é uma acção aberta ao agente. Podem existir opções em situação de certeza e de incerteza. Normalmente distingue-se a posição de agentes perante contextos de decisão em que há **certeza** (caso em que os itens da escolha estão definidos, por exemplo, a lista dos pratos num menu de restaurante) e contextos de **incerteza** (casos em que há acaso envolvido nos itens a escolher, com ou sem probabilidades dadas – por exemplo, apostas e seguros). Estas escolhas são arriscadas de uma forma que as primeiras não o são, e coloca-se então a questão da atitude do agente perante o **risco** (além de desejabilidade e probabilidade, está em causa a aversão ao risco ou procura do risco). Os resultados de determinada opção de um leque de opções são neste caso apenas prováveis e têm determinado valor, e valores diferentes. A racionalidade do agente consistirá em optar pelo curso de acção que lhe trará o melhor resultado possível. **Escolha** é decisão do agente por entre opções, em princípio em função de uma deliberação ou ponderação. O caso clássico é, como fiz notar, o de uma aposta. Cada resultado (*outcome*) de uma opção traz um benefício (monetário) esperado. O valor de uma opção é o benefício esperado para o agente. A Teoria da decisão afirma que um agente racional escolhe de modo a maximizar o benefício esperado: aqui isso significa claramente escolher de modo a ganhar o mais possível.

A grande vantagem, para a concepção do que está em causa em situações de escolha com risco, ao considerar como ‘casos-teste’ as apostas monetárias é o facto de nesse caso as utilidades (valor para o agente) e as probabilidades numéricas se encontrarem facilmente – faz sentido pensar que são quantificáveis de forma precisa. Ora, considerando em geral as situações de escolha (por exemplo, a situação que Maria enfrenta ao ter que decidir casar com João, ou Manuel ou Tiago) é claro que isto raramente acontece. Será que em situações reais os agentes

atribuem e escolhem em função de utilidades e probabilidades? Como devemos pensar no que se passa num agente quando este delibera? Quer Aristóteles quer a teoria da decisão racional visam a questão, Aristóteles de forma descritiva, pondo em relevo as componentes conceptuais da situação, a teoria da decisão de forma quantificada, assumindo que os agentes racionais escolhem em função de utilidades e probabilidades. Deliberar é avaliar diferentes linhas de acção disponíveis (opções) em função das consequências destas, que podem depender de circunstâncias que o agente não prevê nem controla – é isso que se faz em função de utilidades e probabilidades. O princípio de racionalidade que daí ressalta é maximização da utilidade esperada.

Olhemos para um exemplo muito simples, utilizado por R. Jeffrey em *The Logic of Decision*: alguém foi convidado para jantar e ficou encarregado de levar o vinho. Esqueceu-se se ia comer frango ou bife. O vinho branco seria apropriado para frango, o vinho tinto para bife mas ele só pode levar um garrafa. Não pode telefonar nem de alguma outra forma contactar quem o convidou. Qual vinho deve levar? Que tipo de considerações entram nesta deliberação?

Uma primeira consideração diz respeito a combinações: as linhas da matriz representam actos possíveis (levar vinho branco, levar vinho tinto), as colunas representam condições possíveis que podem influenciar o resultado dos actos (o jantar consistir em frango, o jantar consistir em bife)

	Frango	Bife
Vinho branco	Vinho branco + frango	Vinho branco + bife
Vinho tinto	Vinho tinto + frango	Vinho tinto + bife

Matriz de desejabilidade:

0	-1
0	1

Matriz de probabilidade (considera-se igualmente provável cada uma das possibilidades de menu):

0.5	0.5
0.5	0.5

A partir daqui é possível estimar a deseabilidade de cada acto multiplicando as entradas correspondentes das matrizes de deseabilidade e probabilidade:

$(0.5) \times 1 = 0.5$	$(0.5) \times (-1) = (-0.5)$
$(0.5) \times 0 = 0$	$(0.5) \times 1 = 0.5$

É possível agora, somando as linhas, comparar a deseabilidade de levar vinho branco $(0.5) + (-0.5) = 0$ e de levar vinho tinto $(0 + 0.5 = 0.5)$. O convidado conclui por isso pela maior deseabilidade de levar vinho tinto e é isso que faz. O que aconteceu neste processo foi que o decisor estabeleceu um *ranking de preferências* relativamente aos actos disponíveis, ranking no qual a ordem de preferências é identificada com a ordem numérica das deseabilidades estimadas. O 'racionalmente recomendável' é que o acto que se posiciona 'no topo' da ordenação seja o praticado. Claro que fica o problema de saber como se descobre as deseabilidades e probabilidades de que o agente necessita como dados para a deliberação – o caso em análise era um caso especialmente simples, em que fazia sentido pensar que se tinha esses dados na mão. Mas tudo o que se 'vê' quando um agente age, opta, é a sua escolha por entre opções, o seu comportamento de escolha – como se poderá descobrir as crenças e desejos não observáveis, que em princípio subjazem ao comportamento de escolha?

É neste ponto que Ramsey tem uma sugestão, no artigo que referi, *Truth and Probability*, uma sugestão que pode parecer, à primeira vista, irrealista e formalista. Ele sugere que devemos assumir utilidades e probabilidades numéricas precisas quando

um agente delibera e escolhe em situações de incerteza e risco e dá-nos um método para chegar a essas utilidades e probabilidades. A proposta de Ramsey tem a ver com a possibilidade de medirmos a força das nossas crenças (a questão é tratada no Ponto 3 de *Truth and Probability*, sobre crenças parciais). A ideia de Ramsey é identificar o grau ou força de uma crença com uma propriedade causal dessa crença, que podemos exprimir como 'o ponto até ao qual estaríamos preparados para agir com base nela'. Esta ideia de Ramsey, de resto influenciada pelo filósofo e lógico americano C. S. Peirce (1839-1914), é muito importante para compreender, de um ponto de vista histórico, os tratamentos quantitativos da noção de escolha que se tornaram usuais em várias disciplinas (economia, estatística) e não apenas na filosofia da acção.

Para termos um método para medir crenças como base de acções possíveis e para obtermos, no fundo, uma explicitação das utilidades e probabilidades implícitas nas nossas preferências, Ramsey sugere o seguinte. Imaginemos que há um resultado da acção, uma possibilidade específica, que o agente prefere a todos os outros e uma outra perspectiva de resultado, que seria 'pior do que qualquer outro'. Atribua-se-lhes então valores, por exemplo, 100 e 0. Aquilo que Ramsey faz a seguir é sugerir uma forma de localizar um ponto intermédio (50), um resultado de valor intermédio, no 'espectro' de preferências, e depois sucessivamente valores intermédios, até se ter um contínuo de 'utilidades quantificadas'. Cada resultado possível da acção poderá agora ser aí situado. Ramsey mostra que assumindo alguma consistência, as probabilidades se seguem. Tem-se assim então a medida de valores de que necessita a teoria da decisão.

A teoria da decisão, que podemos encarar como uma 'lógica de escolher sozinho', é prolongada na teoria dos jogos, que é uma 'lógica da escolha em situações em que outros agentes vão estar a escolher também e portanto têm que ser considerados como factores da situação'. Além do risco (factor que aparece quando há incerteza no contexto de escolha), entra aqui a questão da

estratégia, importante na tomada de decisão quando as escolhas de mais do que um agente são interdependentes. A teoria dos jogos é a teoria das decisões a tomar num contexto de interacção com um adversário racional. Evidentemente, quando um agente interpreta o comportamento de outro agente, quer chegar a interpretação correcta.

O **Dilema do Prisioneiro** é a situação clássica da teoria dos jogos, uma situação-teste, esquemática, que é muito importante para a teoria social e política, na medida em que serve como ‘análise concentrada’ do que está em jogo na interacção social e cooperação ou não cooperação de agentes (servindo como moldura de referência para analisar situações recorrentes na interacção de agentes tais como interacções não cooperantes e ‘corridas ao armamento’).

A situação é a seguinte: dois criminosos foram presos por terem cometido um crime juntos. Cada um está numa cela separada sem qualquer contacto com o outro. A polícia oferece-lhes as seguintes opções: se um confessar e o outro não confessar, o primeiro pode sair em liberdade e o segundo tem doze anos de prisão. Se ambos confessarem cada um tem dez anos de prisão. Se nenhum confessar cada um cumpre dois anos de prisão. O problema neste caso é que para cada um dos prisioneiros individualmente faz mais sentido confessar (eles pensam assim: se o outro confessou então mais vale eu confessar senão fico com doze anos de prisão; por outro lado, se o outro não confessou mais vale eu confessar e vou em liberdade). Assim, do ponto de vista individual faz sempre mais sentido confessar. No entanto, se ambos pensarem assim e ambos confessarem acabam por ficar dez anos na prisão enquanto que se ambos não confessassem ficariam com dois anos cada. Desta forma, o que parece fazer mais sentido do ponto de vista individual acaba por prejudicar ambos os prisioneiros.

Uma grande quantidade da literatura dedicada a questões da racionalidade prática gira em torno do dilema do prisioneiro. Retomarei a questão quando falar da teoria da racionalidade simbólica proposta por Robert Nozick, no último capítulo.

É muito importante para as questões dos ‘cânones’ e dos ‘modelos’ da racionalidade, referidas na Introdução (e a teoria da decisão é – recorde-se – um dos candidatos a ‘cânone’, pronta a ser utilizada como ‘modelo’), e também para a questão dos desejos e da motivação para a acção, avaliar a pretensão da teoria da decisão, ao ser uma teoria normativa e abstracta, de ser de alguma forma neutra (neutra face à natureza da mente dos agentes e face a caracterizações motivacionais, embora se diga alguma coisa acerca de fins, representações e cálculo, nomeadamente quando se evoca *ranking* de preferências, utilidades, probabilidades e maximização). A teoria da decisão é, admitidamente, uma caracterização formal, abstracta, feita ‘de cima para baixo’ e não ‘de baixo para cima’ (i. e., a partir de uma teoria dos substratos neuronais ou de uma teoria dos processos cognitivos do mental) das escolhas racionais de agentes. Por tudo isto o seu estatuto como modelo do comportamento de agentes reais é problemático. Uma possibilidade é considerá-la irrelevante na explicação psicológica – mas uma tal opção passa por cima da possibilidade de qualquer ‘explicação psicológica’ incluir uma componentes normativa irreduzível.

*Poderá a racionalidade ser empiricamente refutada?
Psicologia cognitiva: juízo e tomada de decisão*

Se queremos um mapa conceptual dos estudos da racionalidade importa ter não apenas uma noção da importância histórica que a lógica, a teoria das probabilidades e a teoria da decisão tiveram e têm – ao oferecerem princípios explícitos de inferência e decisão, princípios de racionalidade, se se entende esta como uma questão abstracta de correcção no pensamento e na acção – como também ter uma noção do estado actual das investigações psicológicas (do âmbito da psicologia cognitiva e da psicologia evolutiva) e uma perspectiva das interpretações filosóficas a que estas têm conduzido. Parece-me que poderíamos

fazer o ponto da situação desta forma: algumas décadas após os estudos pioneiros de Amos Tversky, Daniel Kahneman e outros psicólogos do raciocínio e da decisão cujas descobertas levantaram problemas à ideia tradicional dos seres humanos como 'animais intrinsecamente racionais', há hoje uma nova perspectiva em campo, a perspectiva darwinista da psicologia evolutiva¹⁴, que vem desafiar as interpretações pessimistas que pareciam impor-se. Antes de avançar para alguns dados relativos aos programas de investigação referidos, queria propor que se considere, de modo a manter as ideias claras, que as investigações em psicologia estudam experimentalmente o raciocínio e decisão de agentes, mais do que propriamente a 'racionalidade'. Feita essa proposta, o que encontramos nos estudos psicológicos sobre raciocínio e decisão? Basicamente, encontramos dados – o grande interesse das investigações psicológicas nesta constelação consiste no seguinte: se podemos à partida contar com uma postura normativa ou idealizadora em áreas como a lógica ou a teoria da decisão, é por excelência na psicologia que encontramos um interesse directo pela forma como se movem de facto os agentes cognitivos em raciocínios e decisões, e pela recolha de dados. A psicologia é por isso o ponto de referência da interpretação descritiva da ideia de uma 'teoria da racionalidade'

Uma vez que este livro não está escrito do ponto de vista da psicologia, aquilo que vou procurar fazer é dar as referências centrais e alguns exemplos das investigações empíricas da racionalidade¹⁵. Referi na Introdução (Ponto 3) um importante paradigma de inves-

¹⁴ Dois dos nomes mais salientes são os de Leda Cosmides (psicologia) e John Tooby (antropologia) (Center for Evolutionary Psychology, University of California – Santa Barbara). Para uma declaração dos princípios da investigação aí conduzida, cf. <http://www.psych.ucsb.edu/research/cep/primer.html>

¹⁵ Para uma introdução à psicologia do raciocínio e para a descrição das experiências e resultados dos estudos referidos, ver, por exemplo, NISBETT & ROSS 1980, KAHNEMAN, SLOVIC & TVERSKY 1982, BARON 1988, DAWES 1988 e PIATELLI PALMARINI 1994.

tigação em psicologia do raciocínio e da decisão, conhecido por *heuristics and biases*. Recordo o que disse então: a investigação psicológica sobre juízos visa os processos utilizados por sujeitos para tirar conclusões a partir do conhecimento e evidência disponíveis, a investigação sobre tomada de decisão visa a escolha entre opções. A ideia central destas investigações é que a maior parte das pessoas não tem, ou pelo menos parece não utilizar, princípios correctos de raciocínio e decisão. Pelo contrário, os juízos e as decisões dos agentes reais processam-se com base em princípios mais simples, chamados **heurísticas**. Algumas heurísticas do domínio dos juízos são, por exemplo, julgar pela representatividade (é o caso do famoso exemplo de Linda, descrito em seguida), pela disponibilidade ou por ancoragem e ajustamento. No domínio da tomada de decisão, são comuns, por exemplo, efeitos de *framing* ou enquadramento, que ocorre quando o que é essencialmente o 'mesmo problema de decisão' dá lugar a escolhas diferentes, em função da formulação do problema perante os sujeitos, que se comportam assim como se não tivessem preferências determinadas fixas. Estes estudos, e é esse o núcleo do seu interesse se os considerarmos de um ponto de vista filosófico, redundam na 'descoberta' de princípios psicológicos (que 'explicam' descobertas empíricas acerca da forma como as pessoas inferem e decidem) que são contra-intuitivos e incompatíveis com análises normativas da racionalidade. Uma forma de interpretar, por exemplo, os estudos empíricos da tomada de decisão é considerar que, ao contrário do que a teoria da decisão supõe, as pessoas não tem sempre preferências bem ordenadas, antes abordam os problemas de decisão como problemas a resolver e moldam ou determinam as suas preferências em grande medida em função do contexto de decisão.

O que a investigação inicial de Kahneman e Tversky revelou foi precisamente que a necessidade de emitir juízos em condições de incerteza faz com que os sujeitos 'convoquem' heurísticas determinadas, que produzem distorções ou enviesamentos (*biases*). Por exemplo, as pessoas esperam que observações

futuras sejam como as observações passadas, mesmo baseando-se num pequeno número de observações – aplicam uma ‘lei dos pequenos números’, com algumas propriedades da lei estatística dos grandes números (de acordo com a qual amostras de grande dimensão serão representativas do universo do qual são recolhidas), mas que não é sensível à dimensão da amostra (cf. Tversky e Kahneman, 1971, *Belief in the law of small numbers*). Os sujeitos dos estudos referidos foram cientistas, e os estudos mostrariam assim que eles sobrestimavam habitualmente a probabilidade de amostras pequenas confirmarem hipóteses investigadas. A tese de Tversky e Kahneman era, de resto, que as pessoas, leigos e cientistas, têm intuições acerca de amostragem, que essas intuições estão erradas, mas as pessoas agem sistematicamente em função delas. Em 1972, em *Subjective Probability: a judgment of representativeness*, Kahneman e Tversky sugeriram que a heurística em causa no exemplo anterior era apenas um caso daquilo a que chamaram em geral heurísticas da representatividade (*representativeness*), pelas quais os sujeitos avaliam a probabilidade de um evento levando em conta a forma como este (supostamente) captura as propriedades relevantes de processos e situações, o que frequentemente produz distorções (o caso de Linda é uma delas). Outras heurísticas estudadas foram a disponibilidade (*availability*) – pela qual eventos são julgados tanto mais prováveis quanto se consegue recordar exemplos ou imaginá-los acontecendo – e a ancoragem e ajustamento (*anchoring and adjustment*) – pela qual, por exemplo, quantidades são estimadas pensando quão maiores ou menores elas serão relativamente a um determinado valor de referência (tipicamente as pessoas ajustam pouco, deixando o valor ‘ancorado’ ao valor de referência, mesmo que este seja totalmente arbitrário). Como também já afirmei atrás, uma possível razão para o grande impacto deste tipo de estudos sobre heurísticas foi o facto de eles tornarem a noção de racionalidade limitada (*bounded rationality*) avançada por Herbert Simon, concreta e experimentalmente manejável, identificando os processos que os agentes utilizam de facto para obviar à finitude dos recursos disponíveis.

Passo a descrever o exaustivamente estudado caso de Linda, inicialmente proposto por Tversky e Kahneman (Tversky & Kahneman 1982b, *Judgments of and by representativeness*). A melhor forma de olhar para o caso é vê-lo como exemplificando a forma como a heurística da representatividade funciona muito frequentemente de modo a produzir a chamada falácia da conjunção. O caso diz respeito a raciocínio probabilístico e a experiência consiste basicamente no seguinte: é pedido a sujeitos que elaborem uma lista ordenada de situações, de acordo com as probabilidades atribuídas à ocorrência destas, a partir das descrições fornecidas. Verifica-se que é muito comum que os sujeitos considerem como mais prováveis conjunções do que as frases componentes dessas conjunções. Por exemplo, é pedido aos sujeitos, após ser fornecida uma descrição de Linda como tendo 31 anos, sendo solteira, frontal e muito inteligente, licenciada em filosofia (*‘31 years old, single, outspoken and very bright. She majored in philosophy’*) que avaliem a probabilidade de várias situações descritas acerca de Linda. Sistemáticamente acontece que os sujeitos consideram a afirmação ‘Linda é bancária e activa no movimento feminista’ como mais provável do que a afirmação ‘Linda é bancária’. Ora, obviamente ninguém pode ser ‘bancária e activista no movimento feminista’ sem ser simplesmente ‘bancária’. O raciocínio comum dos sujeitos é uma violação tão flagrante da teoria das probabilidades, que o facto de ocorrer sistematicamente quer em sujeitos não treinados em estatística quer em sujeitos treinados, se torna interessante. Descrevo em seguida de forma mais completa dois exemplos de casos deste tipo, começando por uma descrição mais completa do próprio caso de Linda.

1) Descrição fornecida aos sujeitos:

‘Linda tem 31 anos, é solteira, tem uma forte personalidade e é muito inteligente. Licenciou-se em filosofia. Enquanto estudante interessou-se profundamente por assuntos como a discriminação e a justiça social e também participou em manifestações anti-nuclear’.

Pedido que é feito aos sujeitos:

Ordene as seguintes afirmações de acordo com a sua probabilidade, atribuindo o número 1 à mais provável e 8 à menos provável.

- a) Linda é professora primária.
- b) Linda trabalha numa livraria e tem aulas de yoga.
- c) Linda participa activamente no movimento feminista.
- d) Linda trabalha como psiquiatra na assistência social.
- e) Linda é membro da Liga da Mulheres Votantes.
- f) Linda é bancária.
- g) Linda é agente de seguros.
- h) Linda é bancária e participa activamente no movimento feminista.

Num grupo sem qualquer formação em probabilidade e estatística, 89% dos inquiridos indicaram que a afirmação (h) era mais provável que a afirmação (f). O problema é que quando a mesma questão é apresentada a sujeitos com preparação académica em estatística – por exemplo, estudantes de pós-graduação do programa de ciência da decisão da Stanford Business School – uma percentagem muito próxima de sujeitos (85%) tem a mesma opinião.

Um outro objecto de estudo tem sido a chamada **negligência das probabilidades prévias**. Em termos bayesianos, a probabilidade de uma hipótese sobre um determinado conjunto de dados depende, em parte, das probabilidades prévias. Contudo, Kahneman e Tversky levaram a cabo estudos empíricos que pelo menos aparentemente demonstram que os sujeitos muitas vezes ignoram simplesmente as probabilidades prévias.

Exemplos de experiências:

1) Informação fornecida aos sujeitos:

Ao primeiro grupo de sujeitos: Um painel de psicólogos entrevistou e administrou testes de personalidade a 30 engenheiros e a 70 advogados, todos bem sucedidos nas respectivas áreas. Com base nesta informação foram escritas breves descrições dos 30 engenheiros e dos 70 advogados. Nos seus formulários encontrará

cinco descrições, escolhidas aleatoriamente das 100 descrições disponíveis. Indique, por favor, para cada descrição e numa escala de 0 a 100, a probabilidade de a pessoa descrita ser um engenheiro.

Ao segundo grupo de sujeitos é apresentado o mesmo texto, mas é-lhes dito que o teste de personalidade tinha sido administrado a 70 engenheiros e a 30 advogados.

Algumas das descrições fornecidas aos sujeitos correspondem ao 'estereótipo do engenheiro' outras ao 'estereótipo' do advogado, outras são concebidas como neutras, i. e., como não oferecendo aos inquiridos qualquer informação que facilite a tomada de decisão.

Eis dois exemplos dessas descrições (a primeira concebida para se assemelhar a um engenheiro, a segunda para ser neutra):

Jack é um homem de 45 anos. É casado e tem quatro filhos. É de um modo geral conservador, cuidadoso e ambicioso. Não demonstra qualquer interesse em assuntos políticos ou sociais e gasta a maior parte dos seus tempos livres com os seus muitos passatempos, que incluem carpintaria, vela e puzzles matemáticos.

Dick tem 30 anos. É casado e não tem filhos. É um homem competente e motivado, que promete vir a ser muito bem sucedido no seu ramo. É muito popular entre os colegas.

Como era de esperar, os inquiridos de ambos os grupos consideraram que a probabilidade de Jack ser engenheiro era bastante alta. Além disso, e no que parece ser uma clara violação dos princípios bayesianos, as diferenças de estruturas dos universos não tiveram influência alguma. A negligência foi ainda mais evidente no caso de Dick. Esta descrição foi formulada de forma a não dar qualquer informação acerca da profissão de Dick. Como tal, a única informação útil que os inquiridos possuíam, era a informação acerca de probabilidades prévias fornecida no texto. Porém, essa informação foi totalmente ignorada. A probabilidade média estimada por ambos os grupos de inquiridos foi de 50%.

Vejam os entanto outra experiência que nos mostra que os sujeitos de Kahneman e Tversky não eram completamente insensíveis à informação de partida.

Instruções: Suponha agora que não lhe é dada nenhuma informação acerca de um determinado indivíduo escolhido da amostra ao acaso. Preencha os espaços.

A probabilidade deste homem ser um dos 30 engenheiros [ou, para o outro grupo de inquiridos: um dos 70 engenheiros] na amostra de 100 é de ___%.

Neste caso os inquiridos confiaram inteiramente na informação que possuíam previamente; a estimativa média era de 30% para o primeiro grupo de inquiridos e de 70% para o segundo.

Os psicólogos R. Nisbett e L. Ross¹⁶ discutem estas experiências e propõem a seguinte interpretação: quando não se fornece aos sujeitos nenhum dado concreto acerca do caso a estudar, as probabilidades prévias são utilizadas adequadamente; quando os dados fornecidos não têm qualquer valor específico, as probabilidades prévias podem ser, em grande medida, ignoradas e as pessoas reagem como se não houvesse qualquer base que lhes permitisse presumir certas diferenças onde encontram algumas verosimilhanças. A capacidade das pessoas compreenderem a relevância da informação de partida deve ser muito fraca, pois, uma vez expostas a informações inúteis muito facilmente deixam de ter essa informação em consideração.

Apresento em seguida a descrição de um caso muito citado que mostra de que modo o fenómeno pode ter sérias consequências práticas. O caso é um problema que Casscells, Schoenberger e Grayboys (1978) apresentaram a um grupo de docentes e alunos do quarto ano da Harvard Medical School.

Se um teste para detectar uma doença, cuja incidência é de 1/1000, tem uma taxa de falsos positivos de 5%, qual é a probabilidade de uma pessoa que apresente um resultado positivo no teste

¹⁶ NISBETT & ROSS 1980:145-6.

ter de facto a doença, partindo do princípio que você não sabe nada acerca dos sintomas ou da história clínica dessa pessoa? ___%

Segundo a interpretação mais credível do problema, a resposta correcta é 2%¹⁷ – no entanto, apenas dezoito por cento das pessoas do auditório de Harvard deram uma resposta próxima dos 2%. Quarenta e cinco por cento das pessoas deste distinto grupo ignoraram completamente a informação base e responderam 95%.

Passemos agora a alguns exemplo relativo à tomada de decisão e aos chamados efeitos de *framing* ou enquadramento. Temos um mesmo problema e duas versões.

Problema A – Versão 1 (inquiridos: 126)

Considere que está \$300 mais rico do que hoje. Tem que escolher entre: (i) um ganho seguro de \$100, (ii) uma probabilidade de 50% de ganhar \$200 e 50% de não ganhar nada

Resultado: 72% dos sujeitos escolhem (i), 28% escolhem (ii)

Problema B – Versão 2 (inquiridos: 128)

Assuma que está \$500 mais rico do que hoje. Tem que escolher entre: (i) uma perda certa de \$100, (ii) 50% de probabilidade de não perder nada e 50% de perder \$200

Resultado: 36% dos sujeitos escolhem (i), 64% escolhem (ii)

O problema é que o problema é essencialmente o mesmo: quando os \$300 ou \$500 iniciais são adicionados aos resultados, ambos os problemas representam uma escolha entre \$400 seguros *versus* uma hipótese equiprovável de \$300 ou \$500. Simplesmente, um dos problemas está formulado como uma escolha entre

¹⁷ A informação de que se dispõe à partida permite-nos saber que 999 pessoas em cada 1000 não sofrem da doença. No entanto, quando testadas, 50 pessoas em cada 1000 (5%), teriam como resultado um falso positivo. Portanto, 50 vezes mais pessoas têm um resultado falso positivo do que verdadeiro positivo (1 pessoa). Portanto, há apenas 2% de probabilidades de uma pessoa cujo teste foi positivo de facto ter a doença.

ganhos, e o outro como uma escolha entre perdas. No primeiro caso as pessoas tendem a evitar o risco, no segundo a procurar o risco (e estes 'princípios' são regulares).

Um outro caso é a tarefa da selecção (*selection task*). A tarefa da selecção, que foi proposta pela primeira vez por Peter Wason em 1966 e que é um dos casos mais estudados na psicologia do raciocínio, testa em situações variadas, o raciocínio com condicionais¹⁸. A tarefa de selecção consiste basicamente numa situação em que sujeitos estão perante objectos de escolha (por exemplo, quatro cartas em cima de uma mesa) e um princípio geral é fornecido (por exemplo: '*Se uma carta tem uma vogal de um dos lados, então tem um número ímpar no outro lado*'). As faces expostas das cartas revelam, por exemplo, E, C, 5, 4. É proposta uma tarefa prática aos sujeitos: Indique que cartas têm que ser verificadas para determinar se o princípio é verdadeiro. A tarefa não é uma tarefa pura de raciocínio dedutivo: podemos olhar para ela como o teste de uma hipótese usando um condicional como regra. Uma percentagem significativa de sujeitos falha. Mas o que significa falhar? As interpretações iniciais foram no sentido de que os sujeitos têm, neste tipo de situações, tendência para confirmar (por oposição a infirmar) hipóteses.

Olhemos para o exemplo: mostra-se aos sujeitos quatro cartas. Elas têm números ou letras na face que está à vista. Diz-se-lhes que cada carta tem um número numa das faces e uma letra na outra. E que eles têm que dizer que cartas têm que ser voltadas para testar a seguinte regras: Se uma carta tem uma vogal numa das faces, então há um número ímpar na outra face.

E C 5 4

¹⁸ Em EVANS, NEWSTEAD & BYRNE 1993 a tarefa de selecção é mesmo referida como sendo pura e simplesmente 'o problema mais intensivamente estudado na história da psicologia do raciocínio' (EVANS, NEWSTEAD & BYRNE 1993: 99).

Aquilo que Wason e muitos outros investigadores depois dele descobriram foi, por exemplo, que a maior parte das pessoas diz que é a carta E que deve ser virado, mas também acha que a carta 5 que deve ser virada. Ora, a carta 5 não pode falsificar a afirmação de partida independentemente daquilo que esteja do outro lado. Por outro lado, uma maioria dos sujeitos pensa que a carta 4 não precisa de ser virada. No entanto, se a carta 4 não for virada não se poderá saber se tem uma vogal do outro lado. E se tiver a afirmação de partida será falsificada.

Os resultados mudam muito, como se verá à frente, em tarefas de selecção 'concretizadas', em que não há números e letras mas situações concretas e consequências.

Um outro conjunto de fenómenos investigado por psicólogos do raciocínio tem a ver com o grau de confiança que as pessoas têm nas suas próprias respostas a perguntas concretas.

Dois exemplos:

Considere os seguintes pares de cidades.

Qual delas tem mais habitantes?

- | | |
|---------------|----------------|
| (a) Las Vegas | (b) Miami |
| (a) Sydney | (b) Melbourne |
| (a) Hyderabad | (b) Islamabad |
| (a) Bona | (b) Heidelberg |

Considere os seguintes pares de factos históricos.

Qual deles aconteceu primeiro?

- | | |
|---------------------------------|------------------------------------|
| (a) A assinatura da Magna Carta | (b) O nascimento de Maomé |
| (a) A morte de Napoleão | (b) A compra da Louisiana |
| (a) O assassinato de Lincoln | (b) O nascimento da Rainha Vitória |

Os sujeitos respondem, e em seguida pergunta-se-lhes, apresentando como alternativa percentagens de 50% 60% 70% 80% 90% 100%: Que confiança tem nas respostas que deu? O que se verifica frequentemente é que o grau de confiança dos sujeitos nas suas respostas é sistematicamente superior à correcção das

mesmas (em casos em que os inquiridos dizem estar 100% confiantes, apenas 80% das respostas estão correctas; em casos em que dizem estar 90% confiantes, apenas cerca de 70% das suas respostas estão correctas; em casos em que dizem estar 80% confiantes, apenas cerca de 60% das suas respostas estão correctas, etc.). Parece verificar-se algo como uma 'tendência para a confiança excessiva'. Aparentemente, avisar os sujeitos de que as pessoas são frequentemente demasiado confiantes ou oferecer recompensas pela exactidão não tem efeito significativo. O fenómeno já foi observado numa grande variedade de sujeitos, desde estudantes universitários a médicos e até mesmo analistas da CIA¹⁹.

Como interpretar os resultados da psicologia cognitiva?

Apresentei alguns exemplos de estudos empíricos acerca de raciocínio e decisão. O problema que estudos deste género colocam de um ponto de vista filosófico é o seguinte: será que tais estudos provam a existência de deficiências na nossa capacidade de raciocínio e decisão? Deveremos tirar conclusões pessimistas? Deveremos concluir que Aristóteles estava profundamente errado quando definiu o homem como 'animal racional' e que as propostas abstractas de lógicos, teóricos da decisão e teóricos das probabilidades não têm nada a ver com a forma como as pessoas de facto raciocinam e decidem? De acordo com algumas das interpretações avançadas quanto a estes resultados, dado o seu carácter sistemático, não estaríamos aqui perante meros erros de performance. O que é aqui mostrado é que a capacidade de raciocínio da maior parte das pessoas é 'irracional', i. e., a própria **competência** é normativamente problemática. Introduzi,

¹⁹ Para uma revisão da bibliografia ver LICHTENSTEIN, FISCHOFF & PHILLIPS 1982.

nesta última formulação, um conceito decisivo na discussão, o conceito de competência. De facto, os conceitos de competência e performance, introduzidos por Chomsky na ciência cognitiva por via da linguística foram recrutados por algumas pessoas para analisar e interpretar os resultados da psicologia do raciocínio e decisão. Procurarei em seguida explicar como.

O ponto importante é que podemos perfeitamente imaginar que uma dada performance do sujeito ou agente cognitivo pode ser desviante ou faltosa, sem que a competência – uma espécie de correcção em potência – deixe de estar presente. Ora, assim como Chomsky distingue a competência linguística de um falante de uma língua natural (que diria respeito a regras e princípios de uma gramática interna, mentalmente representada) das performances deste (contam como performance a produção de frases, os juízos de gramaticalidade) poder-se-ia avançar uma distinção análoga para nos permitir pensar não na forma como as pessoas falam mas na forma como elas raciocinam e decidem. Teríamos então uma distinção entre *competência de raciocínio e decisão* e *performance de raciocínio e decisão*.

Notemos antes de mais que não é difícil encontrar analogias entre o comportamento linguístico das pessoas e o comportamento de raciocínio e decisão: em ambos os casos (i) estamos perante processamento espontâneo e amplamente inconsciente de uma classe aberta de inputs; (ii) em ambos os casos, estamos perante sujeitos capazes de fazer juízos intuitivos de 'validade e correcção' (em termos de linguagem, trata-se de juízos de gramaticalidade, ambiguidade, etc., em termos de raciocínio trata-se de juízos de validade, probabilidade, etc.). Assim como a competência linguística seria a superfície de um mecanismo simbólico interno, não acessível à consciência do falante, existiria uma competência de raciocínio e decisão, não directamente acessível à própria pessoa e no entanto presente. O ponto do conceito de performance neste contexto é obviamente permitir-nos admitir que em ambos os casos podem ocorrer 'interferências' relativas ao funcionamento global do sistema – devidas

a atenção, motivação, memória a curto prazo, etc. Tendo esta origem e este estatuto, os erros de performance dos agentes não lançariam sombra sobre a competência subjacente.

Apesar das semelhanças que tornam pertinente o uso das noções de competência e performance para falar de raciocínio e decisão não podemos esquecer uma diferença importante: duas pessoas que falam duas línguas naturais diferentes têm competências linguísticas diferentes, no sentido em que as regras e princípios que supostamente constituem a competência linguística de um falante do português são consideravelmente diferentes das regras e princípios da competência linguística de um falante do chinês, e não faz sentido perguntar qual é a melhor. No entanto, no caso do raciocínio e da decisão parecer fazer sentido perguntar se um sistema de regras e princípios é melhor do que outro. Um outro tipo de 'correção' está aparentemente em jogo.

Voltemos à interpretação dos estudos psicológicos. A interpretação pessimista dos resultados da psicologia do raciocínio e da decisão utiliza as noções de competência e performance para dizer o seguinte: os erros que os sujeitos cometem nestas experiências não são erros de performance, i. e., desvios casuais e pontuais de uma competência normativa intacta. Pelo contrário, os casos mostram o que, nos sujeitos, está no lugar da competência, e o que se demonstra é que nesse lugar não estão 'regras adequadas' de raciocínio e decisão. O que está lá, no lugar da competência, são regras mais simples, ou "heurísticas", que frequentemente conseguem a resposta correcta, apesar de também frequentemente não a conseguirem. Noutras palavras, de acordo com a interpretação pessimista, é a 'competência' dos sujeitos que é normativamente defeituosa; as regras de raciocínio 'em acção' nos humanos não são plenamente racionais. Kahneman e Tversky não defendem abertamente esta interpretação, mas ela encontra-se, por exemplo, em Slovic, Fischhoff e Lichtenstein: «as pessoas não têm os programas adequados para muitas tarefas de raciocínio importantes (...) Não tivemos a oportunidade de desenvolver um intelecto capaz de lidar conceptualmente com

a incerteza»²⁰. Basicamente, isso significa que se teria demonstrado experimentalmente que somos irracionais, e esse seria um constrangimento impossível de ultrapassar²¹. Há alguma forma de recusar esta interpretação? Sim, existe uma forma, que é para muitas pessoas convincente e que está ligada à psicologia evolutiva. Antes de passar à descrição dos princípios desta segunda interpretação, queria recordar que, nos termos da Introdução, os problemas que estão aqui em jogo dizem respeito àquilo a que chamei a questão 2, a questão das teorias normativas formais e dos agentes reais, à questão 3, a questão dos modelos de racionalidade e sua suposta refutabilidade empírica, e, eventualmente às próprias arquiteturas cognitivas dos agentes e ao estudo que a teoria da mente tem nesta arquitectura, e portanto também às questões 4 e 5 (a questão da suposição de racionalidade na teoria da mente e a questão do estatuto da mente e dos conceitos mentais na realidade).

Psicologia evolutiva e racionalidade – resultados e interpretações

A **psicologia evolutiva** é basicamente a abordagem da mente de acordo com a qual esta é constituída por um grande número de módulos” ou “órgãos mentais” (sistemas especializados), os

²⁰ SLOVIC, FISCHHOFF & LICHTENSTEIN 1976: 174.

²¹ Veja-se a expressiva descrição de S. J. Gould: «Gosto particularmente do exemplo de Linda, pois eu sei que a conjunção é menos provável, contudo, um pequeno homúnculo continua a saltitar para cima e para baixo na minha cabeça gritando-me – “mas ela não pode ser apenas caixa de um banco; lê a descrição.” ... Por que é que constantemente cometemos este simples erro lógico? Tversky e Kahneman afirmam, correctamente julgo eu, que as nossas mentes não foram construídas (por qualquer motivo) para trabalhar segundo as regras da probabilidade» (S. J. Gould 1992, *Bully for Brontosaurus*, London, Penguin, p. 469, citado em SAMUELS, STICH & TREMOULET 2003).

quais, da mesma forma que outros dispositivos biológicos, teriam sido configurados pela selecção natural para desempenharem funções específicas, neste caso para resolverem problemas de processamento de informação importantes no ambiente em que os nossos antepassados evoluíram. Vou descrever em seguida a forma como os estudos do raciocínio e decisão atrás referidos podem ser interpretados do ponto de vista da psicologia evolutiva. Seguirei aqui Samuels, Stich & Tremoulet 2003 e também a declaração dos princípios da psicologia evolutiva de Cosmides e Tooby atrás referenciada. De acordo com os autores, a Hipótese da Modularidade Maciça (HMM) é essencial à psicologia evolutiva. A HMM é assim descrita por J. Tooby e L. Cosmides: «A nossa arquitectura cognitiva assemelha-se a uma confederação de centenas ou milhares de computadores funcionalmente empenhados (normalmente conhecidos por módulos) desenhados para resolver problemas de adaptação dos nossos antepassados caçadores-colectores. Cada um destes instrumentos tem a sua própria agenda de actividades e impõe a sua própria organização a diferentes fragmentos do mundo. Existem sistemas especializados em indução gramatical, em identificação de rostos, em calcular a posição de objectos que vêm pelo ar (...), em reconhecimento de emoções nas expressões faciais. Existem mecanismos para detectar a animalidade, a direcção do olhar e para detectar fraudes. Existe um módulo da “teoria da mente”... uma variedade de módulos de deduções sociais... e um grande número de outras elegantes máquinas»²². Noutras palavras, desta vez palavras de um outro autor que tem estado próximo deste tipo de posições, Steven Pinker: isto significa que «a mente humana... não é um computador com um objectivo genérico, mas antes uma colecção de instintos adaptados à resolução de problemas evolutivamente importantes»²³. Pinker chama a esta ideia a ideia de ‘mente como um Canivete Suíço’.

²² TOOBY & COSMIDES 1995: xiv.

²³ PINKER 1994.

O conceito de módulo é bastante corrente na teoria cognitiva e é por vezes considerado de formas diferentes. Samuels, Stich e Tremoulet pensam que o conceito que interessa para interpretar o alcance das hipóteses da psicologia evolutiva é o conceito de módulo darwiniano. Em geral módulos são estruturas mentais ou componentes da mente que podem ajudar a explicar capacidades cognitivas, componentes que são funcionalmente específicos, por oposição a serem aplicáveis a todos os domínios da cognição (domínio-geral). De forma mais específica, “módulo” pode significar ainda duas coisas diferentes (i) sistemas de representações mentais (como as Gramáticas Chomskyanas) e (ii) mecanismos computacionais. Samuels, Stich e Tremoulet falam de *módulos chomskyanos*²⁴ e *módulos computacionais*.

Se os módulos chomskyanos são sistemas de representações, os módulos computacionais são instrumentos computacionais de domínio-específico, i. e., que apenas se dedicam a resolver problemas de um domínio específico porque só são capazes de processar um restrito leque de *inputs*, nomeadamente representações das propriedades e dos objectos encontrados num domínio particular (por exemplo, a fonologia ou a aritmética, casos em que um módulo computacional fonológico apenas fornece análises de *inputs* relativos a objectos e propriedades fonológicas, e um módulo computacional aritmético apenas dará soluções para problemas aritméticos). A outra característica dos módulos computacionais é serem componentes da mente relativamente autónomos.

O filósofo americano Jerry Fodor é talvez o mais conhecido teórico da modularidade (Fodor 1983). De acordo com Fodor, módulos são mecanismos computacionais de domínio específico

²⁴ São módulos chomskyanos de acordo com Samuels, Stich e Tremoulet não apenas as gramáticas chomskyanas como também, por exemplo, teorias mentalmente representadas – sistemas de princípios – específicas de um domínio (por exemplo, a física, a psicologia, a matemática), como aquelas de que falam psicólogos do desenvolvimento como Susan Carey e Elizabeth Spelke (cf., por exemplo, CAREY & SPELKE 1994).

co (1) fechados à informação, (2) obrigatórios, (3) rápidos, (4) superficiais, (5) localizados ao nível neuronal (6) susceptíveis a uma degradação característica, e (7) amplamente inacessíveis a outros processos. A concepção de modularidade de Fodor teve grande impacto na ciência cognitiva, mas não necessariamente entre os psicólogos evolucionistas (sobretudo por causa das ideias de fechamento à informação e *outputs* superficiais). Os psicólogos evolucionistas falam também de módulos, no entanto é complicado dizer exactamente de que tipo de módulos (ou seja, não é simples saber que características devem ter os módulos evocados nas suas teorias). Samuels, Stich e Tremoulet procuraram discernir um protótipo de *Módulo Darwiniano* na psicologia evolutiva e avançam o seguinte conjunto de características: (1) especificidade de domínio (ii) serem mecanismos computacionais (os módulos darwinianos não são, assim, módulos chomskyanos mas módulos computacionais, que podem, no entanto, ao resolver problemas, utilizar sistemas de conhecimento de domínio-específico (ou seja, módulos chomskyanos). Módulos Darwinianos são uma espécie de módulos computacionais/chomskyanos. Um exemplo importante de módulo darwiniano seria o módulo da “teoria da mente”, de que falarei em seguida, inclusivamente para interpretar os resultados que nos interessam. Os módulos darwinianos são além do mais (iii) estruturas cognitivas inatas cujas propriedades características seriam em grande medida determinadas por factores genéticos, bem como (iv) resultado da selecção natural, (v) universais nos humanos.

Resumindo e analisando: os módulos darwinianos em causa na psicologia evolutiva são inatos, especula-se que vieram a sê-lo via a selecção natural, estão lá não por tornarem os agentes aptos no ambiente presente mas por o terem feito quanto aos antepassados dos agentes no ambiente destes (i. e., são *adaptações*, mas não são necessariamente adaptivos no presente). Uma outra *parte importante da proposta* é esta: por oposição à modularidade fodoriana, essencialmente periférica, de acordo com a qual os sistemas centrais dos agentes cognitivos – essencialmente res-

ponsáveis pela fixação de crenças – envolvem processos não modulares, a modularidade da psicologia evolutiva vai até ao âmago da mente, *all the way in*, até ao que Fodor chamaria os sistemas centrais não modulares. É esta a Hipótese da Modularidade Maciça (HMM) que referi no início deste ponto. A hipótese é obviamente apenas uma hipótese, para a qual há que encontrar provas e argumentos. É no entanto uma hipótese incontornável para pensar na natureza da cognição e especialmente relevante para a interpretação dos resultados psicológicos de que partimos.

Temos agora na mão a chave para uma nova abordagem interpretativa dos resultados da psicologia do raciocínio e da decisão. A pergunta-chave é a seguinte: será a mente humana constituída em grande parte, ou mesmo inteiramente, por módulos darwinianos? E se for esse o caso, de que forma é que tal facto influenciaria o raciocínio e decisão? Um dos objectivos centrais da psicologia evolutiva é elaborar e testar hipóteses acerca dos módulos Darwinianos que supostamente compõem grande parte da mente humana. Uma vez tendo a análise evolucionista avançado uma hipótese plausível, o próximo passo da estratégia de pesquisa da psicologia evolutiva é *testar* essa hipótese, procurar provas de que os humanos contemporâneos possuem um módulo com as propriedades em questão. Como Samuels, Stich e Tremoulet notam, as provas de que os psicólogos evolucionistas necessitam podem vir de lugares muito diferentes: desde estudos experimentais sobre os processos de raciocínio dos seres humanos normais²⁵, a estudos respeitantes ao aparecimento e desenvolvimento de capacidades cognitivas em crianças²⁶, ao estudo de deficiências cognitivas em algumas populações não-normais²⁷, aos estudos de antropologia cognitiva²⁸, da história, e até mesmo de áreas tão insuspeitas como o estudo comparativo de tradições

²⁵ COSMIDES 1989, COSMIDES & TOOBY 1992, COSMIDES & TOOBY 1996, GIGERENZER 1991; GIGERENZER & HUG 1992.

²⁶ CAREY & SPELKE 1994; LESLIE 1994.

²⁷ BARON COHEN 1995.

²⁸ BARKOW 1992.

legais. Quando provas recolhidas em várias áreas apontam numa mesma direcção, podemos argumentar que a hipótese sugerida pela análise evolucionista a favor da existência de determinado um módulo mental se torna cada vez mais sólida.

Mas exactamente de que forma é possível reinterpretar neste contexto os dados acerca de raciocínio e decisão que referi atrás? Tínhamos ficado com a interpretação pessimista segundo a qual as pessoas não usam regras ou princípios normativos adequados para raciocinarem acerca de determinados problemas simplesmente porque não dispõem deles. O que a abordagem da psicologia evolutiva nos pode levar a fazer é desenhar variações dos problemas em causa de modo a aproximá-los do tipo de problemas com que os nossos antepassados se terão confrontado. Ora, o que aparentemente se observa é que nessas circunstâncias as performances dos sujeitos melhoram. Para perceber por que razão isto acontece nos casos que nos interessam aqui foi avançada a hipótese da existência de pelo menos dois módulos darwinianos: um módulo para lidar com raciocínios probabilísticos quando a informação é apresentada num formato frequentista, e um módulo para lidar com raciocínios relativos a fraudes em contextos de intercâmbio social.

O que se passa quanto ao primeiro caso é o seguinte. Tínhamos visto que as pessoas se saem frequentemente mal em raciocínios de probabilidade. De acordo com a interpretação pessimista, o que acontece é que as pessoas utilizam métodos heurísticos simples (“rápido e sujos”) para lidar com estes problemas, não tendo acesso aos princípios normativamente correctos adequados de raciocínio probabilístico. Mas de um ponto de vista evolucionista esta falta de acesso aos ‘princípios do raciocínio probabilístico’ seria surpreendente: agentes reais como os humanos vivem num ambiente em que têm que tomar decisões baseadas na incerteza e lidar com o aspecto estatístico do mundo. É pelo menos o que vários autores argumentam²⁹. Em alternativa sugere-se o seguinte:

²⁹ Cf., por exemplo, GIGERENZER 1994, COSMIDES & TOOBY 1996.

os nossos antepassados viram-se certamente confrontados com informações probabilísticas, que lhes colocaram problemas de adaptação. A chave aqui é no entanto perguntar que género de informação probabilística teria estado disponível para se terem desenvolvido determinados mecanismos de raciocínio indutivo. Essa informação não teria certamente a forma com que hoje nos confrontamos (quando nos são fornecidas probabilidades de acontecimentos únicos, ou informação expressa em termos percentuais: «Aquilo que estava disponível no ambiente em que evoluímos era a frequência com que realmente nos deparávamos com determinados acontecimentos – por exemplo, quando tínhamos sido bem sucedidos 5 vezes das últimas 20 que caçámos no desfiladeiro norte. Os nossos antepassados homínidos estavam submersos num imenso fluxo de acontecimentos observáveis que podiam ser usados para aperfeiçoar as suas decisões. Assim, se temos adaptações para o nosso raciocínio indutivo, elas devem ter como *input* a frequência da informação»³⁰. Cosmides e Tooby conjecturam por isso que os nossos antepassados desenvolveram mecanismos que tomavam como *input* a frequência de acontecimentos, mantinham essa informação como representações frequentistas e usavam essas representações frequentistas como base de dados para um raciocínio indutivo eficaz. Uma vez que se espera que a mente contenha muitos módulos especializados, eles admitem encontrar outros módulos envolvidos no raciocínio indutivo, trabalhando de outras maneiras.

O que podemos esperar a partir desta hipótese? Basicamente que (1) A performance no raciocínio indutivo irá variar consoante for perguntado aos inquiridos para avaliarem a frequência ou a probabilidade de determinado acontecimento, (2) A performance nas versões frequentistas dos problemas será superior à das versões não frequentistas (3) Quanto mais inquiridos forem mobilizados para formar uma representação frequentista, melhor será a performance. (4) Nos problemas frequentistas a performance irá satisfazer alguns dos constrangimentos que o cálculo

³⁰ COSMIDES & TOOBY 1996: 15-16.

de probabilidades específica, tal como a regra de Bayes. Isto aconteceria porque alguns mecanismos de raciocínio indutivo na nossa arquitectura cognitiva incorporam aspectos de um cálculo de probabilidades³¹.

Cosmides e Tooby levaram a cabo uma série de experiências (nomeadamente em torno do problema do diagnóstico médico acima descrito³²) e crêem ter confirmado as suas expectativas. No próprio problema de Linda, de Tversky e Kahneman, reformulado em termos frequentistas, a percentagem de inquiridos que cometem a falácia da conjunção pode ser reduzida drasticamente. A versão seria qualquer coisa como:

Linda tem 31 anos, é solteira, sincera e muito inteligente. Licenciou-se em filosofia. Enquanto estudante, Linda interessava-se profundamente por assuntos relacionados com a discriminação e a justiça social; além disso, participava em manifestações anti-nuclear.

Existem 200 pessoas que encaixam na supradita descrição. Quantas delas são:

bancárias?

bancárias e participantes activas em movimentos feministas?

A 'ilusão cognitiva' do excesso de confiança parece igualmente poder desaparecer mediante reformulações frequentistas³³. Tudo isto é no entanto contestável, e foi contestado por Tversky e Kahneman. Para Samuels, Stich e Tremoulet, «a versão hipótese frequentista sugerida por Gigerenzer, Cosmides e Tooby é demasiado simplista. Nem todas as representações frequentistas activam mecanismos que produzam bons raciocínios bayesianos, da mesma forma que a apresentação de informação num formato sequencial, a partir do qual a distribuição da frequência pode

³¹ COSMIDES & TOOBY 1996 : 17.

³² CASSCELLS, SCHOENBERGER & GRAYBOYS 1978.

³³ GIGERENZER 1991: 89.

facilmente ser retirada, nem sempre activa mecanismos que façam um bom trabalho na detecção de correlações. Será preciso mais trabalho experimental para que se consiga determinar os factores que permitem a activação de bons raciocínio bayesianos e que possibilitam uma boa detecção de correlações³⁴.

Vimos a forma como mecanismos para lidar com a dimensão estatística do mundo poderão intervir na forma como raciocinamos. Outra proposta interessante de Cosmides e Tooby é que um outro mecanismo, um módulo da mente, também o faz, nomeadamente em casos como a célebre tarefa de selecção de Wason. O ponto de partida são versões da tarefa em que as performances melhoram drasticamente. Vejamos um exemplo.

Na sua luta contra os condutores alcoolizados, as autoridades judiciais de Massachussets estão a cancelar licenças de venda de bebidas a torto e a direito. Você é segurança num bar em Boston, e perderá o seu emprego se não aplicar a seguinte lei:

“Quem estiver a beber cerveja tem de ter mais de 20 anos.”

Os cartões aqui apresentados contêm informação acerca de quatro pessoas sentadas numa mesa no seu bar. Cada cartão representa uma pessoa. Um dos lados do cartão indica o que a pessoa está a beber e o outro lado diz a idade dessa pessoa. Indique apenas os cartões que terá necessariamente de virar para saber se essa pessoa está a infringir a lei.

A beber cerveja	25 anos de idade
A beber Coca-Cola	16 anos de idade

A primeira coisa a ter em conta é que a estrutura do problema é exactamente idêntica à estrutura da primeira tarefa de Wason que descrevi. O conteúdo no entanto é diferente, e esse facto tem uma importante repercussão nos resultados obtidos (cerca de 75% dos estudantes universitários inquiridos responderam

³⁴ SAMUELS, STICH & TREMOULET 1993, 5.

correctamente a esta versão da tarefa de selecção, enquanto que apenas 25% tinham respondido correctamente à outra versão). Um grande número de estudos têm explorado estes “efeitos de conteúdo” na tarefa de selecção, e eles são sólidos. O problema é interpretá-los. Ora, Cosmides e Tooby, de novo avançam uma curiosa proposta³⁵, a que Samuels, Stich e Tremoulet chamam a hipótese da detecção do batoteiro (*cheater detection*).

Sobre o pano de fundo de um ambiente dos nossos antepassados que incluiria a capacidade de encetar trocas sociais estáveis, Cosmides e Tooby avançam a hipótese de acordo com a qual existiriam nas nossas mentes um ou mais módulos darwinianos cuja função seria reconhecer trocas recíprocas e detectar batoteiros, i. e., indivíduos que aceitam os benefícios das trocas mas não pagam os custos. Em suma, Cosmides e Tooby admitem a existência de um ou mais módulos de detecção de batoteiros, e seria a intervenção de tais dispositivos na tarefa de Wason reformulada aquilo que provocaria os atestados efeitos de conteúdo. Algumas versões da tarefa de selecção utilizam os módulos mentais desenhados para detectar batoteiros em situações de trocas sociais. As versões da tarefa que não accionam os módulos de detecção de batoteiros e de trocas ou contratos sociais são consideradas muito mais difíceis e a performance das pessoas é bastante pior. O problema do segurança no bar de Boston accionaria o hipotético módulo de detecção de batoteiros, enquanto que a primeira versão que descrevi, o problema que utilizava vogais e números ímpares, não o faria.

Um autor que se tem dedicado a esta linha de investigações é o psicólogo alemão Gerd Gigerenzer³⁶. As experiências e respectivas interpretações são controversas e outros autores avançam hipóteses bastante diferentes acerca destes ‘módulos de sociabilidade’.

³⁵ COSMIDES 1989, COSMIDES & TOOBY 1992.

³⁶ Cf. Center for Adaptive Behaviour and Cognition, do Max Planck Institut for Human Development de Berlim (<http://www.mpib-berlin.mpg.de/en/forschung/abc/index.htm>).

À parte a correcção de alguma particular teoria psicológica do raciocínio e da decisão, o que é que estas discussões nos devem fazer pensar sobre racionalidade? Bom, antes de mais fazem-nos encarar a necessidade de descer a discussões cognitivas, e de conjugar possíveis conclusões a retirar a esse nível com as teorias normativas da racionalidade (trata-se das condições de possibilidade para utilizar os cânones utilizados como modelos, na terminologia que tenho vindo a utilizar). Mostram-nos também a necessidade de repensar o estatuto dos *folkconcepts*, neste caso para falarmos das nossas capacidades mentais: eles podem perfeitamente aplicar-se a uma dada realidade em resultado de uma dada história evolutiva que nos faz ‘tomar’ o mundo cognitivamente de certa maneira, sendo portanto de um certo ponto de vista um ponto de partida e não a última palavra daquilo a que queremos chamar racionalidade.

Pensar sobre a natureza da racionalidade obriga-nos ainda a considerar a forma como um sistema atribui mente a outro sistema; por exemplo, uma pessoa a outra pessoa, ou a si própria. E o que as investigações psicológicas sugerem é que o comportamento interpretativo de agentes, que os filósofos capturam usualmente de forma exclusivamente conceptual e descritiva, pode bem-estar relacionado com dispositivos cognitivos internos aos agentes, resultantes da evolução, dispositivos para *mindreading*, que consituiriam então a ancoragem cognitiva da linguagem da racionalidade considerada pelos filósofos como condição para a descrição de um sistema cognitivo físico como mental. A ser assim, a suposição de racionalidade evocada nas teorias da mente de raiz quineana (a interpretação radical do próprio Quine, a tradução radical de Davidson, a estratégia intencional de Dennett) teria uma realidade cognitiva sólida, independente de, e anterior a, qualquer interpretação consciente feita pelos agentes. A atribuição de mentalidade a si próprio e a outrem, relacionada com os mecanismos de raciocínio e decisão, seria assim um ‘fundo cognitivo’, antes da racionalidade consciente e subjacente a esta. Outras conclusões, que partem

destes assuntos para uma discussão meta-epistemológica geral (acerca do que fazemos quando abordamos empiricamente a racionalidade teórica) e que conduzem a uma inflexão em direcção ao pragmatismo ficam para o Capítulo 3, *Filosofia e racionalidade teórica: em que devemos acreditar?*, em que considerarei as propostas de Stephen Stich acerca da natureza da racionalidade avançadas no seu livro *The Fragmentation of Reason* (1993).

Capítulo 2

*Filosofia e racionalidade prática: o que devemos fazer?*¹

Afirmei no capítulo anterior que a teoria da decisão constituía uma referência normativa central dos estudos da racionalidade. Ela é também a referência das investigações sobre racionalidade prática de muitos filósofos contemporâneos, embora não pertencendo, como é óbvio, estritamente à literatura filosófica. Na literatura filosófica, e aqui considero mais especificamente a filosofia analítica, podemos considerar que o tratamento dos problemas da racionalidade prática se estende desde discussões acerca da natureza da acção, da intenção na acção, da racionalidade na acção até discussões acerca das boas razões para agir em situações específicas (e à questão de saber se essas 'boas razões ou razões normativas' são ou não motivadoras para os agentes). Na parte final do capítulo anterior procurei mostrar a necessidade de ir 'mais abaixo', em termos de mecanismos cognitivos subpessoais, seu funcionamento e origem evolutiva, para estudar o raciocínio e decisão, antes de demasiado pron-

¹ Agradeço ao Pedro Madeira a sua colaboração no Projecto *Racionalidade, Desejo, Crença* do GFMC e o incentivo que as suas leituras e propostas acerca destas questões representaram para mim. Essa colaboração traduziu-se em vários artigos e uma tradução publicados no n.º 9 da revista *Intelectu*, na *Revista da Faculdade de Letras – Série de Filosofia* e na revista *Trólei*.

tamente assumirmos que podemos avaliar normativamente raciocínios e decisões de agentes reais por transposição directa de cânones abstractos determinados. Neste capítulo tratar-se-á de uma outra forma de 'ir mais abaixo', desta vez não ao nível subpessoal dos agentes através de um inquérito acerca da origem, por exemplo, evolutiva, de determinados funcionamentos cognitivos, mas à análise dos conceitos necessários para falar de racionalidade prática, começando com o próprio conceito de acção. Sem conceitos como 'acção', 'intenção', 'razões para agir' não há estudos da racionalidade. Ora, a análise de tais conceitos traz consigo problemáticas específicas, que se relacionam, por exemplo, com questões de filosofia da mente (falar de 'agentes' e de 'acção' envolve ter ideias acerca de assuntos tais como a natureza da representação, da consciência e eventualmente da identidade pessoal, assuntos habitualmente tratados na filosofia da mente). Neste Capítulo procurarei esquematizar aqueles que me parecem ser os elementos essenciais para compreender a forma como os filósofos analíticos abordam usualmente as questões da racionalidade prática. Isso será feito em três momentos: considerando antes de mais, (i) as questões da natureza da acção, da explicação da acção e da intenção na acção, em seguida (ii) a questão da racionalidade na acção, tendo no centro a teoria instrumental da racionalidade, e finalmente (iii) a questão das razões para agir e portanto propriamente a filosofia moral, considerada como dizendo respeito à forma como podemos e devemos conformar e moldar o nosso querer, de forma virmos a querer e a agir de modo justificável pelas nossas próprias e explícitas razões, para além de, ou por sobre, o capricho e o instinto. Estes três estratos organizam em grande medida a forma como a literatura filosófica trata a questão da racionalidade prática.

Acção, modelo crença-desejo, intenção na acção, razões e causas. E. Anscombe e D. Davidson

Os filósofos chamam teoria da acção, ou **filosofia da acção**, a uma área da filosofia em que se procura analisar (i) em que consiste uma acção, i. e., o que é que faz com que eventos, acontecimentos no mundo, constituam acções de agentes, e (ii) em que consiste 'explicar' uma acção, nomeadamente através de razões fornecidas pelo agente. Parte do interesse desta discussão² é vir a saber o que constitui actividade e eventualmente responsabilidade de agentes, o que é e o que não é deliberado nos comportamentos de humanos. Imaginemos uma situação concreta: uma pessoa A tem um revólver na mão, acontece um disparo, acontece algo à pessoa B que está uns metros à sua frente, ela cai morta. Pelo menos duas coisas diferentes podem aparentemente ter sido o caso: A pressionou o gatilho com a intenção de matar B, e B morre de facto ou deu-se o acaso infeliz de um movimento incontrolado do dedo de A fazer o gatilho disparar, provocando a morte de B³. Poderá a diferença entre as duas situações dever-se a algo que se passa (ou não se passa) no interior do agente, e que terá a ver com entidades mentais como intenções e razões? Procuo em seguida responder a esta pergunta, e ao responder-lhe ficará também claro o que se entende por acção, por contraste com meros eventos. O filósofo americano Donald Davidson (1917-2003) e a filósofa britânica Elizabeth Anscombe (1919-2001) são dois nomes incon-

² Que se situa na convergência de domínios tais como a filosofia da mente, a metafísica, a teoria da decisão racional e a psicologia (é aliás preferível falar aqui em geral de ciência cognitiva, de forma a incluir também a neurociência, cujos tratamentos da voluntaridade e iniciação de comportamento são obviamente relevantes).

³ Ou pessoa B pode ter tido um ataque cardíaco fatal no instante em que a bala começa a penetrar no seu peito (agradeço a Ricardo Fernandes, aluno de Filosofia da Linguagem na FLUP (2003-2004), pela sugestão de variações possíveis deste exemplo).

tornáveis na tentativa de resposta. Grande parte do que digo a seguir baseia-se nas propostas de Davidson e Anscombe quanto à natureza das acções, da explicação das acções e das acções intencionais. Vale a pena recordar que Davidson considerava a obra de Anscombe *Intention* (1957) o mais importante trabalho sobre raciocínio prático desde Aristóteles e que uma das intenções de Anscombe em *Intention* é afastar aquilo que considera serem as más interpretações estabelecidas da caracterização do raciocínio prático feita por Aristóteles⁴.

Queremos então saber antes de mais **o que é uma acção**. Em seguida perguntar-se-á **em que consiste a explicação de uma acção** e em seguida **o que é uma acção intencional**. Apenas se existem essas coisas (acção, intenção) se poderá sequer falar de agentes (ou pelo menos de certo tipo de agentes), que deliberam, decidem e agem intencionalmente e sem agentes desse tipo não há teoria da racionalidade prática possível. Vejamos um novo exemplo. Pense-se na seguinte diferença: vemos um vaso de flores que cai de um segundo andar em cima da cabeça de alguém que passa na rua. Uma opção: o vento fez com que o vaso caísse. Outra opção: alguém contratado para matar a pessoa que passa na rua empurrou certamente o vaso. Não

⁴ Anscombe pensa em interpretações 'intellectualistas', que vêm o conhecimento prático sob o prisma do conhecimento teórico, o que conduz nomeadamente a considerar o silogismo prático como prova de uma conclusão (com o estranho resultado de que, como vários comentadores notaram, nada parece seguir-se quanto a fazer alguma coisa...). 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? (i. e., não somos necessariamente movidos a acção alguma). Cf. ANSCOMBE 1957: 58. Anscombe quer criticar quem vê o raciocínio teórico sob a perspectiva do raciocínio prático, por isso escolhe esse exemplo. Podemos determinar no entanto que o silogismo prático – entendido como modelo de raciocínio para agentes racionais que escolhem as suas acções – tem que ter como premissa a enunciação de um desejo do agente.

diríamos que se trata de uma acção em ambos os casos, mesmo que um registo filmado do vaso que cai e mata o transeunte nos mostrasse exactamente 'a mesma sequência de eventos num e noutro caso'. O que é que faz com que num caso falemos de acção mas não no outro?

Admitamos um mundo de eventos, particulares e concretos, que simplesmente acontecem, de acordo com leis físicas. O que constituiria neste mundo uma acção? Na terminologia associada a Davidson (e em geral ao modelo crença-desejo⁵) mostrar que uma acção é uma acção, que nesse mundo, 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**. Para racionalizar uma acção 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**. Sem isso não poderemos falar de acções. Acções são então 'acontecimentos intencionais' e de acordo com o modelo crença-desejo o que temos que fazer para tornar inteligível a acção como acção é portanto olhar para acontecimentos no mundo de um ponto de vista mentalista, que faça sobressair desejos e crenças relevantes de um agente. Diremos portanto que existe uma acção se eventos no mundo são 'racionalizáveis' a partir de crenças e desejos de um agente.

Na teoria da acção está em causa ainda uma outra questão: é necessário não apenas especificar o que distingue acções de meros acontecimentos mas também saber que força explicativa tem o discurso acerca de intenções de agentes e razões do agente para as acções. Noutras palavras, é necessário saber se

⁵ Convém aqui perguntar o que poderia opor-se ao modelo crença-desejo. A resposta mais plausível é que o que se opõe ao modelo crença-desejo, ou está mais próximo de se lhe opor, é defender que apenas crenças ou apenas desejos chegam para motivar à acção (cf., por exemplo, NAGEL 1970, *The Possibility of Altruism*).

'coisas mentais' como intenções e razões que se encontram em agentes – e às intenções e razões chega-se a partir das crenças e desejos – e causas se excluem mutuamente, ou se pelo contrário razões podem explicar causalmente acções. Poderão razões, 'coisas mentais', causar acções que são de alguma forma 'intervenção no mundo'? Como é que isso seria possível? Mas o que são afinal 'razões'? Usando a terminologia do modelo crença-desejo, **razões são conjuntos crenças-desejos** de agentes⁶. Ora, uma proposta que por vezes vai em conjunto com a primeira no chamado modelo crença-desejo (é esse o caso em Davidson, por exemplo, e ele constitui referência no pensamento destas questões) é que o par formado pelo desejo e pela crença relevantes para a racionalização da acção também constitui a causa dessa acção. Davidson defendeu esta tese no seu artigo *Actions, Reasons and Causes* (1963) contra uma posição wittgensteiniana muito difundida na altura de acordo com a qual causas e razões eram conceitos de âmbitos diferentes e consistia um erro categorial pensar nas razões como causas. A tese que Davidson desenvolveu obrigou-o a esclarecer posições ontológicas acerca do lugar da mente no mundo.

Vou procurar descrever o caso de Davidson contra os wittgensteinianos. Temos então que os wittgensteinianos consideram uma falácia naturalista falar de 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. No seu artigo clássico de 1963 *Actions, Reasons and Causes* Davidson avança razões para a oposição à posição dos wittgensteinianos. Como já tínhamos visto, para falarmos de acção precisamos de agentes com estados mentais tais como crenças e desejos. Embora crenças e desejos sejam noções mentalistas, cuja relação com o mundo físico é difícil de conceber, Davidson defende que podemos e devemos continuar

⁶ Alguns autores, por exemplo, F. Schick, acrescentariam ainda o entendimento – *understanding* – da situação.

a falar de causas quando falamos de acções, e especificamente podemos falar das razões como causas das acções. Embora descrições psicológicas de eventos constituam um 'modo de explicação sem eco na teoria física' (descrições de acções são intensionais, contextos referencialmente opacos), é possível falar causalmente de razões para agir porque os eventos descritos como razões são também eventos físicos que causam algo no mundo.

Para defender que as razões podem ser causas, Davidson terá que esclarecer o que é 'explicar' no âmbito da acção humana (*versus* 'explicar' no âmbito das ciências físicas). Ora, Davidson pensa que a explicação de uma acção (ao contrário da explicação da ocorrência de eventos por leis físicas) envolve a racionalidade. As acções são racionais, as leis da física não são racionais. No entanto, racionalidade é uma noção problemática, nomeadamente por ser, de acordo com Davidson, uma noção mentalista e holista. A proposta que avança é a seguinte: a ligação entre razão e acção (nos casos em que a razão é de facto a razão para a acção) é uma conexão que obtém entre dois eventos (por um lado, o facto de o agente acreditar e desejar e por outro o facto de o agente agir) e que pode ser descrita de maneiras diferentes. Essa conexão é (i) *racional*, tanto quanto o par crença-desejo (aquilo a que Davidson chama razão primária) especifica a razão para a acção, (ii) *causal*, tanto quanto um evento causa o outro. Aliás para Davidson não há outra maneira de distinguir exactamente por qual razão um agente age a não ser dizendo que foi essa a razão, de entre as muitas possíveis, que causou a acção⁷.

⁷ Além de *Actions, Reasons and Causes* (1963) Davidson publicou vários outros artigos importantes na teoria da acção e da racionalidade, tais como por exemplo, *How is weakness of the will possible?* (1970) (cujo tema são as patologias da racionalidade prática tais como a *akrasia* e o auto-engano), *Agency* (1971), *Intending* (1978). Todos estes artigos se encontram em DAVIDSON 1980, *Essays on Actions and Events*. Um outro artigo que merece referência neste contexto é *Rational Animals* (DAVIDSON 1982). Mais recentemente foi publicado um volume de ensaios intitulado *Problems of Rationality* (DAVIDSON 2004).

Como afirmei, foi precisamente a teoria da acção de Davidson que o levou a especificar a sua célebre proposta ontológica, o monismo anómalo. O monismo anómalo é a resposta de Davidson à questão ‘Como é que o mental existe no mundo físico?’. De acordo com Davidson, o mental deve ser pensado em termos de ‘superveniência’. Em *Mental Events* (1970) Davidson coloca o problema da seguinte forma: «Eventos mentais tais como percepções, recordações, decisões e acções resistem a ser capturados na rede nomológica da teoria física. Como é que este facto pode ser reconciliado com o papel causal de eventos mentais no mundo físico?»⁸. A ideia de Davidson, como já vimos, é que aquilo que ocorre (eventos que são ocorrências datadas-localizadas irrepitíveis) pode ser descrito enquanto físico e enquanto mental. No primeiro caso estamos perante um mundo regido pela causalidade. Se aquilo que ocorre é descrito enquanto mental há uma ‘anomalia’, e os conceitos a utilizar são de superveniência. As leis físicas não se aplicam a eventos-descritos-como-mentais. Não há, não pode haver, segundo Davidson, leis psico-físicas. Em concreto explicações racionais, explicações acerca da nossa natureza mental, não podem ser dadas no vocabulário das ciências físicas.

‘Monismo’ significa aqui obviamente que o mundo é de uma só natureza (e é físico), ‘anómalo’ refere-se ao que não cai debaixo de uma lei. Toda a esfera do mental é ‘anómala’ relativamente ao mundo regido por leis que é o mundo físico. Mas o facto de o mundo, que é físico, ser regido por leis da física, não é suficiente para nos permitir falar do pensamento e da mente em termos físicos. As leis da física não servem para explicar a mentalidade. De resto, o holismo que permite o acesso interpretativo ao mental não encontra sequer eco na física, admite Davidson. Para compreendermos esta posição temos que especificar o que se quer dizer com explicação. Note-se que esta discussão generaliza a discussão acerca de razões como causas de acções.

⁸ DAVIDSON 1970: 207.

Quando se trata de ciências físicas e naturais é fácil dizer o que é explicar: *explicar é subsumir eventos físicos a uma lei*. É esta forma de pensar que não faz sentido transpor para o pensamento sobre acções de humanos. Quando falamos de explicar uma acção o que queremos não é descrever eventos físicos sob leis mas sim dar razões, dar as razões que conduziram à acção.

As posições defendidas neste âmbito – trata-se de saber até que ponto coisas mentais podem ser explicativas daquilo que acontece – determinam a forma como vemos o poder causal de explicações psicológicas, explicações de acontecimentos no mundo em termos de razões de agentes, sendo obviamente relevantes para o tratamento da voluntariedade na acção. Saber de que acção falamos e o que a provocou levar-nos-á ainda à necessidade de individuar acções, a procurar saber se os mesmos eventos sob diferentes descrições mentalistas podem constituir diferentes acções, com as quais o agente tem diferentes relações de voluntariedade (a intuição é, basicamente, que aquilo que é ou não é voluntário no que o agente está a fazer tem a ver com o que o agente sabe, com a forma como o agente configura a situação).

É muito importante repetir que racionalizar uma acção, nesta linguagem davidsoniana e que se generalizou, não quer de forma alguma dizer mostrar que essa acção é racional. ‘Racionalizar’ não tem também o sentido freudiano de justificação *post hoc* (i. e., explicação coerente para uma acção cujos verdadeiros motivos não se conhece), de que falei na Introdução: racionalizar uma acção no sentido davidsoniano e da teoria da acção em geral quer dizer apenas mostrar que se trata de uma acção, por contraste com meros acontecimentos.

Se o filósofo americano Donald Davidson é uma referência na teoria da acção e um dos defensores do modelo crença-desejo atrás mencionado, o clássico de referência para o tema da acção intencional é E. Anscombe (1957, *Intention*). Resumo em seguida brevemente as suas posições.

Antes de mais, perguntemos claramente o que é a 'intenção'. Intuitivamente, a intenção é qualquer coisa como um plano do agente, que este tenta realizar através as suas acções. Intenções envolvem assim desejos de fazer alguma coisa e crenças acerca do que é o caso e acerca da forma como actos realizarão o que se pretende. Mas muita coisa pode 'meter-se pelo meio' desta caracterização (por exemplo, eu sei que o meu amigo adora morangos, compro-lhe morangos, ofereço-lhos e ele vai parar ao hospital com uma reacção alérgica inesperada – havia essa intenção? Faz-se uma guerra para libertar um país e mata-se milhares de pessoas – havia essa intenção?). Anscombe estava de resto preocupada com casos como este último e a subtileza da sua abordagem consistiu precisamente em procurar delimitar conceptualmente aquilo de que falamos quando falamos de intenção. Os pontos principais da caracterização da intenção feita por Elizabeth Anscombe, que levam em conta a intuição de que aparentemente uma acção pode ser descrita de muitas maneiras⁹, são os seguintes: (i) não é possível falar de intencionalidade de movimentos corporais tout court, mas apenas da intencionalidade de movimentos sob uma descrição, e o agente deve aperceber os movimentos sob essa particular descrição (ii) não é suficiente

⁹ Um célebre exemplo de *Intention* é o seguinte: temos um homem, que move o braço para bombear a água para encher o depósito de água de uma casa para envenenar os habitantes dessa casa (alguém encontrou uma forma de contaminar sistematicamente a fonte com um veneno cumulativo cujos efeitos são imperceptíveis até não poderem já ser combatidos). Quantas acções e intenções existem aqui? Será que há uma descrição que é 'A descrição correcta'? 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 interesse deste particular exemplo de *Intention* tem obviamente a ver com a distribuição da 'acção' ao longo do tempo e com o não contacto entre supostos agentes e aqueles que sofrem as consequências da acção. O que é que o homem do exemplo está a fazer? Como nota Anscombe, ele está a ganhar um ordenado, a sustentar a família, a gastar as solas dos sapatos, a provocar movimentos do ar, a ajudar a libertar o país, etc...

que exista apercebimento pelo agente de comportamentos próprios sob uma descrição para haver intencionalidade na acção do agente, essa descrição ou apercebimento pelo agente tem que ser feita sem observação, (iii) uma acção será intencional se o agente puder fornecer a razão para a sua acção. Anscombe chamou ainda a atenção para a diferença na direcção de ajuste entre crenças e intenções e propôs uma terminologia ainda hoje comum: nas crenças trata-se de representar o mundo tal como este é (a direcção de ajuste é 'mente-mundo', no sentido em que cabe à mente ajustar-se ao mundo), nas intenções trata-se de 'fazer o mundo ser de uma determinada maneira' (a direcção de ajuste é mundo-mente, no sentido em que cabe ao mundo ajustar-se à mente). Anscombe compara as intenções com plantas de arquitectura, pelas quais se concebe o desenho de um edifício: ter uma intenção é estar disposto a agir no sentido de fazer o mundo ajustar-se-lhe, por contraste com ajustar a mente ao mundo (pense-se, por exemplo, na diferença entre o uso de uma lista de compras para fazer compras numa loja e a elaboração de uma lista de objectos no local do crime feita por um detective). Parte do interesse pessoal de Anscombe nestes problemas tinha a ver com uma característica importante das acções humanas: uma pessoa que pensa que está a fazer uma coisa está, frequentemente, a fazer outras coisas também (embora fosse uma pessoa extravagante e de certa forma heterodoxa, Anscombe era uma católica convicta, preocupada com questões, como por exemplo, as implicações filosóficas da contracepção).

Anscombe defendeu que toda esta caracterização das intenções deveria ser desligada do mundo das causas físicas. Este é um ponto com o qual Davidson não concorda, como já se viu. Davidson adopta no entanto de Anscombe o princípio segundo o qual as acções não são simplesmente intencionais ou não intencionais mas apenas *sob uma dada descrição* podem ser assim consideradas. É aliás apenas sob uma determinada descrição que uma razão explica uma acção. Esta descrição é uma descrição mentalista. Por exemplo: uma mulher sentada à

mesa come intencionalmente um bife. Mas ela não come intencionalmente o seu animal de estimação (no entanto, aquela carne é a carne do seu animal de estimação). Duas acções diferentes podem ser descrições dos mesmos eventos. Dois sentidos de intencionalidade estão aqui em causa: (i) *aboutness* (ii) voluntariedade. A ideia é que o mundo das acções é um mundo mental (ou o mundo considerado/descrito enquanto mental) e só aí se pode falar de voluntariedade.

Recordo que, embora adopte os princípios de Anscombe quanto à distinção do que é e não é intencional, Davidson opõe-se aos wittgensteinianos. Ora, Anscombe é uma discípula de Wittgenstein tanto quanto aqueles pensam que constitui uma falácia naturalista falar de causas de acções.

O que é que interessa reter de tudo isto para pensar sobre a natureza da racionalidade? Há obviamente semelhanças e coincidências nas descrições de Anscombe e de Davidson acerca de acção, intenção, razões para a acção. Vamos reter o seguinte: para eventualmente virmos a falar de racionalidade prática precisamos de acções, e para falar de acções precisamos da mente, i. e., de desejos e crença relevantes de agentes. Sabemos que é apenas sob uma específica descrição que uma acção pode ser intencional. Davidson acrescenta no entanto algo na sua teoria da acção, que Anscombe tinha prescrito, a ideia segundo a qual alguns dos pares formados por um desejo e por uma crença relevante além de racionalizam uma acção também constituem a causa dessa acção. É obviamente polémico falar de causas de acções (afinal acções são entidades que só intencionalmente, de forma mentalista, ‘aparecem’ no mundo), no entanto falar de razões como causas de acções é precisamente a marca davidsoniana em teoria da acção, e deve ser concebida em conjunto com as suas posições metafísicas e de filosofia da mente. As razões pelas quais é problemático falar de causas de acções são várias, a ponto apenas uma. Desde logo é preciso saber quando começam acções e que acções começam – o que pode ser complicado nomeadamente porque aquilo a que queremos chamar acção

só existe sob uma determinada descrição intencional, mas essa descrição intencional ‘refere’ um comando motor, um esquema de acção que tem que ser cerebralmente iniciado (e pode haver desfasamento, ou antecipação da iniciação cortical de um movimento relativamente à intenção na acção apercebida pelo próprio agente – podemos falar aqui de um Problema de Libet¹⁰).

Esta problemática metafísica e mental está subjacente à questão da motivação para a acção. Davidson está interessado em razões e causas, e explicação de acções, e pensa que razões (pares crença-desejo) causam acções – resta saber como se relaciona isso com a questão da motivação do agente à acção. Em princípio, razões para a acção que podem ser fornecidas pelo próprio agente funcionaram para este como ‘razões motivantes’. Mas as razões pelas quais o agente age não são necessariamente boas razões, razões normativas: podemos perfeitamente imaginar que há boas razões para agir de uma certa maneira mas elas não motivam o agente, que age ainda assim por determinadas razões, que são as suas razões, que são motivantes, mas não boas no sentido de normativas.

Vários debates de filosofia moral e teoria da acção se iniciam aqui, não apenas a questão das razões motivantes ou motivadoras *versus* razões normativas¹¹, como o debate acerca de razões internas ou externas¹², e o debate acerca do poder motivacio-

¹⁰ O problema de Libet (chamo-lhe assim em referência ao neurocientista Benjamin Libet) diz respeito ao atraso da consciência de intenção relativamente à iniciação motora da acção no cérebro do agente (cf. MIGUENS 2002: 331-335).

¹¹ Cf. DANCY 2000, apêndice ao Capítulo 1, para uma história da distinção entre razões motivantes e razões normativas. De acordo com o autor, a distinção aplica-se a contextos (respectivamente de deliberação e de justificação), e não a tipos de razões. Razões motivadoras são razões pelas quais fazemos alguma coisa, razões normativas são razões que damos como boas razões para justificar uma acção.

¹² A questão está marcada pela abordagem de Bernard Williams (WILLIAMS 1980). Internalismo e externalismo são termos que dizem respeito ao que acima está chamado razões normativas. O internalismo é a

nal de desejos *versus* crenças (alguns autores defendem, por exemplo, que apenas os desejos motivam, representações ou crenças não têm força motivacional alguma, outros que crenças motivam), o debate em torno da existência de razões para agir que não são desejos.

Racionalidade na acção: a teoria instrumental

Imaginemos no entanto que estamos satisfeitos com as definições mínimas dos conceitos de acção, intenção e razões como causas de acções. Pelo menos temos disponíveis definições, que podem servir-nos como ponto de partida. Passemos então à seguinte questão: em que consiste a racionalidade na acção? Se o ponto de referência da resposta à pergunta *O que é uma acção?* é o modelo crença-desejo, o ponto de referência da resposta à pergunta *O que é racionalidade na acção?* é a teoria instrumental da racionalidade¹³.

A teoria instrumental da racionalidade – frequentemente chamada ‘humeana’, em referência ao filósofo escocês David Hume (1711-1776) – é uma teoria dos critérios que tornam uma acção racional, i. e., ela diz-nos o que faz com que consideremos uma acção racional. Retomemos um dos exemplos que utilizei atrás: vemos um vaso de flores que cai de um segundo andar em cima da cabeça de alguém que passa na rua, sabemos agora que alguém contratado para matar a pessoa que passa na rua

tese segundo a qual um agente A só tem uma boa razão para fazer algo F quando, se conhecesse todos os factos relevantes e deliberasse racionalmente, ele ficaria motivado a fazer F. O externalismo é a tese segundo a qual é possível o agente A ter uma boa razão para fazer F e não se sentir motivado. Internalismo e externalismo são posições acerca da relação entre boas razões, razões normativas e motivação.

¹³ Assim como perguntamos o que se poderia opor ao modelo crença-desejo como teoria da acção, é uma boa estratégia perguntar o que se opõe à teoria instrumental como teoria da racionalidade na acção.

empurrou certamente o vaso. Há um certo sentido em que dizemos que essa acção foi racional, mesmo se achamos que não deveria ter sido praticada – que sentido é esse? É o sentido da teoria instrumental da racionalidade. De acordo com a teoria instrumental (e não é a primeira vez que encontramos esta definição, ela foi avançada no Capítulo 1), **uma acção é racional se e só se mobiliza os meios adequados para atingir os fins que satisfarão os desejos do agente.** Se o assassino contratado do nosso exemplo resolvesse empurrar um esponja do segundo andar pretendendo alcançar o mesmo objectivo a sua acção não seria racional. A teoria instrumental não é uma teoria das (boas) razões para agir – ela diz-nos em que consiste a racionalidade na acção, não é uma teoria moral. O facto de termos razões para considerar a acção racional do assassino contratado como, por exemplo, imoral ou injustificada ou condenável mostra que depois da questão da acção e da racionalidade na acção, vem a questão das razões para agir tal como os filósofos morais a tratam quando perguntam ‘*o que devemos fazer?*’. Temos boas razões para agir assim ou assim? (a pergunta deve dirigir-se, no exemplo que temos tido em mente, ao assassino contratado). Quais são as razões que movem e devem mover os agentes? Essa será outra questão, a questão do estatuto das **razões para a acção** e aí algumas pessoas (os internalistas, como o filósofo inglês Bernard Williams¹⁴) defendem as razões para a acção estão ‘ancoradas nas motivações dos agentes’, outras que elas são independentes, de modo a que poderemos dizer ‘Há razões para agir assim ou assim’ independentemente de os agentes se sentirem motivados (é, por exemplo, a posição do filósofo inglês Derek Parfit). Neste último caso poderíamos dizer, por exemplo, há coisas que seria objectivamente bom fazer, independentemente de os agentes se sentirem motivados a isso.

Tendo na mão uma definição de racionalidade na acção, o que nos falta? Bom, antes de mais notemos que a definição nos

¹⁴ WILLIAMS 1980.

diz em que pode consistir a racionalidade na acção mas não o que fazer em cada circunstância, quais os meios a mobilizar, por quais razões agir.

Outra coisa que definitivamente não temos ainda ao ter na mão a definição instrumental de racionalidade na acção é uma ideia do que leva alguém, algum agente a agir (aliás, Hume só pôde propor uma teoria instrumental da racionalidade na acção porque pensava que a razão era 'impotente' na motivação para a acção, só entra quando os fins estão decididos e estes são postos pelas paixões e desejos, que são ao contrário da razão, algo que nos move a agir). Passemos então à questão da motivação para a acção. Vou tomar as teorias de Hume e Kant como teorias em confronto quanto à motivação dos agentes para a acção.

Filosofia moral: D. Hume versus I. Kant

Repare-se que o facto de ser racional realizar uma acção se temos determinados desejos – até aí leva-nos a teoria instrumental da racionalidade na acção – não constitui por si só razão para realizarmos essa acção: imagine-se que temos um desejo enorme de ter um diamante, está na nossa frente um belíssimo diamante, ninguém nos observa e basta esticar a mão pegarmos nele e ficamos a ter um diamante. E no entanto dá-se o caso de nos impedirmos de fazer tal coisa.

Pegando exactamente neste último exemplo, será que ele nos mostra um dos grandes problemas aqui presentes, o facto de os desejos poderem ser contrários à razão? Mas poderão de facto os desejos ser contrários à razão? Existe uma coisa tal que mereça ser chamada um 'desejo irracional'? Podemos qualificar os nossos desejos, que por hipótese nos assaltam, invadem, sobrevêm, como racionais ou irracionais? Podemos pelo menos falar de desejos motivados (desejos a que chegamos após deliberação, como autores desde Aristóteles a Thomas Nagel pretenderam), em contraste com os mais comuns desejos não

motivados que nos acometem? Por outras palavras, temos assim tanta possibilidade de valorização e intervenção no nosso próprio querer, tanto espaço conceptual de observação?

Hume e Kant, autores que representam – ou pelo menos são tomados em grande parte da literatura filosófica contemporânea como representando – posições antitéticas no âmbito da filosofia moral, dão-nos respostas totalmente diferentes a estas questões. Descrevo em seguida a forma como cada um deles concebe a motivação para a acção. Ambas as descrições têm cabimento no seio de uma dada concepção de mente. Essas concepções de mente são diferentes, e são também elas que estão em causa. De facto, noções como voluntaridade, liberdade, unidade do agente e questões acerca de desejos e racionalidade estão suspensas da teoria da mente, e a adopção de uma determinada teoria da mente e da acção pode pesar de forma definitiva nas posições substantivas defendidas em filosofia moral.

Para se ter uma ideia da posteridade de posições de Hume e Kant no pensamento contemporâneo, podemos ter em mente dois dos mais importantes filósofos contemporâneos, Thomas Nagel e Simon Blackburn. Thomas Nagel é, na forma como concebe os fundamentos da moralidade e a relação entre a ética e a motivação para a acção um kantiano (e isto não significa que Nagel aceite os dualismos kantianos, mas sim que defende que há motivações para a acção especificamente morais, – por contraste com uma dependência face à psicologia – e que as nossas razões para agir estão sujeitas a um constrangimento formal de objectividade¹⁵) enquanto Simon Blackburn é um humeano que se opõe às concepções racionalistas dominantes em teoria da mente e da acção.

¹⁵ Nagel atribui um papel central à motivação ética para a acção na concepção metafísica que o agente faz de si próprio. Kant falava de liberdade, Nagel fala da concepção de si como uma pessoa entre outras pessoas.

Hume: motivação para a acção

David Hume defendeu, de uma forma que ficou célebre, que a razão só por si não nos move a fazer coisa alguma. Sem o suporte das paixões (desejos) a razão é incapaz de nos fazer preferir a ou b, ter um ou outro objectivo. É este o contexto da afirmação do *Tratado da Natureza Humana* (1739) segundo a qual não é contrário à razão eu preferir a destruição do mundo inteiro a um arranhãozinho no meu dedo (Livro II, Parte III, Secção III). É devido ao seu carácter inerte, à sua impotência para nos fazer tender para a ou b ou c, para nos dar fins ou objectivos, que a razão é e deve ser, segundo Hume, 'a escrava das paixões', não deve nem pode pretender outro ofício senão servir e obedecer (Tratado, Livro II, Parte III, Secção III). Para compreender melhor esta posição é preciso saber o que Hume quer dizer com 'paixões' e com 'razão'.

Em geral, na teoria humeana da mente, paixões são impressões de reflexão. Nos termos da teoria humeana da mente, as impressões de reflexão procedem das impressões mais directas, originárias, que são as impressões de sensação (estas por sua vez têm causas naturais e físicas, que não são objecto da teoria da mente). As impressões de reflexão surgem na mente de duas formas: ou imediatamente, provocadas pelas impressões de sensação ou pela intermediação das ideias destas. As paixões são impressões de reflexão do segundo tipo e são de alguma forma primitivas na nossa natureza. Ao contrário das representações de factos pela razão, que são susceptíveis de acordo e desacordo com a realidade, as paixões ou desejos não podem ser verdadeiras nem falsas. Sob o título de paixões Hume analisa no *Tratado da Natureza Humana*, sobretudo no Livro II que lhes é precisamente dedicado, por exemplo, o desejo, a aversão, o medo, o amor, o ódio, a tristeza, a alegria, a ambição, o orgulho, a humildade, a vaidade, a generosidade, etc.

Quanto à razão, que Hume define como a descoberta da verdade e falsidade, que consistem num acordo ou desacordo

com a realidade, seja quanto a relações de ideias seja quanto a existência e factos reais, ela é afinal o aspecto representativo da mente. Ora, o que é importante na teoria humeana da natureza mental é o facto de Hume considerar que não é esta dimensão representativa – a que poderíamos acrescentar a dimensão 'instrumental-calculadora' da mente, que nos permite inferir a partir das situações representadas, avaliar probabilidades e ajustar meios a fins – que nos conduz à acção. Apenas paixões ou desejos nos movem. Não podem existir razões para a acção independentes dos desejos. Daqui decorre, por exemplo (e este é um argumento fundamental no *Tratado da Natureza Humana*, Livro III), que, na medida em que a moral nos move à acção (Hume pensa que ela excita paixões e produz ou impede acções), ela não pode ser derivada da razão. Na prática isto significa que para Hume nem as 'regras' da moral são conclusões da razão nem a 'virtude' é concebível como conformidade com esta.

Para compreender melhor a imagem humeana da nossa natureza mental dinâmica, convém compará-la com o seu oposto. A imagem antitética pertence a um imaginário filosófico e psicológico intemporal, e faz-nos pensar na Razão como um condutor, um cocheiro, que controla os desvairados 'cavalos' das paixões e dos impulsos irracionais¹⁶. Esta imagem corresponde ao clássico dualismo acerca da nossa natureza mental a que podemos chamar 'Razão *versus* paixões' ou 'Apolo *versus* Diónisos'. Ora, de um ponto de vista humeano, quando temos aí uma pessoa a que queremos chamar sensata e razoável, não devemos concebê-la como funcionando (motivacionalmente, dinamicamente) da maneira retratada. À imagem do agente como um decisor racional, que controla as suas próprias paixões insensatas, Hume contrapõe a seguinte ideia: todos os trajectos motivacionais, todos os impulsos ou tendências do querer, são determinados por paixões, preocupações ou cuidados. Noutras palavras, para querermos, para que o nosso querer tenda, temos

¹⁶ Cf. BLACKBURN 1998.

que nos preocupar, importar, ser movidos. Ora, nós só podemos ser movidos por desejos. A razão ela própria não inclina a nada, apenas nos informa acerca das situações e permite prever resultados possíveis, mais ou menos prováveis, das nossas acções. De um ponto de vista humeano, um agente totalmente racional não faria em última análise absolutamente nada, não seria movido ou motivado para a acção. O que acontece pelo contrário com as pessoas, considerando casos não patológicos de pessoas que vão tendo tendência para agir, é que em qualquer situação em que vem a existir um querer conducente à acção há paixões ou interesses que ‘contrabalançam’ outras paixões ou interesses, sendo que alguma vem a dominar, num dado momento, sem no entanto extinguir as restantes. A própria ética funda-se em sentimentos. Assim, de um ponto de vista humeano, uma pessoa sensata ou razoável é assim aquela que é movida por paixões de alguma forma ‘acertadas’ e não a impossível pessoa ‘movida puramente pela razão’.

Repare-se que a estas duas imagens antitéticas de pessoa sensata e racional, a da vitória e a do equilíbrio pontual, correspondem duas imagens muito diferentes das pessoas que estamos dispostos a considerar ‘irracionais’. Na primeira imagem uma pessoa irracional é aquela na qual o condutor, i. e., a Razão, perdeu o controlo das paixões. Mas de um ponto de vista humeano não podemos sequer afirmar com boa consciência de uma pessoa que consideramos, por exemplo, maligna, malévola ou perversa que ela é pura e simplesmente irracional e que qualquer humano racional deve ver isso que nós vemos, ao considerar as suas acções. Apenas podemos afirmar que o curso de acção dessa pessoa é guiado pelo tipo errado de paixão. Mas isso é um defeito de sentimento, não podemos simplesmente afirmar como sendo um facto que essa pessoa não superou a sujeição aos (baixos...) desejos, passando a auto-controlar-se racionalmente.

Por tudo isto a **teoria** humeana da nossa natureza mental é normalmente classificada como **sentimentalista**, por oposição a racionalista. Note-se que embora uma teoria humeana da na-

tureza mental não nos impeça de criticar moralmente as outras pessoas, ela não nos permite fazê-lo evocando a Razão: o que acontece quando emitimos juízos morais é exprimir aprovação ou reprovação. Por mais vilipendiada que a posição humeana seja, nomeadamente devido ao que parece ser a sua natural associação com o relativismo moral, é preciso notar que a tese *central de Hume segundo a qual só uma paixão contrária, equilibra ou suporta outra paixão não é assim tão invulgar no pensamento ético*¹⁷.

*Kant: motivação para a acção*¹⁸

O sentimentalismo de tipo humeano não é a teoria da motivação para a acção mais ‘à mão’ ou mais conveniente em áreas da filosofia prática tais como a ética, a política ou em teorias da acção e da interpretação. E há razões óbvias para isso. Por exemplo, um filósofo como Kant, cuja influência em específicas teorias filosóficas da motivação humana como as teorias éticas e políticas é absolutamente incontornável¹⁹, defendeu uma teoria acerca da estrutura da motivação totalmente oposta à humeana e fê-lo por razões de peso, nomeadamente relacionadas com intenções a que poderíamos chamar democráticas: de facto, a teoria kantiana da motivação para a acção visa colocar todos os agentes como iguais entre si na sua motivação (racional) para a acção.

¹⁷ Por exemplo, Espinosa (1632-1677), na *Ética*, defendeu algo de semelhante, nomeadamente que um afecto só pode ser removido se se lhe opuser outro afecto. Esta é uma boa razão para a ‘simbólica’ escolha, feita por A. Damásio, da figura de Espinosa por oposição à figura de Descartes, e tem como pano de fundo precisamente o problema da motivação para a acção.

¹⁸ Muito do que se segue é influenciado por BLACKBURN 1998.

¹⁹ É possível, por exemplo, considerar que a filosofia política contemporânea tem por referência a obra de J. Rawls, e o pensamento de Rawls é profundamente marcado por Kant.

Precisamente porque Kant é um autor central no pensamento ético e político contemporâneo vou tomá-lo como referência de uma concepção racionalista da motivação para a acção. Gostaria que se partisse da seguinte 'intuição' que podemos atribuir a Kant: se se quer fundamentar o estatuto das leis e da obrigação à qual estas submetem o querer de agentes, coisas tão contaminadas e sensíveis como os desejos não podem ser considerados como a fonte de todo o processo. A obrigação é um conceito moral fundamental para Kant, e Kant concebe-a como uma motivação pela Razão, a Razão maiúsculada, aquela que seria comum a todos os homens, e os faria querer voluntariamente a Lei, uma vez libertos dos desejos.

Mas por que razão necessitará Kant de afastar os desejos para ter uma boa teoria da motivação para a acção? A razão é que Kant considera todos os desejos como sendo, precisamente, 'sensíveis e contaminados' e fá-lo em grande parte porque os vê como sendo não apenas individuais mas também egoístas, i. e., desejos de prazer pessoal. Saber se isto é correcto, se de facto nós somos criaturas tais que todos os nossos desejos são desejos de prazer pessoal, saber se é o interesse próprio que nos move sempre, é fundamental se estamos interessados em compreender a estrutura do nosso querer.

Para Kant, os desejos são basicamente desejo de prazer pessoal e é precisamente isso que há de errado com eles. É por isso que para Kant uma coisa é a motivação por desejos, outra coisa totalmente diferente e muito mais importante, é a motivação pela Razão, que merecerá o nome mais nobre de 'Vontade'. Kant pensava que as leis, concebida como obrigação racional para o querer, não podem ser criaturas do apetite, se estão lá para constrangê-lo. A Lei tem que ser independente da contingência dos desejos dos agentes. Assim, para Kant, não importa, por exemplo, que as inclinações empíricas dos agentes, os seus desejos, sejam generosos ou não, não importa a diferença entre o desejo de fazer algo de amável a alguém e o desejo de matar essa pessoa. Ambos são inclinações subjectivas, logo

não podem ser fonte de Lei. Um sentimento que inclina (por exemplo, que inclina à simpatia para com outrem, um caso que é tão importante para autores como Adam Smith (1723-1790) e David Hume é uma mola motivacional de natureza totalmente diferente da boa mola motivacional que é a Razão.

Hume, Kant e a deliberação

Dada a divergência das suas concepções acerca de desejos e 'Razão', Hume e Kant apresentam-nos imagens muito diferentes da deliberação. Quando pensamos em deliberação pensamos no querer no momento em este se prepara para a cristalização concreta numa decisão. Imaginemos, uma pessoa, um agente. Segundo Hume, aquilo que nela encontramos nesse momento são paixões, inclinações. Aquilo que determina a decisão é a resolução pontual dos conflitos sempre existentes entre desejos. A deliberação faz-se portanto a partir de, ou por sobre, este conflito de desejos: aqueles que perdem podem ser definitivamente afastados, ou apenas provisoriamente silenciados. Além disso, há outros desejos que têm um papel mais estável ou durável, um papel de segunda ordem, nomeadamente encorajar e silenciar outros elementos²⁰.

Consideremos agora a deliberação de um ponto de vista kantiano. Numa pessoa encontramos, como Hume também admite, os desejos. No entanto, para Kant, existe algo cuja natureza é totalmente diferente da natureza dos desejos: a Razão. A razão é livre e está sempre pronta a impedir que as coisas corram mal por causa dos desejos. De facto, o ideal num agente kantiano puro seria que os desejos nem sequer existissem: a condição ideal do querer para Kant é a independência total relativamente a desejos e inclinações. Ser livre de desejos e inclinações é segundo Kant o desejo universal de todo o ser racional, já que desejos e inclinações são um peso, uma escravidão, um afastamento do agente

²⁰ BLACKBURN 1998: 245.

relativamente à sua verdadeira natureza. É certo que no agente kantiano, como no humeano, os desejos estão presentes, mas eles estão passivamente presentes (no máximo serão ‘incentivos’). O que é importante na determinação do querer é a Razão, que legisla. Esta legislação é a ‘fonte da normatividade’ (esta formulação do problema da natureza e condições do querer encontra-se, por exemplo, em Christine Korsgaard, filósofa americana, de Harvard, que em *Sources of Normativity* (1996) lhe dá precisamente uma resposta kantiana). O agente kantiano é, assim, uma imagem do auto-controlo racional absoluto. De um ponto de vista kantiano, o nosso destino moral está exclusivamente nas mãos da Razão. Só a rectidão, a capacidade de agir segundo princípios, conta. *No feelings* é a ‘lei’ (aliás, é a única situação que possibilita a Lei). Além do mais, só essa situação pode ser ‘democrática’, tornar-nos iguais uns aos outros, livres de inclinações subjectivas. Estranhamente isso significa, se se pensar duas vezes, que inclinações parciais da vida moral tais como o amor ou a amizade dificilmente encontram lugar na imagem racionalista kantiana do querer. Mas isto significa por outro lado que para Kant a moralidade pode ser fundada na Razão, o que a des-subjectiviza, des-sensibiliza, universaliza, tornando-a propícia a sustentar a Lei de uma forma que a ‘mera psicologia’, que veremos em Hume e Smith, seria – pelo menos do ponto de vista de um kantiano – incapaz de assegurar.

*Da filosofia moral à filosofia política.
Da motivação individual para a acção
a concepções de sociabilidade e coordenação
das acções de indivíduos – a tradição sentimentalista
(D. Hume e A. Smith)*

O que é que aquilo a que Blackburn chama “a tradição Hume-Smith em filosofia moral” tem a contrapropor à forma racionalista de conceber a motivação para a acção, e, especi-

ficamente, à forma racionalista de conceber a obrigação e a Lei como ‘motivadoras’ para agentes morais humanos e fontes de coesão social? Antes de mais, de acordo com Blackburn, a tradição Hume-Smith em filosofia moral envolve uma ideia razoavelmente optimista da natureza humana. Nomeadamente, parte do princípio de que existe nos humanos um mecanismo de simpatia natural, pelo qual as dores e alegrias dos outros ‘ressoam’ em cada um de nós, e em função do qual naturalmente elogiamos e promovemos acções que promovem prazeres e evitam dores. Mesmo que este mecanismo funcione antes de mais em grupos muito restritos, por exemplo, ligados por laços de parentesco, ele é um primeiro indício do tipo de laços que de alguma forma virão a sustentar sociedades. A ideia é que a partir de um mínimo de virtudes naturais serão possíveis numa sociedade virtudes muito mais artificiais (para falar a linguagem de Hume), tais como a justiça, sem que seja necessário vir a falar de algum pacto ou contrato racional explícito.

Hume, nomeadamente, vê o referido mecanismo de simpatia em funcionamento na forma como os humanos consideram os sentimentos dos outros nos juízos que fazem acerca de si próprios, chegando a afirmar no *Tratado* que as mentes dos homens são como ‘espelhos’ umas das outras. Esse mecanismo incipiente virá a possibilitar a coordenação entre as acções dos indivíduos através de **sentimentos morais** (tais como a vergonha, a culpa, o desejo de ser louvado e admirado, etc.), sem evocações racionalistas e calculadoras de um contrato. Hume exprime esta situação no *Tratado da Natureza Humana* (Livro II, Parte II, Secção II) e na *Investigação acerca dos Princípios da Moral* (Ap. III) com a imagem dos remadores que remam coordenadamente por interesse comum, sem qualquer promessa ou contrato explícito entre si (Hume considera, por exemplo, que é assim que a propriedade é respeitada, que o ouro e a prata são feitas medidas de troca e que os significados das palavras da linguagem são fixadas).

Adam Smith, por sua vez, na sua obra sobre moral e paixões (*The Theory of Moral Sentiments*, 1759), defende as seguintes

ideias acerca da natureza e funcionamento dos sentimentos morais. Existe nos indivíduos humanos uma voz interior, um homem dentro do peito, que os ‘controla’ e faz fazer (aquilo a que normalmente se chama a ‘consciência moral’). Ora, esse escrutinador das acções individuais, esse espectador imparcial como lhe chama Smith, não tem nenhuma origem sobrenatural ou divina. O homem dentro do peito não é nenhum estranho: a sua jurisdição, o seu poder de nos fazer fazer, funda-se no desejo que os indivíduos têm de serem louvados, na sua aversão a serem inculcados e recriminados, no seu desejo de possuírem as qualidades que admiram nos outros e no seu temor de possuírem os traços que neles desprezam. O observador dentro de cada um, supostamente estrutural em todos os indivíduos humanos, está assim ligado à capacidade de ver a acção própria sob uma luz ‘geral’, i. e., como ilustrando determinadas qualidades, que quando encontradas nas outras pessoas despertam amor, admiração, desprezo, ódio, etc. Repare-se que o que está aqui em potência é um mecanismo de ajustamento mútuo das acções através de sentimentos morais individuais. Isto significa que sentimentos morais como a culpa e a vergonha de alguma forma servem para coordenar os humanos entre si, na medida em que ligam as nossas atitudes para connosco com aquilo que os outros pensam de nós.

De acordo com Blackburn²¹, todo este ‘mecanismo’ põe em relevo algo de essencial para os assuntos morais e políticos, a que já Hume, antes de Smith, chamava a nossa capacidade de assumir o ‘**ponto de vista comum**’. O que é esse ponto de vista comum? Algo de diferente da razão kantiana. Para o explicar, Hume chama a atenção para o contraste entre avaliações feitas em função do interesse próprio (por exemplo, considero alguém meu inimigo) e avaliações nas quais espero que os outros concorram (considero essa mesma pessoa viciosa ou depravada). Neste último caso há uma espécie de ‘subida’ para o ponto de

vista comum, que se relaciona com a capacidade de avaliar de uma maneira imparcial. Esta capacidade não é aliás específica da ética (também podemos ser capazes de julgar que um homem que detestamos tem uma magnífica voz ou é um grande pintor, ou um hábil advogado). A ideia de auto-escrutínio, comum a Hume e a Smith e fundamental para despoletar o mecanismo de coordenação de que tenho vindo a falar, não é, como nota Blackburn, de todo nova (o próprio Hume aponta como predecessores os Estóicos, Cícero e Calvino). O que é inovador na tradição Hume-Smith é mais propriamente a ligação entre o auto-escrutínio, a internalização do olhar dos outros, um mecanismo de ajustamento dos indivíduos entre si através dos sentimentos morais, e a sustentação da sociedade enquanto sociedade.

²¹ BLACKBURN 1998: 201.

Capítulo 3

Filosofia e racionalidade teórica: em que devemos acreditar?

*Quem deve estudar a racionalidade –
o cientista cognitivo ou o filósofo?*

A. Goldman e S. Stich

acerca de questões meta-epistemológicas

No capítulo anterior tratei da questão da racionalidade prática, a questão ‘o que devemos fazer?’, e sugeri que a questão envolve assuntos que vão desde a análise daquilo em que consiste uma acção, por contraste com ‘meros eventos’ e movimentos de corpos, até propostas acerca da explicação da acção em agentes individuais e propostas acerca daquilo que agrega e coordena socialmente diferentes agentes. Neste capítulo vou retomar a questão da racionalidade teórica, a questão ‘em que devemos acreditar?’, a partir de uma particular perspectiva, que passo a explicar. Afirmei atrás, no Capítulo 1, que certas teorias são candidatas naturais à tarefa de propor os princípios da racionalidade. A ideia, que procurei sintetizar com a expressão ‘cânones de racionalidade’, é a ideia segundo qual os princípios correctos de raciocínio e decisão são aquilo que são, mesmo se raciocínios e decisões de agentes reais nem sempre são conformes com esses princípios. Noutras palavras, se, nos nossos raciocínios e decisões acontecem ‘erros’, esses erros seriam erros de performance, não

tocando o estatuto dos princípios. De acordo com esta suposição, quando, por exemplo, se ensina lógica, teoria das probabilidades ou teoria da decisão a uma pessoa aquilo que se está a fazer é conduzir a um estado explícito, nessa pessoa, os princípios correctos de raciocínio e decisão. No entanto, a verdade é que falar em cânones de racionalidade é apenas apontar um ponto de referência: nada é dito ainda acerca da realidade cognitiva desses cânones, não sabemos o que é realmente o caso no agente quando este raciocina e decide. É verdade que uma tal questão não tem por que interessar, por exemplo, os lógicos mas ela interessa cientistas cognitivos. Arriscamo-nos então a ter por um lado estudos da forma como as pessoas de facto raciocinam e decidem e por outro lado cânones de racionalidade apontados por filósofos. A pergunta que se coloca é a seguinte: quem estará a caracterizar a natureza da racionalidade, o cientista cognitivo ou o filósofo? A questão não se coloca aliás apenas relativamente à racionalidade, mas relativamente a todas as questões tradicionais da epistemologia.

Vamos supor que alguém se interessa hoje pelas questões epistemológicas tradicionais da filosofia e que coloca a si próprio problemas tais como: *será que podemos conhecer o que quer que seja? quais são as fontes do conhecimento? o que é inato na mente? o que é 'conhecimento'? o que conta como justificação e prova? qual é a melhor teoria da justificação epistémica? Será que existem crenças básicas e não básicas? O que é que essa pessoa devia fazer? Procurar as investigações científicas sobre cognição ou a filosofia? Procurar teorias, modelos e investigações relativas a percepção, memória, raciocínio, decisão, correlatos neuronais de funções mentais, evolução por selecção natural de hipotéticos módulos computacionais da mente, ou textos e argumentos sobre cepticismo, definição de conhecimento, problemas de Gettier, teorias da justificação epistémica (fundacionalismo, coerentismo, fiabilismo, evidencialismo...) e outros temas da literatura epistemológica? É curioso notar que a ciência cognitiva veio admitidamente modificar a forma como se faz filosofia da*

mente, mas nem tanto a forma como se faz epistemologia. No entanto, quando os epistemólogos investigam o conhecimento consideram-no como 'crença verdadeira e justificada', e pelo menos a Condição-Crença desta definição tripartida faz necessariamente recair as discussões em ciência cognitiva e filosofia da mente sobre a epistemologia.

Claro que várias pessoas já pensaram nesta questão (como relacionar explicitamente as questões epistemológicas tradicionais com os estudos científicos da cognição, como formular as modificações para a tradicional tarefa epistemológica que essa relação envolve). Vou falar aqui de duas dessas pessoas, Alvin Goldman e Stephen Stich, essencialmente dos seus livros *Epistemology and Cognition* (1986) e *The Fragmentation of Reason – a preface to a pragmatic theory of cognitive evaluation* (1993). No primeiro Goldman desenvolve uma proposta acerca da natureza da justificação, no segundo Stich desenvolve um argumento contra a epistemologia analítica (nomeadamente contra Goldman) e avança uma proposta divergente. Tudo isto pode ser tomado como uma polémica em torno do que Quine entendeu – ou deveria ter entendido – por 'epistemologia naturalizada' (eu diria que esse é o ponto zero da epistemologia contemporânea) e também como uma discussão acerca da natureza da justificação, e do modo correcto de abordar tal problema (por análise, ou de outra forma qualquer). É conveniente notar que neste ponto, as questões da racionalidade teórica se cruzam com a discussão que no seio da literatura epistemológica se passa acerca daquilo que constitui 'crenças justificadas'. Não vou aqui dar toda a atenção a essa literatura, uma vez a mudança de plano para um segundo nível (das questões epistemológicas acerca de justificação para as questões meta-epistemológicas acerca da forma como se deve fazer epistemologia) nos dá um acesso mais directo às questões da racionalidade. É precisamente aqui que as posições de Goldman e Stich, e o diferendo que existe entre elas relativamente ao que fazemos quando estudamos a racionalidade, constituem uma referência possível.

A. Goldman – a natureza da justificação

Começo por Goldman. Como epistemólogo ele é sobretudo conhecido pelas propostas fiabilistas¹. Aqui interessa-me sua posição meta-epistemológica, de acordo com a qual apenas uma abordagem epistemológica feita à luz das investigações cognitivas nos permite pensar acerca da natureza da justificação. Esta ideia ergue-se contra duas outras formas de investigar a justificação: (i) analisando a linguagem comum e (ii) recorrendo a teorias formais que são candidatas naturais a constituir cânones de racionalidade, tais como a lógica, a teoria da decisão ou a teoria das probabilidades (devo dizer que, ao contrário de Stich, o próprio Goldman não concebe as suas investigações epistemológicas como dizendo centralmente respeito à racionalidade – ele considera ‘racionalidade’ um termo de louvor epistémico bastante vago, preferindo considerar que investiga a natureza da justificação (*justifiedness*)).

Das três condições da definição tripartida de conhecimento (crença, verdade, justificação) é a justificação que constitui o foco de Goldman em *Epistemology and Cognition*. É costume dizer, tendo em conta a definição tripartida de conhecimento, que a condição-Justificação é a condição que garante que, em casos nos quais consideramos existir conhecimento, a condição-Verdade não está apenas acidentalmente ligada com a condição-Crença: não falamos de ‘conhecimento’ nos casos

¹ O fiabilismo foi atrás definido como a ideia segundo a qual uma crença é justificada se e só se é produzida por processos fiáveis do sistema cognitivo, sendo mencionado que a tese é usualmente acompanhada com a referência à ratio de crenças verdadeiras produzidas pelo sistema. O fiabilismo é assim uma teoria da justificação epistémica (alternativa, por exemplo, ao fundacionalismo e ao coerentismo) de acordo com a qual o conceito de justificação é pelo menos em parte causal. Para exemplos importantes de teorias fiabilistas, cf. ARMSTRONG 1973, *Belief, Truth and Knowledge*, DRETSKE 1969, *Seeing and Knowing*, GOLDMAN 1979 *What is Justified Belief?* e NOZICK 1981 *Philosophical Explanations*.

em que algo de verdadeiro é acreditado arbitrariamente, i. e., sem base em provas legítimas, por um sujeito (por exemplo, A acredita *porque isso lhe foi dito por um vidente* que B, o seu patrão, é um criminoso). Quanto às restantes condições para o conhecimento, a crença e a verdade, Goldman pronuncia-se mais brevemente e de forma que podemos à partida nem sequer aceitar: contra posições eliminativistas em filosofia da mente, Goldman considera as crenças estados mentais reais, quanto à verdade, Goldman defende uma concepção realista e não epistémica, de acordo com a qual quando uma determinada asserção é verdadeira ou falsa, aquilo que a torna verdadeira ou falsa transcende o conhecimento e a verificação.

Basicamente Goldman pensa que quando se trata de desenvolver uma teoria epistemológica, a ‘verdade’ necessita de menos explicação do que a ‘justificação’ (aliás, permitirá mesmo compreender o que entendemos por ‘justificação’). De resto, Goldman considera que conhecidas abordagens da verdade e do realismo, tais como as de M. Dummett e H. Putnam, viraram o problema da justificação e da verdade às avessas ao considerarem que aquilo que necessita mais de explicação em metafísica e epistemologia é a verdade, e pressupondo que a justificação e a racionalidade (evocadas na teoria da verdade) não são problemáticas. Apenas uma tal postura torna legítimo definir verdade, como faz, por exemplo, Hilary Putnam em *Razão, Verdade e História*², como ‘idealização da aceitabilidade racional’. Não vou discutir aqui as posições de Goldman quanto a crença e verdade, o que me interessa é a forma como Goldman vê a reformulação da tarefa epistemológica devido ao contacto com a ciência cognitiva e a noção importante para isso é a de justificação. Contrariamente ao que Quine ou Rorty de alguma forma proclamam (Quine 1969, Rorty 1979), para Goldman a tarefa epistemológica é filosófica, a filosofia não é pura e simplesmente afastada do seu papel tradicional pela investigação empírica da cognição. O problema é outro:

² PUTNAM 1992.

temos que compreender e formular o que se segue do facto de admitirmos que as investigações epistemológicas não podem ser levadas a cabo de forma exclusivamente apriorista.

A segunda contribuição de Goldman para a epistemologia, além do fiabilismo, foi a noção de 'epistémica' (definida como o empreendimento que liga a epistemologia tradicional com a ciência cognitiva por um lado e com disciplinas das ciências sociais e das humanidades que exploram os processos interpessoais e culturais que influem sobre crenças e conhecimento por outro.) A epistémica deverá ter, segundo Goldman, duas partes, uma individual e outra social. É a parte individual da epistémica que requer a contínua aportação da ciência cognitiva, daquilo que esta vai descobrindo acerca da arquitectura e dos processos da mente-cérebro. Goldman não tem presunções de originalidade: ele considera como antecedentes da epistémica (mais especificamente da epistémica primária³) a epistemologia psicologista dos séculos XVII-XVIII e a própria proposta quineana de uma epistemologia naturalizada. Apesar destes antecedentes, Goldman considera que grande parte da epistemologia recente foi, por razões várias, dominada por um espírito anti-psicologista. Que razões foram essas? Para Goldman as seguintes: (i) o facto de a epistemologia ser vista como metodologia e portanto como um estudo anterior às ciências e autónomo relativamente a estas, envolvendo, por exemplo, a lógica, a teoria das probabilidades

³ **Epistémica primária** é a expressão que Goldman utiliza para referir a epistemologia normativa individual, que envolve a avaliação dos *processos* psicológicos, por contraste com a **epistémica secundária**, que avalia não já processos mas *métodos*. A distinção entre processos e métodos é importante neste contexto (em que se trata de saber o que se entende por 'epistemologia naturalizada'): processos são características psicológicas básicas da arquitectura cognitiva dos agentes, métodos são formas explícitas de lidar com estes processos, i. e., 'regras para a direcção do espírito', a observar em inquéritos localizados e controlados, por exemplo, quando se faz ciência. A epistémica secundária trataria assim, segundo Goldman, do tipo de problemas metodológicos usualmente considerados como pertencendo ao domínio da filosofia da ciência.

e a estatística, disciplinas nas quais não há necessidade de levar em conta a realidade psicológica da cognição, (ii) o facto de a epistemologia ser vista como análise puramente filosófica de conceitos epistémicos tais como 'conhecimento', 'justificação', etc., (iii) o facto de a epistemologia ser vista como uma dissolução wittgensteiniana de puzzles e paradoxos conducentes ao cepticismo, (iv) o facto de a epistemologia ser vista como normativa, por oposição ao carácter descritivo de ciências empíricas.

É por não aceitar nenhuma destas razões que Goldman considera ser imperativo redirigir o campo da epistemologia. Isso far-se-á começando por admitir que os resultados em ciência cognitiva são relevantes para a epistemologia. O que isto quer dizer é que questões cognitivas relativas à arquitectura da mente e às operações, poderes e limitações da mesma, têm necessariamente que ser levadas em consideração quando se procura delinear princípios de racionalidade e justificação. Mas o facto é que a caracterização da arquitectura e dos processos da mente-cérebro recai em grande medida no âmbito da ciência cognitiva: assuntos tais como a percepção ou a memória ou a evolução de um módulo da mente são tratados pela ciência cognitiva sem qualquer apelo à filosofia. Onde fica então a filosofia segundo Goldman? Ou melhor, o que há de especificamente filosófico no inquérito epistemológico? Segundo Goldman, o que a epistemologia faz é avaliar processos de formação e revisão de crenças e o que faz desta tarefa uma tarefa filosófica é a sua natureza valorativa. Até aqui nada de novo: não é novidade considerar que a epistemologia está interessada em justificação epistémica nem que a noção de justificação encontra o seu foco nas crenças de agentes. A novidade reside no facto de Goldman não se limitar a considerar que há agentes e que estes têm crenças. Ele dá um passo em frente e defende a necessidade de olhar para as crenças não como um ponto de partida não problemático e sim como o resultado de processos cognitivos no seio de uma arquitectura cognitiva determinada. Quando se pergunta a um agente cognitivo 'Como é que sabes?' é da maior importância

para a avaliação de respostas tais como 'Vi', 'Ouvi', 'Recordo-me que..', 'Inferi de A e B que C...', saber o máximo acerca da arquitectura e processos cognitivos desse agente.

Tomemos como exemplo a memória, uma das principais fontes de crenças. Embora a memória não origine crenças, os processos inferenciais reais, pelos quais se formam novas crenças, dependem da disponibilidade na memória dos agentes de determinadas crenças. Ora, o armazenamento na memória humana está sujeito a determinadas condições (por exemplo, distinções entre memória de curto e longo prazo, estrutura compartimentalizada, etc.). Estas limitações e especificações são empiricamente estudadas e constroem a disponibilidade das crenças que constituem ponto de partida de inferências em agentes humanos. Temos assim que a ciência cognitiva faz descrições e caracterizações, neste caso avança com modelos e teorias da memória, e que o objectivo do epistemólogo é compreender as repercussões epistémicas das descrições e caracterizações (por exemplo, analisando o que constitui evidência disponível e acessível a um particular agente num dado momento temporal, e vendo a essa luz as questões que na literatura epistemológica são chamadas 'obrigações de inferência' imputáveis ao agente, 'reunião de evidência epistemicamente responsável', ou mesmo reformulando o que deve entender-se por 'crença', incluindo ou não o armazenamento 'não-activado' de informação). Só assim se poderá dizer, por exemplo, que um agente é epistemicamente 'culpado' (imagine-se, por exemplo, um médico que não aplica os últimos conhecimentos disponíveis ao tratamento de um doente).

A intuição básica é a seguinte: grande parte das caracterizações da racionalidade teórica ao longo da história do pensamento não consideraram as condições e limitações específicas da cognição humana tendo sido portanto 'cognitivamente irrealistas'. Autores como C. Cherniak e G. Harman (Cherniak 1986, Harman 1999), entre outros, sublinharam este ponto e chamaram a atenção para o facto de tarefas aparentemente tão essenciais à racionalidade de um agente tais como a verificação

da consistência das crenças do corpo de crenças por algum processo como tabelas de verdade ou a manutenção em memória da evidência que conduziu originalmente a cada uma das crenças do corpo de crenças, podem ser computacionalmente intratáveis ou cognitivamente inexecutáveis. Assim sendo, elas não são razoavelmente exigíveis aos agentes, não devendo estes ser considerados irracionais por não as executarem.

Temos então a divisão de tarefas: (i) a ciência cognitiva obriga o epistemólogo ao 'realismo' quanto a arquitectura e processos cognitivos de agentes, (ii) só então se põe a questão normativa, a questão do direito que os agentes têm às suas crenças. Essa é a parte filosófica da empresa epistemológica. Definida esta concepção meta-epistemológica, a ideia de Goldman acerca dos objectivos da própria epistemologia é, em traços largos, a seguinte. O projecto da epistemologia é desenvolver uma teoria que explique o que entendemos por 'justificação'. Goldman pensa que existe um núcleo comum àquilo que consideramos 'justificado' e é esse núcleo comum que tenta capturar. Como? Analisando a justificação em termos de regras: o primeiro passo de uma teoria da justificação é articular o sistema de regras justificativas (a que Goldman chama **Regras-J**⁴) que permitem ou proíbem crenças, que especificam os modos como um agente cognitivo pode mover-se ao formar ou actualizar estados cognitivos. As Regras-J 'permitem' e 'proíbem' não no sentido de fornecerem instruções que um agente cognitivo voluntária e conscientemente deve seguir na formação de crenças (até porque a formação e actualização de crenças não está assim tão submetida à vontade do agente – pense-se nas crenças perceptivas) mas no sentido em que capturam a semântica

⁴ De então para cá, Goldman reformulou e sofisticou a sua posição. A proposta de *Epistemology and Cognition*, feita em termos de Regras-J, parece-me no entanto apropriada para pôr em relevo as questões em causa (para além de que constitui uma boa entrada para os posteriores tratamentos da natureza da justificação por Goldman, os quais podem ser entendidos como pretendendo resolver os problemas desta proposta).

da linguagem de avaliação epistémica, o sabor deôntico desta: quando dizemos que uma crença é 'justificada' o que estamos a fazer é considerá-la um estado cognitivo apropriado, algo a que o agente tem direito (*is entitled to*). É esse 'direito a', o 'ser permitido', o 'ser apropriado', que as Regras-J procuram formular explicitamente. As Regras-J não são regras para a direcção do espírito, instruções cuja aplicação voluntária e explícita o epistemólogo aconselharia (não são 'regras do método', nos termos da diferença entre processos e métodos explicada na nota 4). O objectivo do *rule framework* da epistemologia de Goldman é analisar ou explicar (*explicate*) aquilo de que falamos quando falamos de justificação, não é 'fornecer instruções'. A formulação das Regras-J é um passo preparatório: diferentes pessoas podem ter ideias totalmente diferentes sobre os processos que geram crenças justificadas, o que as conduz a fixar crenças de acordo com conjuntos de Regras-J diferentes ou incompatíveis, ou mesmo Regras-J absurdas (por hipótese: Todas as crenças formadas à terça-feira são justificadas, ou Todas as crenças baseadas no testemunho do homem de preto que vive no centro da cidade C são justificadas). O passo seguinte é investigar quais são as regras correctas (*right*). Mas como saber qual sistema de Regras-J, dos vários que são formuláveis, é o correcto? Para enfrentar esse problema Goldman faz apelo, e esse é um dos núcleos da proposta epistemológica de *Epistemology and Cognition*, a um critério de ordem superior a que chama *criterion of rightness* (**critério de correcção**). Um critério de correcção é um standard que especifica condições factuais, substantivas, para um conjunto de Regras-J ser correcto. Essas condições não podem, de acordo com Goldman, evocar noções epistémicas (como por exemplo, 'evidência' ou 'provas'). Para termos um exemplo de critério de correcção na relação com regras consideremos um exemplo moral:

Um sistema R de regras morais é correcto se e só se R for um sistema decretado por Deus⁵.

⁵ GOLDMAN 1986:64.

Em *Epistemology and Cognition* Goldman considera os seguintes critérios de correcção alternativos para um sistema R de Regras-J:

- (i) ser derivável da lógica e teoria das probabilidades
- (ii) ser escolhido por alguém que acreditasse em todas as verdades da lógica e teoria das probabilidades e fosse ignorante de todos os factos contingentes
- (iii) ser aceite pelos jogadores do nosso jogo de linguagem (Wittgenstein)
- (iv) ser aceite pelos membros da nossa matriz disciplinar (Kuhn)
- (v) ser aceite pelos nossos pares (Rorty)
- (vi) garantir um conjunto coerente de crenças
- (vii) permitir atitudes doxásticas proporcionais à força da evidência disponível
- (viii) ser tal que a conformidade com R maximiza o número total de crenças verdadeiras do agente

O problema torna-se, neste ponto, saber como decidir entre critérios de correcção alternativos. Goldman propõe-se fazer isso através do que chama a 'concepção Goodman-Rawls dos juízos considerados em **equilíbrio reflectido**'. Talvez seja conveniente recordar, antes de mais, o problema para o qual o equilíbrio reflectido pretendeu historicamente constituir uma solução. O problema era o seguinte: como podem standards cognitivos ser descobertos e defendidos? qual é a natureza desses standards? Como justificamos, por exemplo, uma dedução? Como justificamos uma inferência indutiva? A sugestão do proponente original do equilíbrio reflectido, o filósofo Nelson Goodman (em *Fact, Fiction, Forecast*, 1965), foi a seguinte. Antes de mais, aquilo que é visado é a análise da noção comum de inferência justificada. Goodman descreve depois um processo, um teste, que forneceria essa análise e que consistia em pôr de acordo uns com os outros *juízos acerca de particulares inferências e princípios gerais de inferência*. Propõe então que toda a justificação neces-

sária e possível para princípios inferenciais residiria nesse acordo. J. Rawls, a quem se deve a expressão *reflexive equilibrium* para nomear o processo de ajustamento, utilizou, na sua teoria da justiça, uma versão do processo de Goodman para caracterizar a relação entre princípios e juízos morais particulares.

Resumamos: a ideia básica do equilíbrio reflectido é a ideia segundo a qual (i) se explica a justificação analisando o uso comum de ‘inferência justificada’, (ii) a explicação ou análise faz-se através de um teste que consiste num processo de ajustamentos mútuos entre regras gerais formuladas e inferências particulares aceites. Uma regra que mantemos será emendada se produzir uma inferência que não estamos dispostos a aceitar, uma inferência será rejeitada se violar uma regra que não estamos dispostos a emendar. Voltando a Goodman e à formulação original do equilíbrio reflectido: dizemos, por exemplo, que uma dedução (uma inferência particular) se justifica pelo facto de ser levada a cabo conforme regras gerais de inferência dedutiva. Um argumento que esteja de acordo com as regras é considerado válido mesmo que a conclusão seja falsa, um argumento que viole as regras é considerado falacioso mesmo que a conclusão seja verdadeira. Mas o que justifica as regras elas próprias? Não é suposto que estas sejam arbitrárias, mas sim de alguma forma também elas válidas. Poderíamos proclamar que as regras são evidentes ou que elas estão inscritas na própria natureza da mente humana. Mas o interesse da proposta de Goodman é a ideia segundo a qual a resposta ‘está mais próxima da superfície’ – é esse o espírito da proposta do equilíbrio reflectido.

A sugestão de Goldman vem nesta linha: ele propõe-se (i) examinar sistemas de Regras-J que seriam gerados por cada um dos critérios candidatos acima listados, (ii) reflectir sobre as implicações desses sistemas de regras relativamente a particulares juízos de justificação e (iii) verificar se os juízos estão de acordo com as intuições pré-teóricas. Desta forma, um critério será fortalecido ou enfraquecido. Como é característico do equilíbrio reflectido, não é suposto tomarmos as intuições como uma úl-

tima palavra, uma vez que elas podem ser ajustadas reflectindo sobre os critérios. De qualquer forma, um sistema de regras considerar-se-á justificado se passar o teste do equilíbrio reflectido (é importante notar que Goldman diz abertamente que este não é o nível no qual a psicologia ‘é chamada’ à epistemologia – a psicologia só entra se algum dos critérios candidatos fizer apelo a processos psicológicos).

Goldman examina os critérios enumerados e afasta-os por diferentes razões. No caso dos Critérios (i) e (ii) a razão fundamental é a incompletude: o máximo que regras lógicas podem fornecer é um subconjunto das regras justificativas, para inferências dedutivas. E mesmo que com a ajuda da teoria das probabilidades se pudesse chegar a propor regras justificativas para o âmbito da indução (embora a epistemologia da teoria das probabilidades seja notoriamente problemática), certamente nem a lógica nem a teoria das probabilidades teriam algo a propor quanto a regras justificativas para o âmbito da percepção e da memória, fontes essenciais das crenças humanas. Aliás e em geral, regras lógicas e verdades da teoria das probabilidades nada afirmam acerca de formação de crenças, nem, de resto, acerca da realidade psicológica do raciocínio. Os Critérios (i) e (ii) teriam ainda que lidar com os seguintes problemas: porquê dar à lógica e à teoria das probabilidades um estatuto especial numa teoria da justificação? Isso quer dizer que as verdades dessas teorias são verdades necessárias? Cognoscíveis a priori? Tudo isso é problemático. Quanto aos Critérios (iii-v) eles partilham uma perspectiva social e ‘local’, associando ‘justificação’ àquilo que é aceite numa dada comunidade. Mas o que garantiria que as Regras-J admitidas pelo critério conferissem justificação? Nomeadamente, não poderia a particular comunidade a que o agente pertence estar errada? Quanto ao Critério (v) é duvidoso que a coerência seja um requisito suficientemente forte para a justificação. Quanto ao Critério (vi), a sua formulação envolve o termo epistémico ‘evidência’, o que vai contra os constrangimentos impostos pelo próprio Goldman. O Critério (vii), um critério

fiabilista formulado em termos de número total de crenças, enfrenta a mesma objecção que todas as teorias fiabilistas ao empregar a 'verdade' para capturar a natureza da justificação. Goldman virá no entanto a aceitar uma versão 'afinada' de critério fiabilista como sendo o critério de correcção que se adapta à noção de 'justificação' envolvida no pensamento comum ou na linguagem comum (ou no núcleo comum à várias concepções que, segundo Goldman, encontram um certo apoio nas nossas intuições). Goldman formula-o assim:

CRITÉRIO DE CORRECÇÃO: um Sistema R de Regras-J é correcto se e só se R permite certos processos psicológicos básicos e a instânciação destes processos resulta numa *ratio* de verdade de crenças que satisfaz um determinado limiar (superior a 50%).

Goldman justifica cada elemento da sua proposta, explicando nomeadamente o que entende por processos (ao contrário de algoritmos, heurísticas e em geral métodos susceptíveis de aprendizagem, os processos são relativos a cadeias causais da cognição), a razão porque aceita argumentos a favor do consequencialismo na avaliação de processos, etc. Saliento apenas que o critério finalmente proposto é um critério fiabilista, e que a acusação mais comum às teorias fiabilistas da justificação é a acusação de circularidade (pelo facto de utilizarem a 'verdade' para dar conta da 'justificação').

Depois de tratar as questões fundacionais da sua epistemologia, Goldman pode passar à prática da epistémica primária e portanto à análise das repercussões epistemológicas de particulares investigações em ciência cognitiva (relativas a percepção, memória, formato de representações, raciocínio, etc.). É isso que faz na segunda parte ('psicológica') de *Epistemology and Cognition*. Mas o que me interessa aqui não são tanto os ensaios de epistémica que Goldman tem vindo a produzir ao longo de toda a sua obra, e sim a forma como ele, apesar da sua concepção meta-epistemológica anti-apriorista, evoca finalmente algo que se pode considerar análise conceptual para capturar a natureza da justificação.

S. Stich – contra a 'epistemologia analítica': justificação e pragmatismo

Pelo menos é isso que pensa Stich, para quem a epistemologia de Goldman é um exemplo de 'epistemologia analítica'. A 'epistemologia analítica' é definida como qualquer projecto epistemológico baseado na ideia de acordo com a qual a escolha entre regras justificativas ou critérios de correcção concorrentes depende de análise conceptual ou linguística e é o alvo da teoria pragmatista da avaliação epistémica de Stich. A principal razão pela qual Stich se opõe à epistemologia analítica é o facto de ele considerar que a avaliação epistémica é uma questão prática: numa teoria da avaliação epistémica trata-se de dizer 'quais processos cognitivos são bons', 'quais processos cognitivos as pessoas devem utilizar'. Em contrapartida para Goldman (pelo menos aos olhos de Stich) a questão da avaliação epistémica é uma 'questão teórica', uma questão conceptual acerca da natureza da justificação, em que se trata de 'revelar', por análise, o conceito de justificação implícito no pensamento e linguagem comuns. Para Stich uma tal forma de conceber a teoria da avaliação epistémica faz da investigação epistemológica um empreendimento conservador e paroquial.

Este posicionamento de Stich reflecte-se na forma como ele olha para o campo histórico da epistemologia. Ele considera que a epistemologia se faz tradicionalmente de pelo menos três maneiras: (i) como avaliação dos métodos do raciocínio e do inquérito (foi o que fizeram Bacon, Descartes, Mill, Carnap, Popper), (ii) como análise da noção de 'conhecimento' (de Platão a Gettier), (iii) como resposta aos argumentos cépticos (de Descartes a Moore). Ora, Stich pensa que (ii) e (iii) são vias de certa forma bizarras, e que apenas (i) tem possivelmente implicações importantes para a maneira conduzir os nossos assuntos cognitivos comuns e científicos.

O que devemos então fazer quando fazemos epistemologia, segundo Stich? Até certo ponto, exactamente o que propunha Goldman: compreender a natureza das avaliações epistémicas.

Mas Stich considera que isso se faz investigando o que fazemos quando consideramos determinados sistemas cognitivos superiores a outros, ou quando consideramos que alguém está a raciocinar mal. Para este foco de análise (trata-se de saber, para os agentes, como conduzir os assuntos cognitivos) a natureza conceptual da justificação não importa, o que importa é o facto de os processos cognitivos poderem eles próprios ser avaliados instrumentalmente. A ligação da epistemologia com as investigações empíricas, nomeadamente em psicologia do raciocínio, entronca aqui.

Stich vê os processos cognitivos de agentes como instrumentos ou tecnologias empregues no decurso da tentativa de satisfazer desejos e alcançar finalidades. Esses instrumentos podem ser avaliados na prática. Avaliar processos cognitivos na prática significa avaliá-los em função do sucesso que têm a provocar estados de coisas desejados, estados que as pessoas intrinsecamente valorizam (tais como ser capaz de prever e controlar a natureza, ou contribuir para uma vida interessante e cheia). Isto pressupõe que as pessoas valorizam determinadas coisas intrinsecamente, que estas coisas são várias, mas não incluem em geral 'a justificação ou a racionalidade tais como estas são concebidas no pensamento e linguagem comuns'. Para Stich os 'standards epistémicos comuns' (a justificação ou a racionalidade tais como são concebidas no pensamento e linguagem comuns, e que constituem o foco da teoria da justificação de Goldman) são paroquiais e culturalmente variáveis e não devem constituir o foco da teoria da avaliação epistémica. À luz da abordagem pragmatista e consequencialista da avaliação epistémica de Stich, a proposta de Goldman, apesar da intenção explícita de reformular o campo da epistemologia à luz das investigações cognitivas, não passou ainda ela própria por reformulação. Só isso explica a ideia de remeter à análise conceptual a incumbência de explicitar a natureza da justificação.

Que alternativa apresenta Stich? Para responder a esta questão vou resumir brevemente a forma como Stich chega à sua teoria da racionalidade ou da avaliação epistémica. O incentivo prático

imediate foi um pedido de ajuda vindo da psicologia, nomeadamente do psicólogo Richard Nisbett. O que o psicólogo Nisbett quis saber do filósofo Stich – e podemos pensar na questão de Nisbett como uma questão geral que o estudo psicológico do raciocínio coloca à filosofia – foi o que queremos dizer quando dizemos que uma inferência é justificada. Sem uma resposta disponível é impossível interpretar os resultados da psicologia do raciocínio. Nisbett tinha por trás investigações que mostravam sistemáticas falhas na tarefa de selecção (*selection task*), inicialmente estudada por P. Wason e P. Johnson-Laird, a frequência da chamada 'falácia da conjunção' revelada pelos estudos de A. Tversky e D. Kahneman, a sistematicidade de 'pseudo diagnóstico' dada a ignorância de probabilidades prévias, a perseverança de crenças mesmo após a revelação da intenção de iludir os sujeitos da experiência (*debriefing*), etc.

E o que os resultados da psicologia do raciocínio pareciam revelar era a violação sistemática de 'princípios de racionalidade', por exemplo, princípios da lógica (caso da tarefa de selecção, em que estão em causa condicionais) e da teoria das probabilidades (falácia da conjunção). Na medida em que a lógica e a teoria das probabilidades constituem candidatos óbvios a 'cânones de racionalidade', os resultados foram interpretados como revelando a irracionalidade do raciocínio comum. Stich estava inicialmente inclinado a subscrever o veredicto de irracionalidade emitido pelos psicólogos. Não teve no entanto qualquer sucesso ao procurar fundamentá-lo. Veio por isso a recuar na sua posição. Vamos directamente ao que interessa: que direito temos nós a afirmar que determinadas pessoas estão a raciocinar mal em determinadas circunstâncias? Repare-se que perguntar o que é uma inferência justificada é muito diferente de perguntar o que é um argumento válido. No último caso tudo é muito mais claro, nomeadamente porque questões psicológicas, evidenciais e ambientais, questões quanto a fontes de crenças, recolha de provas, adaptação do comportamento de agentes ao mundo, pura e simplesmente não se colocam. A lógica simplesmente

não é uma teoria do raciocínio, não nos diz – nem tem por que dizer – em que devemos acreditar, o que devemos tomar como premissas em argumentos, que evidência devemos aceitar. Os lógicos não pretendem dar regras para a formação de crenças, ou uma teoria do raciocínio já feita. De facto não encontramos disponível nenhuma teoria do raciocínio (Goldman diria que, a estar disponível, ela seria um ramo da epistemologia).

Voltando à questão colocada pelos psicólogos, o que faz com que uma inferência seja justificada? Uma hipótese inicialmente colocada pelo próprio Stich e também por muitos outros autores – nomeadamente por Goldman como vimos – é a utilização do teste do equilíbrio reflectido para responder a essa questão. Recordo o que isso significa: tomando como referência a noção comum de ‘justificação’, atestamos o ‘conteúdo’ desta através de um processo de delicados ajustamentos mútuos entre regras formuladas e inferências aceites. Stich recusa no entanto essa via. Vou procurar resumir as suas objecções.

Objecção 1: É contestável que seja possível seleccionar como objecto de análise, no ‘pensamento e linguagem comuns’, uma noção única e coerente de ‘justificação’. Nada nos garante que ‘o conceito comum de justificação’ tenha essas características (de resto, tanto quanto a psicologia dos conceitos nos permite pensar, podem, por exemplo, existir na mente do agente, exemplares prototípicos de justificação que são considerados em cada decisão de considerar uma inferência justificada, sem que existam propriedades comuns a tudo o que é considerado ‘justificado’). De qualquer forma isso só poderá ser descoberto empiricamente e portanto não temos direito de assumir à partida uma ‘noção comum de justificação’.

Objecção 2: Mesmo que a análise do ‘conceito comum único e coerente de justificação’ fosse possível, ela seria inútil, no sentido em que não nos facultaria um critério defensável para distinguir boas de más estratégias inferenciais.

Imaginemos que aceitamos a ideia do teste de equilíbrio reflectido. O primeiro passo prático seria determinar que regras

inferenciais poderiam resultar do processo de ajustamento mútuo entre inferências aceites pelos agentes e regras explicitadas. Ora, Stich pensa, e essa é talvez a sua objecção fundamental, que princípios inferenciais estranhos poderiam passar o teste do equilíbrio reflectido para um determinado agente cognitivo. Pense-se, por exemplo, na falácia do jogador. Imaginemos alguém que pensa: ‘Nas últimas 100 vezes que lancei o dado não me saiu um 6; portanto, o 6 deve estar mesmo a sair’. Temos um indivíduo que intuitivamente julga assim. Quando se articula explicitamente a regra ‘A probabilidade de ocorrência de um evento x aumenta com cada não-ocorrência do evento x’, esse indivíduo aceita que se trata de uma boa regra. Não podemos afirmar a priori que tal não sucederia e a ser assim, estranhas inferências, que não queremos intuitivamente admitir como justificadas, seriam admitidas como justificadas: de acordo com a concepção de equilíbrio reflectido se uma inferência passa o teste do equilíbrio reflectido ela é justificada (é suposto o equilíbrio reflectido ser *constitutivo* da justificação). Stich considera que esta possibilidade é suficiente para duvidarmos da ideia segundo a qual o equilíbrio reflectido é constitutivo da justificação.

Se o teste do equilíbrio reflectido não nos oferece um critério para distinguir boas de más estratégias inferenciais e nos obriga mesmo a considerar como boas para um dado agente estratégias inferenciais que queremos considerar como más, qual é a alternativa? A alternativa ao apelo a conceitos é o apelo a consequências. Essas consequências não são no entanto segundo Stich ‘crenças verdadeiras’ – Stich não vê, à maneira fiabilista comum, que encontramos, por exemplo, em Goldman, os processos cognitivos como processos-para-gerar-verdades. Vê-os antes como tecnologias que podem ser utilizadas para alcançar uma variedade de finalidades de agentes. E podemos afirmar com segurança que os agentes se preocupam com alcançar as suas finalidades (por contraste com a falta de segurança de afirmar que os agentes se preocupam com ‘crenças verdadeiras’

ou com 'a análise do conceito de justificação implícito nas suas práticas de raciocínio e decisão').

Olhar para os processos cognitivos como ferramentas mentais é obviamente uma ideia pragmatista, que se tornou central no pensamento de Stich quando este se persuadiu de que o projecto goodmaniano tinha as deficiências apontadas. É a ideia pragmatista que leva Stich a negar que a racionalidade, a justificação e a verdade tenham em si mesmas valor para o agente cognitivo. Antes de subscrevermos críticas pouco caridosas, como as de Susan Haack, à forma como Stich trata a verdade na sua teoria da avaliação epistémica (Haack pensa que Stich suja o nome do pragmatismo ao dizer que 'a verdade não importa'⁶), convém termos claro o que Stich não está a dizer acerca da verdade. Stich não está a dizer que 'a verdade não interessa', apenas que a verdade não é intrinsecamente valiosa para o agente cognitivo enquanto agente cognitivo, i. e., que aquilo que o agente visa quando age não é obter crenças verdadeiras. Stich aceita perfeitamente que muitas pessoas diriam que valorizam o facto de ter crenças verdadeiras, e talvez dissessem mesmo que consideram ter crenças verdadeiras como intrinsecamente valioso. Mas provavelmente não chegariam a dizer nada de coerente sobre o que é 'a verdade', i. e., seriam incapazes de dizer exactamente o que é isso que valorizam. Isto parece natural: afinal, há muitos séculos se discute a natureza da verdade. Do facto de as pessoas – enquanto agentes cognitivos comuns, o que nos inclui a todos – não conseguirem em geral dizer nada de coerente sobre a verdade não se segue que nada se possa dizer acerca da verdade. É possível dizer muita coisa acerca da verdade, nomeadamente em teorias formais, em metafísica, em filosofia da mente, etc. Imaginemos que temos – não, evidentemente, fornecida pelos agentes cognitivos, mas elaborada por lógicos, filósofos da linguagem e filósofos da mente – uma boa teoria formal da verdade e uma boa teoria psicosssemântica. O

⁶ HAACK 1993: 195-202.

que Stich quer dizer é que essas teorias não têm importância para os agentes individuais nos seus afazeres cognitivos, o que não significa de modo algum que esses afazeres cognitivos não são importantes para o agente, ou que as teorias não são importantes. Significa apenas que o que o agente cognitivo quer com os afazeres cognitivos é atingir os seus fins, sejam estes quais forem, e não atingir 'a verdade', e isso deve ser considerado numa teoria da avaliação epistémica.

Vou procurar exemplificar. Se, como Stich aceita em traços gerais, espécimes de crenças são estados cerebrais que têm valores de verdade, os quais são plausivelmente explicáveis através do apelo a uma função de interpretação que mapeia estados-cerebrais-de-crença numa classe de entidades semanticamente avaliáveis (como proposições, situações, mundos possíveis, etc.), a pergunta que devemos fazer é esta: se é nisso que consiste o facto de uma crença ser verdadeira, será que realmente importa ao agente cognitivo que as suas crenças sejam verdadeiras? A resposta, segundo Stich, é negativa: se essa é a natureza das crenças verdadeiras, não há nada de especial, para o agente, em ter crenças verdadeiras. O resultado para a teoria da avaliação epistémica desta posição acerca do 'valor da verdade' é uma posição pragmatista, de acordo com a qual todo o valor cognitivo é instrumental, não há valores cognitivos intrínsecos. Não podemos falar portanto de estratégias de inferência ou raciocínios 'intrinsecamente bons', mas apenas de mecanismos cognitivos, que são instrumentos a avaliar da mesma forma que outros instrumentos, i. e., pelas suas consequências práticas. Um sistema de processos cognitivos será preferível a outro se, ao usá-lo, for mais provável atingirmos as coisas que queremos. Essas coisas podem ser várias: é a nota pluralista da teoria.

Stich não aderiu facilmente ao pragmatismo que hoje defende. Parecia-lhe que objecções óbvias ao pragmatismo tinham um grande peso, nomeadamente as acusações de (i) relativismo e (ii) circularidade. Não se pode negar que um relato pragmático da avaliação cognitiva é de alguma forma relativista: a avaliação

de sistemas de processos cognitivos é sensível às finalidades e circunstâncias dos agentes que os utilizam. Mas porque é que o relativismo é uma coisa má? Procurando argumentos contra o relativismo, o que Stich encontra é sobretudo o medo de cair nas mãos de 'nilistas epistémicos', pessoas que desistiram de separar boas e más estratégias cognitivas. Mas a posição de Stich não é um **nilismo epistémico**: conceber a avaliação epistémica de forma pragmática não é uma posição mais nilista do que conceber de forma pragmática estratégias de investimento ou técnicas de engenharia. Em nenhum dos casos se trata de encontrar O Bom Raciocínio, A Boa Opção de Investimento, A Boa Técnica de Engenharia, mas de olhar para processos como estratégias que podem ser mais ou menos bem sucedidas na prossecução de finalidades e que podem ser melhoradas.

O relativismo é habitualmente acusado de conduzir ao cepticismo. Se sistemas de raciocínio diferentes podem ser preferíveis para finalidades diferentes e se eles geram crenças significativamente diferentes com base no mesmo input sensorial, então presumivelmente pessoas diferentes podem acabar com crenças diferentes, mesmo partindo da mesma evidência. Stich admite que o relativismo da teoria torna impossível defender que o bom raciocínio conduz à Verdade. Isso não lhe parece no entanto constituir causa de preocupação, a não ser que tenhamos alguma razão para querer que os nossos sistemas cognitivos produzam crenças verdadeiras. Ora, vimos que Stich pensa que não temos tal razão. Para Stich, o cepticismo simplesmente assume a importância de um dado 'bem epistémico' (a verdade) que afirma, ao mesmo tempo estar fora de alcance. Logo, a melhor resposta ao céptico que mantém que não conseguimos atingir a Verdade é: e depois?

Relativamente à circularidade, Stich vem a adoptar argumentos que serviram a Goldman em *Epistemology and Cognition* para defender o fiabilismo da mesma acusação. Quer o pragmatismo de Stich quer o fiabilismo de Goldman envolvem o objecto de explicação na explicação que fornecem dos processos cognitivos.

A defesa perante a acusação segundo a qual esse envolvimento peca por circularidade consiste em admitir que embora na avaliação de sistemas ou processos cognitivos (i. e., na investigação de qualquer critério de correcção) se utilize inevitavelmente processos cognitivos, isso não nos impedirá de concluir que determinados processos e sistemas cognitivos são superiores a outros. Além da defesa contra a acusação de circularidade, Stich vê uma outra característica decisiva na análise que Goldman faz da justificação: ele considera que o próprio Goldman nos mostra que o conceito de justificação que constitui o foco da epistemologia analítica é paroquial e local. Para Stich o conceito comum de justificação é arbitrário e idiossincrático, apenas um membro de uma família maior de noções tais. Mas como fica então a justificação, se nos deixarmos persuadir de que o conceito comum de 'justificação' sob análise, nomeadamente na epistemologia de Goldman, é idiossincrático e paroquial? Poderemos sequer querer mais do que isso? E como é que a admissão do paroquialismo e localidade de noções de justificação alguma vez nos permitirá responder à pergunta colocada à filosofia pelos psicólogos do raciocínio? Stich vê a situação da seguinte maneira. Parece-lhe claro que uma teoria pragmatista da avaliação cognitiva não pode dar uma resposta geral à questão 'os sujeitos estão a raciocinar bem ou mal?'. Mas isso não torna impossível uma teoria da avaliação cognitiva. Porquê? Porque um relato pragmatista das 'virtudes cognitivas' é um relato comparativo: apenas é possível dizer se um dado sistema cognitivo se está a sair melhor ou pior do que outro sistema cognitivo relativamente a uma dada finalidade. Não é possível dizer em abstracto que um sistema cognitivo é um bom sistema cognitivo.

Isto não nos impede de defender que um dado agente está a raciocinar mal numa dada circunstância. O que temos que fazer é ver o veredicto de mau raciocínio como a apologia de uma alternativa possível. Mas exactamente o que constitui uma alternativa, em termos de processo de raciocínio, para um dado agente cognitivo? – seria perverso considerar que os sujeitos

estão a fazer um mau trabalho de raciocínio comparando as suas estratégias com estratégias que requerem um cérebro que eles não têm. A comparação que interessa numa avaliação epistémica é a comparação com alternativas exequíveis. Mas como sabemos o que são alternativas exequíveis? Para Stich, de um ponto de vista pragmatista não podemos senão considerar que a exequibilidade depende das finalidades do sistema e das tecnologias disponíveis.

Podemos neste momento formular algumas conclusões de uma teoria pragmatista da avaliação epistémica.

Conclusão 1. Questões acerca da qualidade do raciocínio não podem ser respondidas de forma abstracta, para um agente cognitivo qualquer, independentemente de arquitecturas e processos cognitivos e das finalidades desse agente. A pergunta abstracta e geral acerca de bom raciocínio faz sentido na epistemologia tradicional mas não na epistemologia feita em contacto com a ciência cognitiva.

Conclusão 2. Se não tem sentido colocar uma questão abstracta quanto a racionalidade ideal, tem todo o sentido colocar uma questão prática relativa ao melhoramento das performances cognitivas de agentes⁷. O que se faz em epistemologia e nomeadamente quando

⁷ Em que consistiria o melhoramento das capacidades de raciocínio? Não poderá dar-se o caso de ser "psicologicamente impossível" melhorar a nossa capacidade de raciocínio? É esclarecedor pensar aqui em termos de línguas naturais e linguagens artificiais. Nós aprendemos a usar linguagens artificiais como o BASIC e o LISP, linguagens que violam princípios que, segundo um chomskyano, todas as linguagens naturais deverão satisfazer. Mas não podemos dizer que aprendemos usamos uma linguagem artificial da mesma forma que aprendemos ou usamos o inglês ou o árabe. São necessários esforço e treino específicos e a linguagem é utilizada apenas em certas circunstâncias. Algo de análogo poderá passar-se em termos de raciocínio: pode ser possível aprender certas regras de raciocínio que violem alguns dos princípios "naturais" ou "psicologicamente possíveis", e usar essas regras em circunstâncias específicas ao mesmo tempo que quando o agente se depara com os desafios ecologicamente naturais para o tipo de agente que é utiliza

se analisa avaliação epistémica pode e deve ter implicações na nossa maneira de conduzir os nossos assuntos cognitivos.

Se o melhoramento das performances cognitivas de agentes for a nossa finalidade quando comparamos sistemas cognitivos, então sim teremos espaço de manobra: alternativas a considerar são aquelas que agentes poderiam ser levados a utilizar. Quais são exactamente não o poderemos saber sem investigação empírica. Levar a cabo essa investigação empírica acerca de alternativas que poderíamos levar pessoas a utilizar parece-me ser a ideia por trás do trabalho prático de filósofos na criação de software para melhoramento e treino de raciocínio (veja-se, por exemplo, o trabalho de Tim Van Gelder, na Universidade de Melbourne (Reason! Research Project, <http://www.philosophy.unimelb.edu.au/reason>)). A obrigação de 'realismo' quanto a arquitecturas e processos de agentes é certamente uma das consequências importantes do cruzamento de investigações cognitivas com questões epistemológicas. Sabemos que estratégias cognitivas muito custosas estão para além daquilo que cérebros como os nossos conseguem realizar. No entanto, o cruzamento de investigações cognitivas com questões epistemológicas permite-nos ver a uma nova luz as situações de limitação (limitação ao nível dos processos, como diria Goldman): nada exclui a possibilidade de treino e inculcação de hábitos cognitivos, nomeadamente o aperfeiçoamento de métodos. É claro que isso torna ainda mais importante – e essa é uma questão para a epistemologia e para a ciência cognitiva – saber, por exemplo, se padrões de inferência determinados têm estatuto de método ou de processo no tipo de agentes cognitivos que somos. A Goldman nota algo de aparentemente tão óbvio como isto: «o *modus ponens* não é mais válido do que outras formas válidas

"heurísticas" menos exigente. (É este o ponto que S. J. Gould torna claro, ao evocar o pequeno homúnculo saltitante na sua cabeça, e isso pode explicar o facto, de outro modo surpreendente, de estudantes de pós-graduação de um prestigiado programa de ciência da decisão não serem melhores do que o resto das pessoas no que toca a evitar a falácia da conjunção).

de argumento. É apenas psicologicamente mais simples do que outras.⁸ Essa constatação é importante não para a lógica mas para a epistemologia.

Voltando às relações entre a filosofia e a psicologia: o resultado de tudo isto é que não é nada claro que os críticos do psicólogo Richard Nisbett estivessem errados quando lhe respondiam que ele não tinha fundamento para afirmar que os sujeitos dos seus estudos estavam a raciocinar mal. Está longe de ser óbvio que os sujeitos estão a raciocinar mal quando ‘falham’ na tarefa de selecção ou incorrem na falácia da conjunção (eles poderiam estar, em alternativa, a utilizar estratégias heurísticas que teriam evoluído para lidar com determinadas tarefas comuns no ambiente da espécie e que seriam cognitivamente mais determinantes do que as leis da lógica; investigações em psicologia evolutiva fornecem o ponto de partida de argumentos nesse sentido). Se queremos argumentar que sujeitos estão a raciocinar mal, teremos que mostrar que existe uma estratégia para atingir as suas finalidades que é superior e que é exequível com o equipamento cognitivo de que dispõem. Uma última conclusão se impõe.

Conclusão 3. É uma consequência do pragmatismo que as questões epistemológicas tradicionais ficam com o destino ligado às investigações empíricas da cognição e portanto à psicologia e à tecnologia.

Goldman tem toda a razão em defender que arquitectura e processos cognitivos são fulcrais numa teoria da avaliação epistémica feita à luz da ciência cognitiva. Mas ao olhar para arquitectura e processos cognitivos de agentes humanos devemos olhar ao mesmo tempo para o passado e para o futuro. Parece-me por isso que a forma correcta de considerar a Conclusão 3 é vê-la como afirmando que (i) arquitectura e processos podem ter ‘razões’ evolutivas específicas e (ii) poderão vir a ser

⁸ GOLDMAN 1986: 89.

suplementados. Há uma tradição em epistemologia que rejeita a dependência das questões epistemológicas relativamente a assuntos ‘impuros’, tais como a psicologia e a tecnologia. Essa, no entanto, é uma tradição que Stich considera estéril e moribunda. Outra tradição epistemológica mais recente, ligada aos pragmatistas J. Dewey e W. James, que Stich quer reclamar como predecessores, considera que a epistemologia está naturalmente ligada à ciência e à tecnologia. No espírito dessa tradição, investigações acerca de racionalidade acompanharão, por exemplo, as investigações da psicologia evolutiva e as ‘experiências’ sobre melhoramento do raciocínio.

Contra a possível apologia da transmissão do testemunho da filosofia para a ciência cognitiva, penso que a presença da ciência cognitiva não justifica o afastamento da filosofia da tarefa epistemológica mas sim uma reformulação desta. Não há como negar que os estudos acerca de cognição são relevantes para o tratamento das questões epistemológicas tradicionais enunciadas no início do Capítulo e que a sua existência constitui mesmo um desafio para o estudo filosófico do conhecimento (senão o que diríamos estar a fazer quando abordamos as questões clássicas da epistemologia? exegese e comentário de autores? uma posição semelhante representa tanto uma retirada da filosofia do campo epistemológico como as interpretações cientistas radicais da ideia de ‘epistemologia naturalizada’). Parece-me que a forma correcta de ver o problema é a seguinte: as investigações cognitivas não satisfazem o propósito central da epistemologia, que é compreender o direito que os agentes têm (ou não têm) às suas crenças, e portanto compreender a normatividade envolvida na justificação. Penso que devemos aceitar a divisão de trabalho e a complementaridade entre ciência cognitiva e filosofia proposta por Goldman. No entanto, a proposta de Goldman quanto à natureza da justificação não me parece igualmente pertinente. Penso que Stich está certo quando ao considerar a análise de Goldman como um exemplo de epistemologia analítica, oferecendo um resultado ‘paroquial e conservador’. Mas o facto de

não restringirmos a 'justificação' ao conceito inscrito no pensamento e linguagem comuns e explicado por procedimentos de análise não nos exime da necessidade de compreender a normatividade envolvida na justificação. O pragmatismo de Stich é precisamente uma proposta de enquadramento para a compreensão de tal normatividade. Como seria de esperar do pragmatismo, Stich remete as questões da racionalidade teórica (em que devemos acreditar?) para as questões para a racionalidade prática (o que valorizamos? o que nos motiva a agir?). A divisão habitual dos estudos da racionalidade em racionalidade teórica e racionalidade prática mostra assim ser mais operacional do que substancial.

Ressalta desta incursão pelos debates meta-epistemológicos em busca de uma teoria da natureza da justificação uma conclusão geral: as questões da racionalidade prática não podem ser dadas como resolvidas no tratamento das questões da racionalidade teórica. As questões da racionalidade prática, relativas nomeadamente ao que move o agente à acção, estão claramente em jogo na forma de conceber a própria tarefa da epistemologia, se essa epistemologia não é concebida de forma apriorista mas tendo em atenção agentes reais, o estatuto de uma teoria da avaliação epistémica e a forma como respectivamente a filosofia e a ciência cognitiva encaram os agentes cognitivos. Um relato pragmatista dá-nos uma visão específica da forma como as questões epistemológicas se relacionam com o que move o agente à acção. E aqui não podemos, pura e simplesmente assumir à partida, apenas porque isso é vantajoso num contexto de teoria epistemológica, considerar que aquilo que move o agente à acção é o facto de valorizar intrinsecamente crenças verdadeiras e raciocínios racionais. Antes de nos pronunciarmos nesse sentido e pensando na nossa natureza de agentes, precisamos de saber o que é valorizar alguma coisa e em seguida como é que esse valorizar se relaciona com as nossas atitudes epistémicas e os nossos inqueritos.

Capítulo 4

Duas teorias filosóficas da racionalidade

Neste Capítulo pretendo apresentar dois filósofos cujo trabalho é suficientemente ambicioso na sua finalidade para pretender ao título de 'teoria filosófica da racionalidade'. Assim, após a descrição geral de questões de racionalidade prática e teórica, que foi objecto dos Capítulos 3 e 4, passo a considerar duas teorias gerais da natureza da racionalidade relevantes no panorama contemporâneo da filosofia analítica, a teoria apresentada por Simon Blackburn em *Ruling Passions* (1998) e a teoria apresentada por Robert Nozick em *The Nature of Rationality* (1993).

Simon Blackburn – a mente e a natureza da racionalidade prática

«Que todos os apetites particulares e paixões se dirigem a coisas externas, distintas do prazer que delas surge, é desde logo manifesto, pois não poderia existir este prazer se não fosse pela prévia adequação entre o objecto e a paixão: não poderia haver gozo nem deleite de uma coisa mais do que de uma outra, de comer comida mais do que de engolir uma pedra, se não fosse uma afeição ou apetite por uma coisa mais do que pela outra»

Joseph Butler, *Fifteen Sermons Preached at the Rolls Chapel*, Sermão 11

«Toda a afeição particular, mesmo o amor pelo nosso próximo, é tão realmente nossa afeição como o amor-próprio; e o prazer que resulta da sua gratificação é tanto o meu próprio prazer como o prazer que o amor-próprio teria, por saber que eu mesmo seria feliz daí a algum tempo, seria o meu prazer»

Joseph Butler, *Fifteen Sermons Preached at the Rolls Chapel*, Sermão 11

Motivação para a acção:

sentimentalismo versus racionalismo.

Natureza das pretensões normativas éticas: expressivismo

O tratamento da racionalidade feito pelo filósofo neo-humano britânico Simon Blackburn em *Ruling Passions* interessa-me aqui como exemplo contemporâneo de elaboração de uma posição sentimentalista quanto ao problema da motivação para a acção (recordo que o **sentimentalismo** é uma posição acerca da natureza e fundamentos da ética, frequentemente associada ao nome de D. Hume, de acordo com a qual a ética se funda em sentimentos). A posição sentimentalista de Blackburn – que pretende além do mais constituir um ponto de vista naturalista na abordagem das questões da razão prática, colocando a explicação do comportamento normativo humano na continuidade da explicação do comportamento de outros animais¹ – é acompanhada por uma posição expressivista acerca da natureza das pretensões normativas éticas. O **expressivismo** é a posição de acordo com a qual a função primária destas pretensões normativas é – por contraste com as expressões de crença – expressar atitudes e emoções. Quando emitimos uma pretensão normativa o que estamos a fazer não é tentar ‘captar a realidade correc-

¹ Um projecto naturalista *realista* em ética não é, segundo Blackburn, exequível. Ser naturalista significa para Blackburn ver os seres humanos como parte da ordem natural e explicar o seu comportamento sem fazer apelo a alguma ‘ordem platónica’ de Formas ou Normas.

tamente’, i. e., não é formular juízos susceptíveis de verdade e falsidade, mas sim exprimir² desejos, sentimentos e atitudes, que são precisamente o tipo de estados-dirigidos-a-fins que motivam as pessoas à acção segundo uma concepção humeana de motivação para a acção.

O afastamento dos preconceitos racionalistas que Blackburn pretende conseguir com a sua teoria da racionalidade prática, a ideia segundo a qual são os sentimentos, e não a razão, que movem as pessoas à acção, tem repercussões que se alargam a todas as teorias filosóficas ou científicas que lidam com a **motivação** humana³. De facto, facilmente verificamos que os problemas acerca da motivação para a acção não são apenas questões da filosofia (filosofia ético-política e filosofia da mente e da acção). Procurar compreender o que está em causa quando um agente, uma pessoa, se sente ou não se sente movido a agir, diz respeito também, por exemplo, a questões médico-psicológicas tais como a apatia, a abulia, a depressão, o an-hedonismo e muito em especial às distinções que devem ser feitas entre elas⁴, a questões

² É importante notar que *exprimir* não é *descrever*.

³ Recordo a noção de **motivação** com que se lida em contextos de psicologia e neurociência: a motivação é ligada à avaliação e selecção de finalidades (*goals*) do comportamento de agentes e aos mecanismos cognitivos/neuroquímicos/hormonais subjacentes a esses processos e é tomada como uma influência moduladora e coordenadora sobre a direcção, vigor e composição do comportamento, com fontes internas, ambientais e sociais. Mudanças de estado motivacional manifestam-se em vários níveis da organização comportamental e neuronal. É pelo facto de as finalidades do comportamento a seleccionar, os objectos de avaliação, poderem ir desde objectos no ambiente do agente até abstracções que é possível em Filosofia, por exemplo, discutir a forma como a deliberação racional motiva, bem como discutir se todas as razões têm que se motivantes, ou pelo contrário algumas razões podem ser boas razões (‘razões normativas’) e no entanto não motivar o agente à acção (cf. WILLIAMS 1980 e DANCY 2000).

⁴ Por exemplo, num indivíduo que não se comporta de forma motivada e finalista relativamente aqueles que seriam potencialmente ‘objectos de desejo’ em torno de si muitas coisas diferentes podem passar-se: ele pode

jurídico-psicológicas tais como o respeito dos indivíduos pelas leis, etc. De acordo com a proposta de Blackburn, precisamos de sentimentos e não apenas de razão (i. e., de representação de situações e de cálculo meios / fins) para nos sentirmos motivados a agir. Do lado ‘adversário’ da proposta de Blackburn estão as **concepções ‘racionalistas’ de motivação para a acção** largamente difundidas na filosofia contemporânea. O facto de estas constituírem um ponto de partida frequente quer na filosofia moral e política quer na filosofia da mente torna especialmente importante verificar se elas são sequer defensáveis.

A compreensão das posições gerais de Blackburn acerca de racionalidade prática, vistas em contraste com as posições de outros autores relevantes no panorama contemporâneo, tais como por exemplo, Thomas Nagel⁵, John McDowell⁶ ou Christine

não ser capaz de se fazer mover, não ser capaz de (se fazer) querer, não ser capaz de experimentar prazer, etc.

⁵ Posições específicas defendidas por T. Nagel permitem-nos ver facilmente em que consiste uma oposição ao humanismo: relativamente à estrutura do nosso querer, Nagel admite, por exemplo, a existência de desejos motivados ou justificados por razões *que não dependem de desejos mais básicos*. O alcance conferido por Nagel à razão na estrutura da motivação é assim claramente mais lato do que aquilo que é permitido a um humeano pensar.

⁶ Para compreender as posições de J. McDowell (McDOWELL 1981) é necessário recordar o que se entende por **cognitivismo** em ética é a ideia segundo a qual nas proposições morais está em causa *conteúdo cognitivo* (e não apenas expressão ou exclamação), e os juízos morais são acerca de *factos morais*. As posições cognitivistas (mas anti-platonistas) de J. McDowell em ética partem basicamente das considerações de Wittgenstein acerca de seguir-regras para defender que não existe distinção entre factos e valores e que, portanto, não existe nada de distintivo nas proposições morais face a outro tipo de proposições. Segundo McDowell, subjacente à ideia segundo a qual existiria algo de distintivo nas proposições morais estaria uma indefensável distinção entre Razão e Paixões. Para McDowell, requisitos morais (como também, de resto, requisitos lógicos) são perceptíveis a partir do interior de ‘formas de vida’.

Korsgaard⁷, é uma boa chave para compreender a forma como a discussão, centrada em indivíduos, acerca de racionalidade e motivação para a acção se reflecte num quadro ético e político mais amplo. Seria possível abordar o pensamento de Blackburn a partir de um ângulo mais técnico, comum nas discussões em filosofia moral⁸, mas vou tomar antes o seguinte caminho. Vou assumir que ao investigar a natureza e fundamentos da ética e a natureza das pretensões normativas éticas, se desenvolve forçosamente posições quanto à motivação dos agentes para a acção, quanto àquilo que se passa quando, por exemplo, nos damos, ou nos dão, razões para agir de uma ou outra forma, razões que nos motivam ou não motivam, e vou considerar a proposta de Blackburn como especialmente relevante para aquela a que chamei na Introdução ‘questão alargada da racionalidade’. Esta questão alargada da racionalidade (Questão 7) diz respeito, recorde, ao facto de questões cognitivas acerca da racionalidade dos indivíduos deverem ser consideradas como a ‘ancoragem’ individual no tratamento das macro-questões da ‘Razão na sociedade, na história e na civilização’. Ora, uma dos objetivos de Blackburn em *Ruling Passions*⁹ é precisamente

⁷ C. Korsgaard (KORSGAARD 1996) defende uma posição kantiana (e conduz uma oposição kantiana ao humanismo) de acordo com a qual embora a razão prática não seja uma capacidade de reflexão sobre um suposto domínio objectivo de factos normativos, ela é governada por constrangimentos normativos genuínos, que são tornados normativos pela sua relação com a vontade de agentes cujas decisões governam. Os princípios da razão prática são princípios constitutivos da agência racional, que nos obrigam tanto quanto nos comprometemos a obedecer-lhes, ao quisermos o que quer que seja. Chama-se **construtivismo** a esta posição pelo facto de o reino do normativo ser construído pelos agentes na sua actividade volitiva.

⁸ Seria de notar então que o sentimentalismo de Blackburn (ao qual o próprio Blackburn chama também quase-realismo) se opõe quer ao realismo moral, de acordo com o qual haveria um reino objectivo de factos normativos e factos acerca do que devemos fazer anteriores à deliberação do agente, quer ao construtivismo.

⁹ Cf. especialmente os Capítulos 5, 6, 7 e 8.

explorar a forma como posições em filosofia moral e política reflectem a forma de conceber a motivação para a acção ao nível dos indivíduos.

Vejamus então a forma como Blackburn aborda a questão da racionalidade prática. De acordo com Blackburn, a forma como falamos de ética reflecte o que se passa de ético na nossa forma de viver e pensar e o que se passa de 'ético' é o seguinte. Antes de mais, nós 'vibramos': indignamo-nos, horrorizamo-nos, apreciamos, respeitamos. Depois vimos a exprimir o que valorizamos. O que está em causa na ética é assim, para Blackburn, saber como havemos de nos conduzir em assuntos práticos, quem admirar e respeitar, quem punir, quando agir, e não questões relativas a saber que tal-e-tal é o caso. Não há aí uma Razão a sustentar o que fazemos nem objectos do pensamento ético (que seriam 'factos normativos', razões para a acção reificadas). O que se passa de ético connosco passa-se na continuidade do nosso comportamento natural. Quando valorizamos alguma coisa e nos expressamos em termos do que é bom, mau ou obrigatório, não estamos a procurar conceitos ou condições de verdade: propriedades e proposições éticas devem antes ser vistas como uma espécie de moeda na qual as nossas transacções valorativas decorrem. Logo, para compreendermos a natureza dos juízos éticos, temos que olhar para as acções e situações que constituem as transacções, e não para as próprias proposições éticas (olhar para as próprias proposições éticas seria como analisar minuciosamente uma nota de banco para compreender a natureza do valor monetário). Este é o núcleo do expressivismo, que é portanto uma posição acerca do que fazemos quando dizemos o que valorizamos, quando emitimos pretensões normativas. Como afirmei atrás, de acordo com o expressivismo as pretensões normativas não são juízos susceptíveis de serem verdadeiros ou falsos mas sim expressões de desejos, sentimentos e atitudes, precisamente o tipo de estados que movem as pessoas à acção.

Reconhecendo uma objecção no horizonte – de acordo com a qual o sentimentalismo é subjectivista, não admite factos mo-

rais objectivos, sendo portanto tendencialmente relativista¹⁰ –, Blackburn elabora a sua posição de forma a defender que ela não tem de ser relativista pelo facto de se querer manter naturalista, não admitindo um reino de factos normativos e sentimentalista, pretendendo fundar a ética não na razão mas em sentimentos de indivíduos. O espaço de manobra de Blackburn é o seguinte: mesmo se não se admite um reino de factos normativos, mesmo se se admite que a natureza é 'sem coração', não quer saber do que acontece ou deixa de acontecer às 'suas' criaturas, o facto é que existem criaturas naturais com desejos (nomeadamente os animais humanos). Ora, tais desejos são susceptíveis de vicissitudes e moldagens, e a história natural dos desejos coloca-nos inevitavelmente perante o problema da educação sentimental dos humanos. Esta perspectiva fornecerá, em termos gerais, a ancoragem psicológica da normatividade que sustenta a teoria da racionalidade prática defendida por Blackburn em *Ruling Passions*.

Os desejos, o eu e os outros: sentimentos morais e crítica ao egoísmo psicológico

De forma a compreendermos o alcance e as origens da posição de Blackburn, convém ter em mente aspectos da tradição sentimentalista em filosofia moral, tradição na qual se inscreve o próprio David Hume bem como um filósofo moral seu contemporâneo, Adam Smith. Os aspectos a considerar são, nomeadamente, posições acerca da estrutura da motivação humana bem como a crítica à (ainda hoje) comum caracterização de todos os agentes como egoístas psicológicos¹¹. Essa crítica é central para

¹⁰ O relativismo, posição meta-ética, corresponderia, na formulação de Blackburn e utilizando a ideia de Nagel de 'última palavra', à ideia segundo a qual a última palavra em ética seria 'é assim que fazemos', 'é apenas a nossa opinião' (BLACKBURN 1998: 279).

¹¹ Distingue-se usualmente o egoísmo psicológico (ideia segundo a qual as pessoas agem sempre em função do seu interesse pessoal) do

a posição acerca da natureza da motivação que está na base da teoria da racionalidade prática de Blackburn e a clássica refutação do egoísmo ético pelo bispo e filósofo do século XVIII Joseph Butler (1692-1752) serve-lhe de referência em *Ruling Passions*.

Os autores sentimentalistas clássicos, D. Hume e A. Smith, defenderam que os **sentimentos morais** são a força motivadora das nossas acções e também o ponto de apoio para a coordenação das acções individuais em sociedade. Sentimentos morais são traços tais como o orgulho, a vergonha, a culpa. Resta, é claro, saber qual é a natureza e origem desses sentimentos morais. Ora, quer Hume quer Smith defenderam uma imagem razoavelmente optimista da natureza humana, cujo fulcro é a noção de **simpatia**, pensada como coordenação com os prazeres e dores de outros (na formulação de D. Hume 'As mente dos homens são espelhos umas das outras'). É esta simpatia que assegura, por exemplo, alguma humanidade no trato dos humanos entre si: como Hume notou, a natureza nunca nos mostrou cenários em que a guerra e o conflito são as únicas relações entre os indivíduos. A partir do momento em temos grupos de humanos termos, do ponto vista da acção e interacção, alguma coisa mais do que a guerra hobbesiana de todos contra todos, ou o egoísmo racional associado à maximização da utilidade esperada. Temos, nomeadamente, **benevolências locais**. É a partir destas que em termos sociais surgirão eventualmente traços tais como a confiança (*trust*) e a coordenação, desenvolvidos ao longo do tempo. De acordo com uma linha de pensamento humeana – e este é um ponto relativamente ao qual é particularmente claro que a forma de nos posicionarmos perante as questões filosófico-psicológicas da motivação para a acção tem reflexos directos na forma de conceber o problema político, pensado como o problema relativo àquilo que mantém agentes humanos em sociedades – o problema de Hobbes (o problema de uma instituição estável da sociedade e da legalidade) não pode ser resolvido racionalmente

egoísmo ético (ideia segundo a qual as pessoas devem agir sempre apenas em função do seu interesse pessoal).

e de uma vez por todas, através de algum tipo de contrato, mas apenas organicamente, com o desenvolvimento de hábitos de **confiança e cooperação** entre indivíduos.

A partir deste núcleo de ideias, Blackburn defenderá o seguinte acerca do mecanismo de motivação que sustenta as acções individuais e a coordenação social destas. Aquilo que sentimentos morais tais como a vergonha, a culpa e o orgulho fazem – e é por isso que eles têm a função que têm na motivação para a acção – é ligar a avaliação do comportamento próprio à **internalização do olhar dos outros** e à sua apreciação (à sua fúria, desdém, aprovação). Esse funcionamento molda a nossa natureza mental dinâmica, i. e., o nosso querer. Que o mecanismo é natural é, de acordo com Blackburn, óbvio: os humanos têm a preocupação obsessiva de observar as acções uns dos outros, de discutir e coordenar as suas próprias acções com as acções dos outros (pense-se na necessidade de narrativas, seja em livros, cinema, telenovelas, notícias de tablóides ou vicissitudes da vida política e económica). A preocupação com o olhar dos outros é um traço humano tão geral que nos leva a pensar que a humeana 'vibração interna em simpatia com os outros' é natural e fundamental para compreender a motivação para a acção.

Mas exactamente como é que tudo isto se processa em cada indivíduo – este facto de 'o outro estar em mim no meu próprio querer' – em termos de 'mecanismos psicológicos' que virão a estar subjacentes à produção de juízos morais? A ideia de Blackburn é que estaria em funcionamento, em todos os humanos, um processo de **educação sentimental**, i. e., de moldagem dos desejos, com as seguintes características. O processo teria um início exteriorista, cultural – aprovamos uma ou outra qualidade nas outras pessoas, possivelmente porque fomos educados para isso. Dada a possibilidade de assumir 'o ponto de vista comum' humeano (descrito no Capítulo 2), essa aprovação torna-se estima. Ao 'olharmos para o interior' apercebemo-nos se essa é uma qualidade que nós próprios exibimos ou não. Sentiremos então auto-satisfação, orgulho, vergonha, etc., correspondendo à nossa avaliação inicial e imaginando a apreciação que os outros fariam de nós.

Quando Blackburn considera que a 'mola motivacional' dos comportamentos humanos é 'sentimental', o que tem em mente são estes mecanismos de elogio e culpa, vergonha e orgulho e o facto de eles estarem envolvidos na internalização de normas por indivíduos e no respeito por elas. Trata-se portanto de qualquer coisa como uma ancoragem psicológica da normatividade, descrita de um ponto de vista naturalista, que tem que existir para as normas serem motivantes para os indivíduos. Blackburn defende no entanto que nada disto nos obriga a pensar que o resulta desta situação são agentes que têm como único objectivo na suas acções obter elogios, evitar atribuições de culpa, etc. Porquê? Precisamente porque, apesar da 'mola sentimental' que 'funciona' individualmente, Blackburn pensa que um indivíduo pode chegar a um ponto de vista comum (de novo evocando Hume, Blackburn faz notar que é muito diferente o caso em que alguém diz 'este homem é meu inimigo, odeio-o' e o caso em que alguém diz ele é ignóbil, pretendendo que os outros concordem consigo).

É muito importante compreender exactamente em que medida é que esta caracterização sentimentalista da motivação para a acção pode ser contraposta à bastante generalizadamente aceite ideia do egoísmo psicológico. Afinal, parece que descrevemos aqui, também, uma centração no eu, apenas menos racional do a que nos é usualmente apresentada. O ponto é o seguinte, e é importante na forma como Blackburn considera o eu e os desejos: a grande razão pela qual não devemos caracterizar de forma simplista todos os agentes como egoístas psicológicos é o facto de não podermos afirmar sempre, em todas as circunstâncias, que o objecto do desejo é o prazer próprio.

A ideia segundo a qual o principal motor da nossa acção é o interesse próprio parece irrecusável e está, segundo Blackburn, mais ou menos abertamente incorporada em variadas teorias, que vão desde o utilitarismo¹² à sociobiologia. É certo que há um

¹² Mesmo que a pedra de toque do utilitarismo seja a 'imparcialidade' na consideração do bem-estar, ainda assim quem age não é algum agente

problema com a formulação do egoísmo psicológico: facilmente se derrapa para uma formulação vazia, segundo a qual cada agente é motivado pela sua própria motivação (e como poderia ser de outro modo?). Mas o que Blackburn pretende fazer ver com a sua crítica ao egoísmo psicológico é o seguinte: mesmo que o querer de um agente seja regido pelo interesse próprio, mesmo que em abstracto as suas acções sejam praticadas de acordo com esse princípio, o facto é que quando um agente age aquilo que ele visa são coisas externas específicas (comida, sexo, salvar as baleias, a vitória do comunismo, escrever um livro, etc.). De acordo com o egoísmo psicológico, todas essas coisas seriam meros meios para o prazer próprio do agente. Butler no entanto argumenta, e Blackburn concorda, que mesmo que o interesse do agente seja próprio no sentido em que é o interesse desse agente, o interesse não tem que visar, que ter por objectivo ou finalidade, estados de prazer próprio do agente. O objecto de um particular desejo (de comida, sexo, vitória do meu partido nas eleições, a morte do meu inimigo) não é a mesma coisa que o prazer sentido pelo agente com a satisfação desse desejo. Os desejos 'apontam para fora'. Claro que o prazer que existe quando o desejo é satisfeito é prazer do agente: é dos interesses e desejos do agente que se fala, e interesses e desejos são sempre referidos a agentes. Mas isso não nos permite pensar que o objectivo último de qualquer acção é o prazer próprio. O facto de os interesses serem interesses do eu não os torna interesses no eu. Antes de mais, nem sequer é óbvio que a satisfação de qualquer desejo, embora representada como sendo de certa forma prazerosa, traga prazer para mim: eu posso saber, por exemplo, que não denunciar alguém redundará na minha própria condenação à morte, e ainda assim querer fazê-lo, e cumprir o meu desejo de o fazer. Será que também nesse caso a satisfação do meu desejo redundará em prazer para mim?

universal e imparcial mas sim agentes particulares, que têm que ser, de algum modo, movidos à acção.

Em alternativa, se nos sentirmos insatisfeitos com a descrição da satisfação dos desejos como culminando sempre no prazer próprio, poderemos talvez defender que a satisfação dos desejos produz o alívio de uma tensão, como por vezes se diz em psicologia. Para Blackburn, também essa definição é insatisfatória: o que é desejado no desejar não é o alívio da tensão. Nós só ficaríamos satisfeitos com a remoção do desejo, com o alívio da tensão, por hipótese através de alguma droga pacificadora, se não nos tivéssemos identificado com esse desejo, se ele fosse, não de alguma forma aquilo que nós somos e queremos ser, mas sim uma 'adição ou perturbação' exterior, um 'desejo implantado' e não subscrito por nós. Mas pense-se no exemplo acima, e no caso em que o nosso desejo seria removido através de alguma droga. Seria isso a satisfação do desejo? Ou pense-se num outro exemplo de Blackburn¹³, relativo a pessoas que sentem desejo sexual por alguém específico, que as atrai, por quem estão por hipótese apaixonadas: a maior parte das pessoas que experimentam esse desejo não ficariam satisfeitas com a possibilidade de serem libertadas dele através de alguma droga pacificadora. O que elas querem é, naturalmente, sexo com a pessoa desejada.

O problema aqui é saber se o egoísmo psicológico é uma boa teoria, uma descrição realista, da psicologia moral humana. Será que podemos dizer que sempre que há satisfação de um desejo, mudança subjectiva, de tensão para não tensão, ou de falta para não falta, essa mudança subjectiva é aquilo que é desejado quando o agente deseja? Butler, nomeadamente, defendeu que a descrição é má: a alteração do estado próprio, a mudança do estado de experimentar o desejo para o estado de não experimentar o desejo não é o objecto de desejo. Toda a situação pressupõe um desejo prévio a que Butler chama apetite ou paixão e que é apetite ou desejo de alguma coisa específica.

¹³ BLACKBURN 1998:140.

Desejos como 'querer ver-se livre dos seus próprios desejos' ou 'ver os seus próprios desejos satisfeitos' são desejos de segunda ordem, que pressupõem outros desejos de primeira ordem. Não se pode passar do facto de o objecto do desejo ser de alguma forma representado como desejado, e nesse sentido prazeroso, para a ideia segundo a qual aquilo que todo o agente deseja é o prazer próprio. Pegando ainda noutro exemplo de Blackburn¹⁴: se aquilo que eu desejo é a sobrevivência das baleias, que as baleias sobrevivam é representado por mim como prazeroso. Mas a sobrevivência das baleias num futuro distante não é idêntica ao meu prazer, ou a prazer para mim.

Blackburn subscreve as objecções de Butler ao egoísmo psicológico e considera que o egoísmo psicológico é uma má teoria da motivação para a acção. Assim, é um passo importante numa teoria da racionalidade prática começar por admitir que os desejos do agente não são por definição desejos egoístas, visando o prazer pessoal. O interesse do próprio é obviamente referido ao próprio mas aquilo que é visado (ao contrário do que é tacitamente suposto em teorias da acção que assumem o egoísmo psicológico, como a sociobiologia ou a teoria da decisão racional) não é necessariamente o prazer próprio. Nomeadamente, desejos de alguma forma benevolentes relativamente a outros (desejos *other-regarding*) são certamente referidos ao próprio (são *self-referential*), infundidos da motivação deste para agir, sem que isso implique que eles são em última análise, ou podem ser desmascarados como, desejos de prazer do próprio, satisfações do interesse próprio. Mas a verdade é que, segundo Blackburn, a concepção errónea analisada por Butler na sua crítica do egoísmo psicológico continua a ser muitíssimo comum nas teorias da acção e da motivação. É como se desmascarar o egoísmo subjacente a toda a acção humana fosse a única forma realista, razoável, não iludida, de conceber a nossa motivação

¹⁴ BLACKBURN 1998: 137.

para agir. Qualquer pessoa que veja claramente verá no fundo da acção de qualquer outra pessoa só e apenas egoísmo. Blackburn considera esta ideia simplista, e em *Ruling Passions* argumenta nomeadamente que uma teoria da motivação para a acção com a forma de egoísmo-psicológico-especificado-nos-termos-da-decisão-racional (o agente racional age porque é um egoísta psicológico que age de modo a maximizar a utilidade esperada) não é nem empírica, nem normativa, nem definicional. De facto, para Blackburn, a ideia de maximização da utilidade esperada é de alguma forma ‘vazia’, o que não rouba importância à teoria da decisão mas a torna insusceptível de ser recrutada sem problemas para a filosofia moral.

Voltemos à questão dos agentes que teriam como único objectivo obter elogios, evitar atribuições de culpa, etc. Agentes semelhantes seriam, segundo Blackburn, ‘crianças mal educadas’ ou em que a educação não chegou a funcionar. Agentes semelhantes não poderiam ser, nem são, a última palavra de uma posição sentimentalista quanto à natureza da ética – a forma correcta de conceber o mecanismo apontado é vê-lo como uma explicação genética da estrutura da motivação. Ora, é fundamental segundo Blackburn separar a perspectiva ‘genética’ acerca de motivação da perspectiva sincrónica sobre a deliberação de um agente. Noutras palavras, mesmo se mecanismos de elogio e culpa explicam geneticamente a nossa natureza motivacional isso não significa que eles figurem sincronicamente nas estruturas motivacionais dos agentes, excluindo tudo o resto. Eles estarão presentes, para usar os termos de Blackburn, de forma *self-effacing* (auto-apagando-se). O resultado sincrónico pode ser apenas, e é isso que Blackburn defende, que nos preocupamos (*we are concerned*), queremos saber das nossas acções, da forma como elas fazem que nós sejamos vistos. As nossas acções são nossas, dizem-nos respeito, não se destacam de nós ‘racionais e puras’ (como de certa forma uma imagem kantiana da deliberação nos faz pensar).

Uma outra especificação de Blackburn quanto aos nossos desejos é a seguinte. É importante, para capturar correctamente (i. e., de forma defensável quando consideramos a questão do ponto de vista da teoria da mente) a natureza da motivação para a acção, considerar que, além de os objectos de desejo serem específicos e diferentes do prazer próprio, os nossos desejos não são objectos passivos, a serem observados e controlados e certificados ou afastados por um eu, um ‘centro decisor de nós próprios’ que seria a Razão. Contrariamente ao que Kant defendeu (ou a imagem caricaturada de Kant na literatura), nós não somos seres tais que alguma vez estejamos acima dos nossos próprios desejos e preocupações, de modo a operar sobre eles de uma forma totalmente activa. Acreditar nisso é pressupor à partida aquilo a que Blackburn chama, em *Ruling Passions*, a divisão kantiana entre a tripulação (os desejos) e o Capitão (a Razão). Ora, Blackburn rejeita essa divisão e propõe as seguintes alterações de perspectiva na nossa forma de pensar acerca dos nossos **desejos**:

(i) quando desejamos nós somos os nossos desejos e o conflito entre estes e não um eu racional que observa e avalia os desejos (embora certamente possamos ter desejos de segunda ordem acerca dos nossos próprios desejos – i. e., nós somos agentes tais que podemos desejar desejar e desejar não desejar e temos por aí um certo poder de intervenção no nosso próprio querer)

(ii) quando desejamos o que se passa não é que estamos ‘virados para dentro’ supervisionando e tentando dominar o nosso tumulto interior; o que acontece é que somos atraídos pelos aspectos de situações que se apresentam como atraentes (i. e., da mesma forma que as percepções, os desejos são estados mentais intencionais, que ‘olham para além de si’).

(iii) quando deliberamos o que fazemos não é supervisionar inclinações dentro de nós, inclinações que seriam passivamente observadas pelo nosso eu racional – há aí na situação de deliberar um visionamento do nosso entorno, das coisas cujas características

atraem ou repelem as nossas inclinações. A primeira forma de conceber a deliberação está estreitamente ligada a uma reificação do deliberador, que está por sua vez estritamente ligada ao kantismo e ao dualismo que, segundo Blackburn, este comporta.

Dito isto quanto a desejos e deliberação, as questões que se colocam a Blackburn são as seguintes: 1) como é que, a partir de uma estrutura motivacional como a descrita, se chega a apreciar traços que merecem ser considerados correctos, moralmente defensáveis e desejáveis, i. e., como é que um agente vem a ser movido não (apenas) pelo desejo de elogio mas pelo desejo de merecer elogio pelos traços que considera moralmente apreciáveis?; 2) que tipo de autoridade poderá ter a instituição de um dever-ser através de um mecanismo 'frágil e psicológico' como este? Para Blackburn a única resposta para a primeira questão, i. e., para o problema da adequação entre a estrutura motivacional sentimental e aquela que se considera ser uma 'boa psicologia moral', que possa nomeadamente construir a real ancoragem de uma sociedade justa, é, como já afirmei, aquilo a que chama uma educação sentimental. Isso significa antes de mais que o problema da conformação da estrutura motivacional dos indivíduos não será nunca resolúvel através de um apelo à Razão. Uma tal resposta atribui obviamente um papel importante à tarefa do educador: não há aqui um caminho único ou uma garantia a priori do resultado, i. e., do desenvolvimento de determinados hábitos e propensões do querer considerados desejáveis num particular indivíduo. Tudo pode sempre dar 'errado'. Não são no entanto impossíveis clareiras de cooperação e confiança entre indivíduos, dada uma determinada educação sentimental. Aquilo que não devemos pressupor é um Tribunal da Razão ao qual todos os agentes comparecem finalmente para justificar as suas acções.

Importância e problemas da proposta de Blackburn

O grande problema da proposta de Blackburn é, como comecei por apontar, o relativismo. Embora o próprio Blackburn considere possível evitar o relativismo, mesmo considerando que emoções, desejos e sentimentos e sua contingente moldagem são o material da ética, fica uma dúvida. A alternativa seria no entanto não olhar de frente, do ponto de vista da teoria da mente, as múltiplas facetas da nossa natureza emocional e motivacional quando partimos para uma teoria ética. Ora, essa não seria, segundo Blackburn, uma alternativa razoável. Mas nestas circunstâncias, e aceitando uma teoria humeana da nossa estrutura motivacional, como podemos exactamente imaginar que são moldáveis e educáveis as atracções que experimentamos, sabendo que nunca podemos apelar à Razão universal? Não há, no sentimentalismo de Blackburn, uma última palavra para além, precisamente (e contingentemente) dessa moldagem. Blackburn insiste, no entanto, no facto de a sua teoria da racionalidade prática não constituir obstáculo para que as 'virtudes', i. e., as conformações desejáveis das propensões do querer, sejam as mesmas de um ponto de vista sentimentalista e de um ponto de vista kantiano (por exemplo, a justiça). A única coisa que difere é o estatuto que lhes é atribuído. Do lado da tradição racionalista kantiana, a razão opera milagres, no sentido em que poderíamos dela esperar que constanja de forma irrecusável o querer de qualquer agente humano, do outro lado, não há 'milagres racionais', apenas psicologia e moldagem de mecanismos de motivação. Mas se, para Blackburn, a preocupação por um terreno comum, por um ponto de vista comum, a que se tem chamado imparcialidade e universalizabilidade é apenas, como qualquer outra preocupação, 'o resultado contingente e frágil das nossas naturezas sentimentais', e não um ditame da Razão, isso não significa que ele seja, por exemplo, insensível à atracção do ideal moral e político kantiano, nomeadamente na versão rawlsiana considerada em *Ruling Passions*. O que Blackburn quer tornar

claro é que o que pensadores da racionalidade moral e social como Kant e Rawls fazem é promover um ideal, não é descrever a nossa psicologia. Não se pode, portanto, legitimamente conferir outro estatuto às suas propostas morais e políticas senão o de exercerem uma contingente atracção em psicologias moldadas de determinada maneira. De resto, como nota Blackburn, nem todos os prolongamentos do ideal kantiano são atraentes: ver todos os humanos como potencialmente racionais da mesma maneira pode fazer-nos confiar demasiado neles. Por exemplo, e para usar os exemplos de Blackburn, confiar demasiado na segurança de situações em que estes lidam com armas, drogas e automóveis ou agem no sacrossanto domínio privado. Um enfoque humeano tornar-nos-ia pelo contrário mais cautelosos e mais atentos às situações concretas e à moldagem dos indivíduos.

Uma 'moral' possível a retirar das propostas de Blackburn é a seguinte: quando lidamos com questões de moralidade, questões acerca do que devemos fazer, o apelo a fontes de autoridade, tais como textos ou teorias ou leis, e à relação destes com a nossa 'racionalidade' não é de forma alguma suficiente: se essas fontes exteriores não encontram apoio na nossa constituição motivacional elas pura e simplesmente não nos moverão.

Resta recapitular quais das tarefas para uma teoria filosófica da racionalidade são tratadas por Blackburn em *Ruling Passions*. Blackburn trata sobretudo a questão 6, a questão da motivação dos agentes para a acção, e a questão 7, a questão alargada da racionalidade. Mas Blackburn trata também as questões 4 e 5, que visam a natureza dos conceitos mentais e da atribuição de racionalidade, e isso é importante nomeadamente quando se especifica a racionalidade atribuída aos agentes pela teoria da decisão racional, sobretudo se esta de facto assume o egoísmo psicológico. De qualquer forma, o que torna a proposta de Blackburn especialmente relevante para os propósitos de mapeamento do presente livro é o facto de ela considerar directamente a dimensão de desejo da racionalidade, a forma como nos desejos, o eu e os outros se relacionam para resultar, num agente, em

motivação para a acção. Investigações cognitivas, nomeadamente toda a investigação em curso sobre emoções e também sobre teoria da mente e *mindreading*, ganhariam em serem vistas à luz da teoria da racionalidade prática de Blackburn.

Robert Nozick

– *uma teoria da racionalidade simbólica*¹⁵

Embora o filósofo americano Robert Nozick (1938-2002) seja conhecido sobretudo pela sua obra de filosofia política *Anarchy, State and Utopia* (1974), ele é também o autor de obras tão pessoais e enciclopédicas – traços certamente heterodoxos na filosofia analítica – como *The Examined Life* (1981) e *Philosophical Explanations* (1989). Nozick escreveu ainda, em 1993, um livro intitulado *The Nature of Rationality*, que constitui, de entre as obras até ao momento referidas, o exemplo mais claro daquilo a que tenho vindo a chamar uma 'teoria filosófica da racionalidade' e a procurar definir pela tentativa de resposta às questões enumeradas na Introdução. Os problemas tratados em *The Nature of Rationality* cobrem um largo espectro. Eles vão desde (i) uma abordagem evolucionista da natureza das 'razões' (Nozick concebe a 'história natural da racionalidade' como um caminho que vai desde os *free floating rationales* – razões para funcionamentos e comportamentos totalmente inapercebidos pelos agentes nos quais ocorrem, que constituem a 'racionalidade' ao nível, por exemplo, do comportamento de animais não humanos – até à preocupação com razões, o cuidado com o raciocínio e a decisão e com a explicitação dos princípios que os regem que *now floats free*¹⁶, i. e., existe livremente, sem qualquer propósito instrumental imediato, nos humanos), até (ii) questões

¹⁵ Baseio-me na recensão de *The Nature of Rationality*, publicada por Sara Bizarro no n.º 9 da revista *Intellectu* (www.intellectu.com), que me ajudou a compreender os pontos principais da obra.

¹⁶ NOZICK 1993: 114.

da racionalidade nas crenças e nas acções, que conduzem a teorias da justificação epistémica e da decisão racional, até (iii) uma proposta acerca do significado simbólico de acções, que permite a Nozick lidar com problemas clássicos da teoria da decisão, tais como o Dilema do Prisioneiro e o Problema de Newcomb, até (iv) propostas acerca das funções dos princípios na acção e no pensamento. Através do tratamento desta última questão, Nozick procura esclarecer as razões que nos motivam a querer determinadas qualidades na nossa vida intelectual e moral (tais como ter crenças verdadeiras, raciocinar e agir de forma racional), enfrentando assim o problema do valor da nossa racionalidade para nós próprios¹⁷. Não pretendo defender que em *The Nature of Rationality* Nozick responde directa e exaustivamente a todas as questões para uma teoria geral da racionalidade que listei no início deste livro, mas penso que ele oferece pontos de partida para tal tarefa, ao enquadrar numa mesma teoria da natureza da racionalidade questões que vão desde o surgimento, por evolução, de um interesse pela verdade em agentes, até à função dos princípios no pensamento humano.

Princípios

Nozick inicia *The Nature of Rationality* com uma admirável discussão acerca de princípios e da função destes no pensamento e na acção. Esta discussão enfrenta, partindo de um ângulo curiosamente instrumental, aquela a que na Introdução chamei a primeira questão para uma teoria filosófica da racionalidade, a questão dos critérios de correcção. Recordo que um critério de

¹⁷ A isto acrescenta-se o facto de Nozick ser, de entre os filósofos contemporâneos até aqui referidos, aquele que estabelece uma ligação entre valor da racionalidade para nós e a filosofia como disciplina. É nesse contexto que deve ser entendido o seguinte comentário: a palavra 'filosofia' significa amor da sabedoria mas aquilo que dá de facto prazer aos filósofos é raciocinar, combater e vencer pelo raciocínio (NOZICK 1993: xi)...

correcção explícita as razões por que consideramos determinados conjuntos de regras, normas, princípios, como cânones da racionalidade. Critérios de correcção são, portanto, justificações para standards de racionalidade empregues. Considerando esses standards como princípios, Nozick faz uma pergunta muito simples: qual é a função dos princípios? É dessa pergunta que nasce o capítulo inicial de *The Nature of Rationality*, intitulado How To Do Things With Principles: trata-se de analisar as funções dos princípios, funções que Nozick considera serem intelectuais, interpessoais, sociais e pessoais. Dizer que os princípios têm funções intelectuais, interpessoais, sociais e pessoais é o início de uma resposta possível para o problema de saber se queremos e para que queremos ser racionais, pelo menos tanto quanto podemos identificar a racionalidade com o uso de princípios de pensamento e decisão. Ora, a resposta de Nozick é diferente da resposta negativa de Stich referida no Capítulo 3, de acordo com a qual não queremos saber de verdade e racionalidade (recorde-se que Stich defendia esta posição com o seguinte sentido: enquanto agentes cognitivos, não queremos saber de verdade e racionalidade, na medida em que não são estas as finalidades da acção). Mas, se as respostas de Stich e Nozick são diferentes, é em parte porque elas não se situam no mesmo nível – Stich fala de cognição, Nozick fala de pensamento. Convém aqui recordar distinções que avancei na Introdução¹⁸: quando falo de pensamento por contraste com cognição, estou a propor que se considere cognição como os vários processos de informação, computação e decisão envolvidos na adaptação de um agente ao seu ambiente, processos esses que podem ser subpessoais e inconscientes; quando falo de pensamento, estou a pôr em relevo as dimensões qualitativa, voluntária e explicitamente normativa que os processos de representação, inferência e decisão podem atingir num ser humano consciente.

Ora, ao contrário de Stich, que fala de cognição, Nozick fala de pensamento quando considera as funções dos princípios. Ele

¹⁸ A distinção foi introduzida a propósito do tópico da Questão 7.

pensa que queremos sim ser racionais, nomeadamente queremos saber de princípios regendo os nossos raciocínios e decisões, e isto antes de mais porque os princípios têm funções específicas. Em última análise, Nozick defenderá que queremos saber de princípios não apenas para cuidarmos da qualidade do nosso pensamento teórico, nomeadamente científico e prático, por exemplo, moral e jurídico, e para coexistirmos socialmente mas também para cuidarmos de nós. Neste último âmbito, uma das razões para a importância dos princípios será a possibilidade que eles nos oferecem de nos porem 'em conjunto connosco mesmos', 'unânicos' (tomando o termo de um exemplo de F. Schick, da bem-afortunada mulher que era unânime consigo própria¹⁹), reconhecíveis para nós mesmos ao longo do tempo nas nossas ações e pensamentos, podendo confiar nos nossos eus futuros.

Um outro aspecto fundamental em *The Nature of Rationality*, para além do tratamento dos princípios, e conseqüentemente, da ligação que acabei de referir entre o problema dos critérios de correção (Problema 1) e a questão pessoal da racionalidade (Problema 9), ligação que pode revelar-se fundamental para a questão alargada da racionalidade (Problema 7), é a proposta de Nozick acerca do 'significado simbólico' (*symbolic meaning*). Esta proposta deve ser compreendida como uma proposta acerca de motivação para a ação (relacionada portanto com o Problema 6 – a questão dos desejos e da estrutura motivacional dos agentes), e permitirá a Nozick, por exemplo, ligar formulações da teoria da decisão à ética. A ideia de significado simbólico de ações é fundamental na teoria da racionalidade de Nozick e subjaz às propostas quanto à teoria da decisão que são centrais no livro. Na medida em que a teoria da decisão é, como tenho vindo a notar desde o início, uma perspectiva normativa de referência, ou mesmo dominante, nos estudos da racionalidade, as propostas de 'reformulação' de Nozick a partir da noção de significado simbólico são exemplos de um possível ângulo de crítica aos 'modelos de racionalidade'.

¹⁹ SCHICK 1997: 57.

Da teoria da decisão à ética: princípios e significado simbólico

Os princípios e as suas funções constituem um ótimo ponto de observação da importância da racionalidade na nossa unidade enquanto agentes cognitivos, e, mais do que agentes cognitivos, pessoas (pessoas são agentes cognitivos com características bastante especiais, nomeadamente capacidade de consciência, vontade, identidade pessoal, pensamento auto-controlado e coexistência em sociedade). Na caracterização que faz daquilo que nós fazemos com princípios, Nozick assume que é importante e desafiador conhecermos-nos a nós próprios e que analisar as funções dos princípios no pensamento e na ação é um passo essencial em direção à compreensão das razões por que queremos (ou pelo menos podemos querer) ser racionais. Mas o que é que Nozick está a fazer quando pergunta pelas funções dos princípios? Para compreender a questão olhemos para duas das definições de racionalidade com que temos vindo a lidar: de acordo com a definição instrumental, é racional o agente que mobiliza eficazmente meios em ordem a fins. Já sabemos que a aplicação da definição vai desde agentes sem consciência e vontade (nos quais o comportamento de acordo com razões é uma questão de *free floating rationales*, razões não apercebidas) até pessoas, i. e., agentes cognitivos capazes de utilizarem e analisarem conscientemente standards de ações e pensamento, e até mesmo de colocarem a si próprios a questão dos critérios de correção. Temos aqui, obviamente, outra definição de racionalidade: racionalidade é conformidade com standards de correção, com princípios (que estão para casos particulares numa relação tipo-espécime). Ora, aquilo que Nozick faz ao perguntar pelas funções dos princípios é ligar a questão da racionalidade instrumental à questão dos standards de racionalidade. A pergunta pode ser reformulada assim: quais são os fins relativamente aos quais a ação e pensamento conformes a princípios constituem os meios?

Nozick chama 'Princípios' a regras gerais de pensamento e acção – são exemplo de princípios leis científicas, leis jurídicas e princípios morais. Estas regras gerais agrupam acções, casos, instâncias e assim mantêm o fio, a unidade, o que há de comum, em procedimentos de pensamento e acção ao longo do tempo (em ciência, no direito, em termos pessoais). A acção (e o pensamento) de acordo com princípios devem ser concebidos em contraste com a acção levada a cabo por casualidade, capricho ou mera maximização do interesse próprio. Como afirmei, o primeiro passo de Nozick em *The Nature of Rationality* consiste em analisar da perspectiva da racionalidade instrumental – perguntando por funções, finalidades, utilidades, benefícios – os princípios. A análise leva-o a defender que os princípios têm funções intelectuais, interpessoais, sociais e pessoais. Como introdução ao esclarecimento dessas funções, consideremos o exemplo do próprio Nozick: os princípios funcionam de forma análoga a uma decisão judicial tomada não à luz do que seria o melhor resultado para o caso particular C em consideração naquele momento mas segundo uma regra geral que enquadre precedentes, casos hipotéticos e o próprio caso em consideração. Tentar formular um princípio geral que se aplique ao caso em consideração de forma aceitável funciona assim como um teste para o juízo particular que se desejaria aplicar ao caso. Este ajustamento entre juízos sobre casos particulares e princípios gerais é o melhor que podemos conseguir quando não estão disponíveis quadros explicativos de aplicação mecânica²⁰. Um dos ganhos da conformidade a princípios é o facto de estes acabarem por funcionar como dispositivos de previsão (da aplicação das regras a novos casos). Daí uma especificação da definição de princípio proposta por Nozick: «Os princípios são dispositivos de transmissão de probabilidade ou suporte, que fluem de dados ou casos, via o princípio, para juízos e previsões

²⁰ Nozick está obviamente a retomar aquele a que chamei o problema do critério de correcção (cf. Introdução, Problema 1, e capítulo 3).

acerca de novas observações ou casos cujo estatuto seria de outra forma desconhecido ou menos certo²¹.

Em termos intelectuais interpessoais, os princípios servem como dispositivos para facilitar a chegada a uma decisão correcta acerca de casos, limitando a influência de factores irrelevantes, tais como preferências e inclinações individuais (por exemplo, em contextos legais proporcionando a equanimidade e impedindo injustiças) e deste modo propiciando a justificação de raciocínios e acções individuais perante outrem. Na medida em que os princípios podem servir aos indivíduos para superar tentações, nomeadamente a tentação de agir segundo os impulsos do momento, e prosseguir os seus fins a longo prazo, eles servem ainda, em termos interpessoais, eventualmente, para tornar os comportamentos de outrem previsíveis. Aliás, os princípios servem em geral para tornar comportamentos socialmente previsíveis, constituindo uma solução para o problema do 'jogo da coordenação' de acções, sem o qual nenhuma sociedade se sustenta. Esta explicação aplica-se ainda à forma como princípios éticos podem funcionar entre pessoas. Mas a verdade é que tudo isto parece explicar apenas a razão por que poderíamos querer que os outros tivessem princípios, pensassem e agissem segundo princípios – por que razão os desejaríamos para nós próprios?

Segundo Nozick, e antes de mais, os princípios só podem ter as funções interpessoais referidas porque têm determinadas funções pessoais. Nomeadamente, os princípios permitem, em termos pessoais, ultrapassar tentações imediatas quando há coisas mais importantes no futuro, comprometem os nossos eus futuros com os nossos projectos presentes, fazendo com que seja razoável contar não apenas com o comportamento dos outros mas também e antes de mais com comportamentos desses eus futuros. Os princípios permitem assim a integração da identidade pessoal ao longo do tempo e ligam-se ao projecto do entendimento próprio, projecto este que é para Nozick uma questão importante,

²¹ NOZICK 1993: 5.

culminante, da racionalidade²². O pensamento de acordo com princípios possibilitará também, eventualmente – e isso pode ser importante para a identidade pessoal de algumas pessoas – o entendimento do entendimento, ou seja, como se vê pelo que faz o próprio Nozick, o próprio aprofundamento teórico daquilo que fazemos com princípios, e daquilo a que os princípios dizem respeito (e que é, basicamente, tudo aquilo sobre que podemos pensar e agir). Em suma, se em termos interpessoais e sociais, os princípios servem para justificar raciocínios e acções perante outrem e para tornar os comportamentos previsíveis, em termos pessoais os princípios servem, segundo Nozick, para a nossa unificação e determinação. Isto é, eles servem para vermos ao espelho um rosto mais determinado por nós do que aquilo que demos por nós sendo (e que tanto quanto vamos sabendo era já um animal racional desenhado pela natureza), possibilitando-nos ainda, em situações de escolha, filtrar determinadas opções, facilitando a tarefa de escolher à criatura de racionalidade limitada que somos, e permitindo-nos reconhecermo-nos como uma e a mesma pessoa ao longo do tempo.

A marca de um princípio é então a ligação de uma particular decisão ou juízo a uma classe de acções ou pensamentos, e também o compromisso de continuar a agir e pensar assim no futuro. Por tudo isto, adoptar um princípio e agir ou pensar de acordo com ele faz ‘uma coisa’ (um pensamento, uma decisão, uma acção) ‘estar por muitas coisas’. Ora, perante o olhar de teórico da decisão de Nozick, isso redundará na alteração da utilidade de uma particular acção, decisão ou pensamento conforme ao princípio (torna-a mais elevada). Esta ideia é central na teoria da racionalidade simbólica de Nozick, e é, como já afirmei, precisamente através da discussão das funções dos princípios que Nozick introduz um tema fundamental de *The Nature of Rationality*, o tema do significado simbólico de opções de acção, que subjaz às suas propostas no âmbito da teoria da decisão. O significado

²² Cf. aquilo a que chamei na introdução o ‘Problema Pessoal da Racionalidade’.

simbólico é uma chave para o tratamento da natureza da nossa racionalidade, uma vez que esta, segundo Nozick, envolve não apenas o uso de princípios no pensamento e na acção mas também a decisão acerca de que princípios usar para pensar e agir e o valor que o uso e a decisão podem ter para nós.

Consideremos aquilo a que Nozick chama ‘significado simbólico’. Com esta noção, Nozick pretende chamar a atenção para o facto de, para além de serem o acto específico que são, os actos poderem simbolizar uma determinada coisa. Este facto tem que ser encarado pela teoria da decisão, quando considera na situação de deliberação utilidades, probabilidades e maximização da utilidade esperada. Se isso não for feito, estar-se-á a ocultar parte da complexidade em jogo na escolha racional dos humanos, e eventualmente a considerar como irracionais opções de acção que são função da imputação de um determinado valor simbólico a uma particular opção. A questão é de facto fundamental quando consideramos a natureza da situação de deliberação. Antes de mais, como pergunta Nozick, sem significado simbólico, o que desejaríamos? Como explicaríamos que tantas coisas fossem desejadas e escolhidas em culturas como as nossas pelas pessoas? Poderíamos dizer que sem significado simbólico desejaríamos simplesmente conforto material, segurança física e prazer sensual – mas Nozick duvida que mesmo essas coisas sejam desejadas ‘a nu’, sem significado simbólico²³. Olhemos para um exemplo da forma como a relação ‘estar por’ pode intervir, através do significado simbólico, na utilidade em causa numa determinada decisão. Por exemplo, porque é que pode acontecer que evitemos quebrar princípios, agir contra eles, uma vez assumidos, mesmo perante fortes tentações? Imaginemos os seguintes cenários: trata-se, por exemplo, de (i) suspender a aplicação de uma pena a alguém que nos é particularmente querido, ou (ii) chegar a uma conclusão que desejaríamos não conhecer – por

²³ NOZICK 1995: 30. Nozick pensa na ligação entre estes aparentemente simples objectos de desejo e questões tais como a libertação da dependência infantil, o sucesso em competição, a relação com os progenitores, etc.

exemplo, a partir das provas disponíveis um pai conclui que o crime foi cometido pelo filho, ou (iii) interromper a aplicação de legislação de apoio aos mais pobres uma vez conhecidos os resultados que nos dão boas razões para pensar que o apoio não está a ter os resultados previstos e desejados. De acordo com Nozick, a razão por que mesmo perante a tentação os humanos frequentemente evitam agir contra os princípios que professam reside no facto de mesmo um só caso de quebra simbolizar o abandono total do princípio, o que parece indesejável. Desta forma, os princípios funcionam como, suspendendo o regime normal de funcionamento da racionalidade como cálculo de utilidades sem atender a este valor simbólico. Isso significa, segundo Nozick, que os princípios marcam uma área-limite para o domínio exclusivo de considerações de utilidade sem valor simbólico, mas de forma alguma o limite de considerações de racionalidade. Vejamos agora a forma como a introdução de considerações acerca de utilidade simbólica pode pesar sobre a forma de encarar questões clássicas de decisão.

Uma das intenções práticas de Nozick em *The Nature of Rationality* é especificar a razão do seu desacordo relativamente a formulações da teoria da decisão que não incluam a consideração do significado simbólico de acções. Basicamente, Nozick considera que é necessário encontrar um lugar para aquilo a que chama 'utilidade simbólica' na teoria da decisão. Esta 'utilidade simbólica' de opções – que de acordo com Nozick intervém constantemente nas decisões que tomamos – não é uma utilidade de tipo diferente mas um tipo diferente de conexão de uma acção aos seus efeitos, que a teoria da decisão deve considerar. A conexão de uma acção aos seus resultados ou efeitos é usualmente, no contexto da teoria da decisão, pensada de duas formas fundamentais, que dão lugar a duas versões da teoria da decisão, a versão causal e a versão evidencial. De acordo com a versão causal da teoria da decisão, a ligação acção-efeitos é causal (um acto é pensado como a causa de um resultado desejado), de acordo com a versão evidencial a ligação acção-efeitos é pensada em termos de

evidência, que faz acreditar que com dada probabilidade um acto específico produzirá o resultado desejado. A utilidade esperada de uma acção será pensada de formas diferentes, de acordo com a versão adoptada. Pensar-se-á assim, respectivamente, que aquilo que o agente considera quando delibera é a 'utilidade causal esperada' ou a 'utilidade evidencial esperada' dos resultados das suas opções. As propostas de Nozick acerca de utilidade simbólica e valor de decisão de uma acção que considera esta utilidade simbólica, devem ser encarada no contexto de discussões entre teóricos da decisão 'causais' e 'evidenciais'.

Em termos gerais, Nozick quer chegar a propor uma nova regra de decisão racional – a maximização do valor de decisão (DV, *decision value*) – de acordo com a qual um agente racional quando age, age de forma a maximizar o valor de decisão. Nozick propõe a seguinte fórmula para o valor de decisão de uma acção A:

$$DV(A) = W_c \times CEU(A) + W_e \times EEU(A) + W_s \times SU(A).$$

A fórmula captura o seguinte: o valor de decisão de uma acção A (e agir racionalmente é, para Nozick, maximizar o valor de decisão) é igual à utilidade causal esperada (CEU – *causal expected utility*) de A mais a utilidade evidencial esperada (EEU – *evidential expected utility*) de A mais a utilidade simbólica esperada (SU – *symbolic utility*) de A, atribuindo a cada uma delas um determinado 'peso' (W_c , W_e , W_s).

Basicamente, e é isso que está inscrito na fórmula, Nozick pensa que há boas razões para pensar a ligação acção-efeitos quer em termos causais quer em termos evidenciais quer em termos simbólicos e que tudo isso deve ser considerado na regra da racionalidade. O que Nozick pensa que não existe são boas razões para optar exclusivamente por uma interpretação única, causal ou evidencial, da ligação. Noutros termos, o que Nozick está a fazer é encarar a própria questão 'Quais serão os princípios de decisão correctos?' como um problema de decisão. Ora, sob essa perspectiva, ele pensa que não é razoável, dada

a complexidade da situação, ter confiança absoluta num único princípio de decisão. É esta a convicção que subjaz à análise que Nozick faz de dois casos centrais e clássicos da teoria da decisão, o Problema de Newcomb e o Dilema do Prisioneiro. Pondo em relevo a oscilação das nossas intuições perante ajustamentos na formulação do problema de Newcomb, Nozick interpreta essa oscilação de uma forma que apoia a sua proposta.

A formulação do **problema de Newcomb** deve-se ao próprio Nozick²⁴. O problema é o seguinte: apresenta-se a uma pessoa duas caixas, a da direita contendo \$1 000, a da esquerda contendo ou \$1 000 000 ou nada. A pessoa tem duas opções: escolher as duas caixas ou só a caixa da esquerda. O factor ‘perturbador’ ou despoletador do problema é um ser que prevê a escolha da pessoa. Se ele prever que a pessoa vai escolher as duas caixas, então não põe nada na caixa da esquerda (a que teria \$1 000 000), se ele prever que a pessoa vai escolher só a caixa da esquerda põe lá um milhão de dólares. A ordem é a seguinte: primeiro, o ser faz a previsão, depois põe ou não põe o milhão de dólares na caixa da esquerda, depois a pessoa escolhe. A pessoa que escolhe não sabe qual a previsão que foi feita.

O problema consiste não apenas em saber qual é a escolha racional, i. e., que caixa deve ser escolhida pela pessoa, mas também o que é que há de errado com as duas formas de pensar sobre o problema, a forma ousada e a forma cautelosa. De acordo com a primeira, a forma ousada, é preferível arriscar escolher a caixa que pode ter um milhão de dólares. De acordo com a segunda, a forma cautelosa, mais vale escolher as duas caixas. Porque é que de acordo com a primeira interpretação é preferível escolher a caixa da esquerda? Porque se a pessoa escolher as duas, essa foi certamente a escolha prevista, e só ganhará 1 000 dólares, mas se escolher apenas a caixa do milhão

de dólares é provável que tenha sido essa a escolha prevista (e nesse caso ganha um milhão de dólares). De acordo com a segunda interpretação é preferível escolher as duas caixas porque a previsão já foi feita de qualquer maneira e se foi previsto que a escolha recairia sobre as duas caixas a pessoa ganha ainda mil dólares, se foi previsto que a pessoa escolheria apenas a do milhão de dólares, ela ganhará um milhão e mil dólares.

Segundo Nozick, de início as pessoas pensam sobre o problema ou de uma forma ou de outra (de forma ousada ou de forma cautelosa). No entanto, se forem confrontadas com diferentes valores, a sua abordagem modificar-se-á. Por exemplo, se na caixa da direita em vez de mil dólares estiverem 900 000, os partidários da primeira interpretação desistem do risco e escolhem as duas caixas. Por outro lado, se na caixa da direita em vez de mil dólares estiver um dólar, os partidários da segunda interpretação preferem agora arriscar num milhão. Que sentido tem esta alteração de estratégia? Ela não é compreensível do ponto de vista da teoria da decisão: mais dinheiro esperado deveria ser sempre preferível a menos dinheiro, independentemente dos valores específicos em jogo.

Nozick pensa que o problema pode ser resolvido considerando a utilidade simbólica das acções. As escolhas são feitas também *por aquilo que simbolizam*. Escolher apenas a caixa que pode ter um milhão simboliza para algumas pessoas gostar de correr riscos para ter recompensas e a acção determina a sua identidade de “aventureiras” (com os valores da segunda versão, escolher um milhão seria mais estupidez do que aventureirismo). Para o segundo tipo de pessoas, escolher as duas caixas simboliza uma certa sensatez – mas se o valor da primeira caixa for reduzido a um dólar, a sensatez passaria a algo de muito pior, como falta de visão. Segundo Nozick, a introdução das considerações de utilidade simbólica explica os comportamentos-tipo. Repare-se que a utilidade simbólica não é apenas mais uma variável a introduzir no cálculo da utilidade esperada mas uma utilidade de outro tipo, a considerar separadamente. E o que Nozick está a defender é

²⁴ NOZICK 1969, *Newcomb's Problem and Two Principles of Choice* (o problema a que o título do artigo de Nozick se refere foi concebido pelo físico William Newcomb e relatado a Nozick por um amigo comum).

que se perguntarmos 'o que se deve escolher?' a resposta será 'devemos escolher conforme o que queiramos ser'.

O Dilema do Prisioneiro é outro problema tratado na teoria da decisão e foi atrás exposto (Capítulo 2). Recordo a situação: dois criminosos foram presos por terem cometido um crime juntos. Cada um está numa cela separada sem qualquer contacto com o outro. A polícia sugere-lhes as seguintes opções: se um confessar e o outro não o primeiro pode sair em liberdade e o segundo tem 12 anos de prisão. Se ambos confessarem, cada um tem 10 anos de prisão. Se nenhum confessar cada um cumpre 2 anos de prisão. O problema neste caso é que para cada um dos prisioneiros individualmente faz mais sentido confessar (eles pensam assim, se o outro confessou então mais vale eu confessar senão fico com 12 anos de prisão; por outro lado, se o outro não confessou mais vale eu confessar e vou em liberdade). Assim, do ponto de vista individual faz sempre mais sentido confessar. No entanto, se ambos pensarem assim e ambos confessarem acabam por ficar 10 anos na prisão enquanto que se ambos não confessassem ficavam com 2 anos cada. Desta forma, o que parece fazer mais sentido do ponto de vista individual acaba por prejudicar ambos os prisioneiros.

O Dilema do Prisioneiro é usualmente tomado como referência para analisar interações sociais, cooperação social, e o que está em jogo na alternativa entre a "acção dominante" (que é não cooperar) e a "acção de cooperação". A escolha dominante (que é no caso do Dilema do prisioneiro, não cooperar) pode parecer vantajosa para o indivíduo mas pode acabar por ter consequências piores do que a escolha cooperante. Em casos como este revela-se o valor de decisão dos princípios, ou a forma como princípios podem alterar a situação de deliberação: ao obrigarem os agentes a incorporar nas suas decisões outras razões para além do cálculo instrumental de utilidade, os princípios permitem uma 'saída' do dilema. No caso dos prisioneiros, a escolha cooperante pode ser preferida, por exemplo, porque o indivíduo tem como princípio ético «não denunciar os amigos». Se ambos tiverem este princípio, ambos acabarão por beneficiar.

Não há, no entanto, forma de quantificar princípios semelhantes, nem forma de saber até que ponto os outros os vão seguir. De qualquer forma, o que Nozick pretende com as suas abordagens do problema de Newcomb e do Dilema do Prisioneiro Nozick é mostrar como a inclusão da utilidade simbólica no 'cálculo de utilidades' alarga aquilo que se entende por 'racionalidade'. Sem entrar no detalhe das análises de Nozick, o que interessa ter em conta é que ele considera que a existência de vários tipos de ligação de uma acção aos seus resultados, (causal, evidencial, simbólica) não pode ser ignorada pela teoria da decisão ao considerar a racionalidade de decisões (daí a fórmula do valor de decisão de uma acção). Ora, agir eticamente, agir por princípios, pode ter valor simbólico para o agente, e é isto o que permite a ligação entre a teoria da decisão e a ética. Em termos concretos e exemplificando, Nozick está a propor introduzir alterações na teoria da decisão que nos permitam conceber a racionalidade de situações como as seguintes: medidas de controlo anti-droga ou legislação relativa a um ordenado mínimo podem, para além da utilidade causal e evidencial esperada, simbolizar reduzir o consumo de droga ou ajudar os mais pobres, e esse factor influi na deliberação racional. Da forma mais geral possível, aquilo que Nozick está a defender é que agir eticamente pode simbolizar ser uma criatura racional e «a utilidade destas coisas grandiosas simbolicamente expressa e instanciadas pela acção torna-se incorporada na utilidade simbólica dessa acção e assim no valor de decisão dessa acção»²⁵.

*Regras práticas para a racionalidade teórica.
A verdade como fim cognitivo*

Tenho vindo a afirmar desde o início que a forma mais comum de falar de racionalidade, que deve sempre ser tomada em consideração, é a racionalidade na acção. A racionalidade na acção

²⁵ NOZICK 1993: 62.

conduz-nos à definição instrumental, de acordo com a qual uma acção é racional quando mobiliza eficazmente os meios que permitem atingir os fins desejados. Mas como é que uma definição 'prática' de racionalidade, que é acerca daquilo que devemos fazer, se relaciona com a racionalidade teórica, característica da vida mental de um agente que acredita naquilo que deve acreditar? Como se relaciona uma definição instrumental, puramente comportamental e portanto 'cega ao interior do agente', ou uma definição da teoria da decisão, da racionalidade como maximização da utilidade esperada, a qual pode também ser cega ao processo pelo qual o agente atinge um fim pela acção (pode ser totalmente acidental), com a 'definição' teórica de racionalidade, a ideia segundo a qual racionalidade é racionalidade nas crenças, ser racional é acreditar aquilo que se tem razão para acreditar, poder justificar as crenças que se tem? Uma forma de ver a ligação entre as duas definições passa pela questão cognitiva e epistemológica da fiabilidade das crenças, que já foi referida no Capítulo 3. A fiabilidade, ao fazer-nos olhar para ligações causais entre representações e processos no agente e a realidade, pode fazer a ligação entre definições 'comportamentais' de racionalidade, 'cegas' ao interior dos agentes, e a problemática da racionalidade teórica, racionalidade nas crenças, abordada a partir do interior dos agentes. A ligação será a seguinte.

Agentes cognitivos têm fins, por exemplo, e Nozick considera que a verdade é um fim cognitivo instrumentalmente baseado no sentido em que a eficácia das crenças verdadeiras terá ajudado à adaptação dos agentes cognitivos ao ambiente: agentes cognitivos funcionam de forma dirigida a acreditar verdades acerca do seu ambiente. Considerar a verdade como um fim cognitivo deixa obviamente em aberto saber o que é a verdade, por que é que ela é um fim valioso para o agente, etc., e Nozick tem o cuidado de afirmar desde logo que o valor da verdade para o agente cognitivo não se identifica com a natureza da verdade (isto permite-lhe nomeadamente considerar que diferentes teorias da verdade – correspondentistas, coerentistas, etc. – são diferentes

hipóteses explicativas da natureza de uma propriedade que está lá, que subjaz à eficácia evocada quando se fala de valor instrumental da verdade). Ora, de acordo com Nozick, o interesse pela verdade, o valor da verdade para o agente cognitivo, teve uma origem instrumental, no sentido em que a eficácia das crenças verdadeiras terá ajudado à adaptação dos agentes cognitivos ao ambiente: de facto, 'meios mobilizados' e 'fins desejados' na acção, na definição instrumental de racionalidade, são o que são mediante crenças do agente, i. e., representações, e é melhor que estas sejam representações verdadeiras e representações em ajuste constante. As crenças são precisamente representações em ajuste constante e num ambiente que muda constantemente, tornam agentes mais aptos do que, por exemplo, o condicionamento operante. Esta é a razão pela qual acreditar verdades é valioso para os agentes, mas não significa que agentes cognitivos, por exemplo, os nossos antepassados, desejassem acreditar em verdades porque isso era instrumentalmente útil mas sim que devido à utilidade de acreditar verdades, algum interesse pela verdade nas crenças, assim como curiosidade pela procura de verdades, teriam sido seleccionados (*selected for*). A ligação entre a fiabilidade de processos e razões para acreditar estabelecer-se-á considerando que agentes cognitivos mais frequentemente chegam a crenças verdadeiras – e o valor das crenças verdadeiras para os agentes foi justificado – se sustentarem crenças suportadas por razões. Basicamente, razões contribuem para fixar mais crenças verdadeiras.

No entanto, Nozick pensa que fins cognitivos instrumentalmente baseados como a verdade podem ganhar uma autoridade própria, e que foi exactamente esse o caso: um interesse (pela verdade) cuja origem foi instrumental expandiu-se de tal forma que aquilo a que hoje chamamos racionalidade nas crenças vai muito além da procura da verdade nas crenças acerca de meios-mobilizáveis-fiáveis e fins-no-ambiente- imediato. De resto, e como em agentes cognitivos do nosso tipo a racionalidade nas crenças, as razões para acreditar, dizem respeito a sistemas de

crenças muito complexos, não se dá sequer o caso de as crenças a fixar poderem depender de um único método ou sistema, tal como a evidência empírica, e assim o problema das razões para acreditar torna-se, e é aí que Nozick quer chegar, um complexo problema de decisão – basicamente formulável como ‘*o que devemos fazer para acreditar aquilo em que devemos acreditar?*’. Nozick vem por isso a analisar o problema da racionalidade nas crenças como um problema relativo a processos como meios para atingir fins, e é por essa razão que coloca questões como estas: uma pessoa cujas crenças fossem formadas por processos com outras finalidades que não a de atingir a verdade (finalidades tais como ser apreciado, ou ser feliz) deveria ser considerada irracional? Crenças verdadeiras adquiridas por processos como levar uma pancada na cabeça ou tomar um comprimido seriam racionais? Noutras palavras: considerando que a racionalidade nas crenças diz respeito por um lado a razões para acreditar e por outro lado à fiabilidade dos processos que originam crenças, considerando que essas razões para acreditar e esses processos não podem estar totalmente sob o controlo de um agente cognitivo complexo como uma pessoa, o que devemos fazer para acreditar naquilo que devemos acreditar?

Em agentes cognitivos como nós, as razões para acreditar são constituídas e funcionam numa rede de crenças interligadas umas às outras, de modo que a racionalidade de uma crença tem que ser avaliada através de complexos processos de raciocínio e evidência. Nestas circunstâncias coloca-se um meta-problema prático para a racionalidade teórica: será que podemos formular regras práticas (*decision-theoretical*) de racionalidade para chegar a boas crenças? Como poderemos aplicar considerações de racionalidade prática, acerca de racionalidade nas decisões, às questões de racionalidade teórica, racionalidade nas crenças, para de alguma forma exercermos controlo sobre aquilo em que devemos acreditar? A proposta de Nozick passa pela noção básica de **valor de credibilidade** de uma asserção a acreditar. O valor de credibilidade de uma asserção é gerado no seio de uma rede

de (muitos tipos) de razões a favor e contra, com pesos diferentes e determinados, a partir da ponderação de probabilidades, valor explicativo, simplicidade, evidência contrária e falsificação, etc. Como Nozick nota, questões metodológicas discutidas na literatura da filosofia da ciência²⁶ poderiam ser reencontradas aqui funcionando dentro de rede, ou como fenómenos emergentes do funcionamento desta. A proposta de regras para a racionalidade nas crenças apresentada em *The Nature of Rationality* tem cabimento precisamente como especificação desta situação, e baseia-se nos valores de credibilidade gerados e no uso destes para chegar a crenças. As regras são considerações de ‘ética da crença’, daquilo que é útil acreditar, considerando em geral como finalidade de agentes cognitivos a preocupação com a verdade das crenças numa situação em que a prossecução de tal finalidade depende da consideração de um grande número de razões, e de razões de muitos tipos. Nozick imagina então um sistema cognitivo, capaz de aprendizagem (i. e., os pesos dos factores que geram valores de credibilidade para cada asserção h a ser avaliada são modificáveis por *feedback*), que pode ser uma pessoa, que se importa em termos globais com a verdade. Nozick acentua o facto de o sistema poder ser globalmente considerado racional sem nenhum dos aspectos do seu funcionamento ser óptimo. As regras que propõe para avaliar asserções candidatas a serem ‘acreditadas’ são as seguintes:

R1. Não acreditar em h se alguma asserção alternativa incompatível com h tem um valor de credibilidade mais elevado do que h.

R2. Acreditar em h apenas se a utilidade esperada de acreditar em h não for menor do que a utilidade esperada de não ter qualquer crença sobre h

²⁶ Referentes, por exemplo, a suporte pela evidência disponível, falsificação de hipóteses, simplicidade de hipóteses, evitamento de hipóteses *ad hoc*, controlo de variáveis, probabilidades, etc.

R3. Acreditar em h apenas se o seu nível de credibilidade for suficientemente elevado tendo em conta o tipo de asserção de que se trata.

R4. Acreditar em h se não existe uma asserção alternativa a h com um valor de credibilidade mais alto, e se o valor de credibilidade de h for suficientemente alto, dado o tipo de asserção que h é, e a utilidade esperada de acreditar em h for pelo menos tão elevada como a utilidade esperada de não ter qualquer crença acerca de h (basicamente: acreditar em h se h não é excluído pelas três primeiras regras).

R5. Acreditar em h se o valor de decisão de acreditar em h é pelo menos tão elevado como o de não ter nenhuma crença acerca de h (trata-se de uma reformulação de R2 de acordo com a formulação da teoria da decisão em termos de valor de decisão – a qual, recorde-se, envolve utilidade causal, evidencial e simbólica, em substituição da ‘utilidade esperada’)

Nozick acrescenta ainda uma sexta regra, a aplicar a inferências dedutivas:

R6. Acreditar q porque q é inferido das premissas $p_1 \dots p_n$ numa inferência dedutiva explícita apenas se cada uma das premissas p_i for acreditada e apenas se a sua conjunção $p_1 \& p_2 \dots \& p_n$ for também acreditada.

Sublinho dois pontos importantes desta exploração de Nozick. Primeiro, algumas regras estão formuladas em termos de valor de credibilidade, algumas em termos de valor de decisão, de qualquer forma quer o valor de credibilidade quer o valor de decisão trazem considerações de racionalidade prática para o âmbito da racionalidade teórica. Segundo, esta exploração faz-nos pensar que as considerações acerca de racionalidade nas crenças, tradicionalmente trabalhadas por filósofos, que procuram formular princípios explícitos, regras de inferência racional, para formas dedutivas e não dedutivas de chegar a crenças, ‘a partir

de cima’, deverão, se de facto todos os factores evocados por Nozick para gerar o valor de credibilidade de asserções devem ser tomados em consideração quando encaramos a questão da racionalidade teórica do ponto de vista da racionalidade prática (e da eficácia), perder o estatuto privilegiado e mesmo, possivelmente, o ‘apelo intuitivo’. De facto, de acordo com Nozick, regras da racionalidade em sistemas considerados sob esta perspectiva levarão em consideração aspectos subpessoais do processamento. Dados os aspectos cognitivos subpessoais envolvidos, considerações acerca de racionalidade nas crenças não cabem apenas à filosofia mas também às várias disciplinas da cognição, podendo tais considerações ressaltar mais facilmente de uma simulação por computador do que de uma análise do género cartesiano para chegar a ‘regras para a direcção do espírito’. Parece plausível que assim seja na medida em que aquilo a que tradicionalmente se chama as regras da racionalidade, princípios explicitamente formuláveis e que seriam supostamente aplicados conscientemente a partir de cima por um pensador e decisor racional (regras *surveyable*, como diz Nozick), mais propriamente ‘emergem’ como descrições globais de um funcionamento que não é ‘supervisionado’ e que é porventura melhor descrito ao nível subpessoal. Voltamos assim ao problema do Capítulo 3: estamos perante mais uma consequência da naturalização da epistemologia.

*Limites da racionalidade instrumental:
natureza e valor da racionalidade.
Desejos e princípios*

A última questão de Nozick em *The Nature of Rationality* é a questão dos limites da racionalidade instrumental. A importância da questão reside no seguinte: a racionalidade instrumental é definitivamente o núcleo da noção de racionalidade, o ponto de intersecção de todas as teorias da racionalidade – isso não

é questionável. A questão, para Nozick, é saber se isso é tudo, se o conceito de racionalidade é esgotado pela racionalidade instrumental. Ora, Nozick pensa que não: a natureza e a origem da racionalidade podem ser totalmente instrumentais, os standards da racionalidade nasceram certamente com impurezas²⁷, mas o valor da racionalidade não é totalmente instrumental. Como vimos, a abordagem inicial da racionalidade em *The Nature of Rationality* – a discussão das funções dos princípios, da utilidade prática específica de seguir princípios – levou a considerações de racionalidade instrumental muito longe, até aos próprios standards de racionalidade. Nozick termina o livro analisando a forma como a reformulação da teoria da decisão proposta, que inclui a utilidade simbólica, se repercute no valor da nossa racionalidade para nós próprios. Segundo Nozick, é simbolicamente importante para nós que nem todas as nossas actividades sejam dirigidas a satisfazer os nossos desejos. Ora, princípios são precisamente uma forma de controlar e reformular desejos em nós. Seguir princípios pode ter uma utilidade simbólica para nós, enquanto humanos. E nós queremos saber desse significado simbólico, independentemente do que o emprego de princípios causa ou produz – é nesse sentido que não somos racionais de forma simplesmente instrumental (repare-se, entretanto, que não seria simples, para o defensor da natureza meramente instrumental da racionalidade, defender que o que Nozick propõe é irracional – por que seria um particular desejo mais irracional do que qualquer outro?).

Vejamos de forma mais específica em que sentido podemos defender que a racionalidade não é meramente instrumental. Tomando como referência a decisão racional (e já vimos que segundo Nozick a nossa racionalidade envolve não apenas o uso de princípios de raciocínio e decisão mas também a decisão

acerca de que princípios de decisão e de raciocínio usar) e recordando a proposta de Nozick acerca de valor de decisão, termos o argumento pelo qual Nozick conclui que a racionalidade não é meramente instrumental. O argumento de Nozick é o seguinte. Uma decisão racional é uma decisão que maximiza o valor de decisão de uma acção, valor de decisão que é uma soma ponderada da utilidade, causal, evidencial e simbólica dessa acção. A racionalidade instrumental é capturada pela noção de utilidade *causal* esperada. Ora, a utilidade causal esperada é apenas um aspecto da racionalidade. Logo, a natureza da racionalidade excede a mera instrumentalidade. Essa excedência da racionalidade relativamente à racionalidade instrumental constituía já, de facto, o núcleo das propostas de Nozick acerca de racionalidade na acção e racionalidade nas crenças

Com este passo além da racionalidade instrumental, Nozick avança, naturalmente, alguns passos para além de Hume, sem no entanto se identificar com o posicionamento kantiano quanto à questão dos desejos, da razão e da motivação para a acção. A fundamental razão pela qual, para Nozick, os seres humanos não são seres humanos (movidos por paixões, sendo a razão apenas uma serva destas) é o facto de utilizarem **princípios** como meio de controlar e formatar os **desejos**. Não podemos dizer que por definição os nossos desejos biologicamente dados, que damos por nós tendo, são racionais – nós somos criaturas, começamos com determinados desejos e isso marca um limite da nossa racionalidade. Como nota Nozick, embora possamos alterar esses desejos, e mesmo perguntar por que é que eles não são satisfeitos, é daí que partimos e é a partir daí que chegamos onde quer que seja. Também não podemos, seguindo Kant, considerar que nos movemos a puros princípios e isso é a racionalidade – não podemos transcender desse modo os nossos desejos, deixá-los simplesmente para trás. A tentativa kantiana de separar princípios da racionalidade dos desejos e finalidades de agentes cognitivos naturais pura e simplesmente falha. Mas isso não significa que fiquemos com os desejos como

²⁷ No sentido em que a natureza da racionalidade inclui a Natureza na racionalidade, i. e., o facto de a racionalidade ser resultado de um processo de selecção natural e adaptação.

dados brutos, como quis Hume. Isto é assim não apenas porque a nossa racionalidade consiste na utilização de princípios como meio de controlar e formatar os desejos (e o funcionamento dos princípios é um passo além da racionalidade instrumental), mas também porque é possível, contrariamente ao que Hume defendeu, explicitar constrangimentos quanto à racionalidade substantiva de desejos e fins, nomeadamente quanto à racionalidade de conjuntos de preferências (Nozick não defende que preferências individuais de agentes sejam irracionais, mas que grupos de preferências podem definitivamente ser irracionais)²⁸.

²⁸ Rational Preferences (NOZICK 1993: 139-151).

Conclusão

Este livro nasceu não apenas da necessidade prática de mapear os estudos da racionalidade na filosofia e na ciência cognitiva mas também da necessidade de pensar as consequências da naturalização da epistemologia no tratamento da questão da racionalidade. Uma vez que a polémica em torno da noção quineana de epistemologia naturalizada gira em torno de saber se se trata ou não de expulsar a normatividade do estudo do conhecimento, e assim de uma possível passagem de testemunho da filosofia para a ciência, nada melhor, para testar essa possibilidade, do que tomar directamente como ponto de referência o estudo, a partir da ciência cognitiva, de alguma coisa que parece ser tipicamente normativa, a racionalidade. A questão da racionalidade é um terreno tradicional da filosofia, no sentido em que é difícil ou impossível nomear um filósofo que não tenha pensado sobre os princípios que guiam a forma como pensamos e como agimos (vários desses filósofos foram aqui convocados, desde Aristóteles a Hume, Kant e Davidson e os princípios em causa vão desde a lógica até à filosofia moral). Mas a racionalidade é também um terreno no qual a ciência cognitiva (via a psicologia cognitiva, a psicologia evolutiva, a neurociência, etc.) definitivamente entrou. Noutras palavras, a questão da racionalidade é uma questão antiga, é a questão da avaliação do pensamento pelo pensamento – o que é novo no contexto que constitui o horizonte do presente livro é a forma como essa avaliação pode e deve ser feita considerando dados

acerca de cognição. Nesta situação, deparamo-nos com uma alternativa: podemos pensar ou que a questão da racionalidade muda de mãos, da filosofia para a ciência cognitiva, ou que, sendo a racionalidade definitivamente objecto para a ciência cognitiva, ela requer de novo a filosofia. Penso que as questões enumeradas na Introdução, e que estruturaram o desenvolvimento do livro, apontam razões que podem ser evocadas a favor da diferença entre estudos científicos da racionalidade e uma teoria filosófica da racionalidade. A intuição básica é a seguinte: uma coisa é apontar princípios como cânones de correcção para pensamentos e acções e evocar esses princípios como modelo para investigações empíricas, outra coisa é articular os critérios de correcção pelos quais determinados princípios são apontados e colocar questões quanto à justificação destes, bem como quanto à possibilidade, origem e funções dos próprios princípios. Fazer isso conduz-nos a uma investigação sobre a natureza da justificação, investigação essa que se ramifica em problemas que vão desde a análise do próprio conceito de justificação e das condições da sua aplicação, até investigações acerca do estatuto das atribuições de mentalidade supondo a racionalidade e acerca da estrutura da motivação para a acção (motivação para a acção que em nós, humanos, pode dizer respeito a acções que vão desde a satisfação de uma necessidade como a fome até à procura da verdade, do bem, da beleza – ou do nada). Em suma, se os estudos da racionalidade em ciência cognitiva visam questões específicas relativas nomeadamente a emoções, raciocínio e decisão, que são certamente essenciais para o tratamento empiricamente informado da questão, eles não constituem por si só uma posição perante problemas gerais da racionalidade que envolvem a relação pensamento/realidade, a relação pensamento / linguagem, relação entre crenças e desejos, a diferenças entre mente e pensamento¹, etc., alguns dos quais

¹ Não me parece correcto nem metodologicamente pertinente identificar 'mente' com 'pensamento'. A distinção é importante para o tratamento da questão da racionalidade. Os problemas da mente dizem respeito a mentes

procurei articular na Introdução sob o título de 'questões para uma teoria filosófica da racionalidade'.

Gostaria neste momento de recapitular, a partir de Nozick, a formulação do problema filosófico da racionalidade. Temos um problema filosófico da racionalidade (ao qual uma teoria filosófica da racionalidade, por contraste com estudos empíricos, procurará responder) antes de mais porque não apenas usamos princípios para pensar e agir mas temos que decidir que princípios devemos usar para pensar e agir, de uma forma que devemos poder justificar (é esse o sentido da citação de L. Carroll com que usei como epígrafe na Introdução). Para uma tal justificação será necessário considerar as questões acima mencionadas (relação pensamento / realidade, a relação pensamento / linguagem, relação entre crenças e desejos, a diferenças entre mente e pensamento, etc.). A principal intenção deste livro era fazer um mapeamento das questões de tipos diversos a considerar no tratamento do problema da racionalidade e dos pontos na literatura onde elas são abordadas. Esse mapeamento deverá ser o ponto de partida para iniciar investigações mais específicas, quer no âmbito da filosofia, quer no âmbito da ciência cognitiva. Como procederiam tais investigações?

Antes de mais, começando por descrever princípios em funcionamento no pensamento e na acção, em áreas específicas. Depois, perguntando pelas funções desses princípios e procurando formular explicitamente critérios de correcção. A

(mentes humanas, animais, e talvez outras), i. e., à possibilidade de representar, à consciência, à acção intencional, à identidade pessoal, à relação entre essas características e o mundo físico. Basicamente, os problemas da mente dizem respeito àquilo a que chamei cognição, descrita 'a partir do interior'. O pensamento envolve antes de mais toda a dimensão linguística, normativa, que constitui razão para defender que apenas existe pensamento humano, e não animal ou em máquinas, e que é esse pensamento que subjaz a produções teóricas, desde científicas a artísticas. É certo que o pensamento é presumivelmente dependente de mentes, as mentes humanas, mas não é 'acerca delas' nem redutível a elas, mas acerca da realidade em geral.

investigação acerca de critérios de correcção dirá respeito, obviamente, à natureza da justificação, e desenvolver-se-á para além da própria evocação de cânones de racionalidade. Recordo que foram aqui analisadas posições quanto à natureza da justificação² que constituem marcos e poderão servir como referência. Foi sublinhada também a relação do próprio posicionamento da investigação acerca da natureza da justificação com concepções de epistemologia que são elas próprias controversas. A partir das posições alternativas quanto à natureza da justificação, nomeadamente a partir da posição pragmatista, tornou-se além disso visível um outro passo, gerado pela seguinte questão: de um ponto de vista pragmatista, as questões da racionalidade teórica ficam subsumidas às questões da racionalidade prática. Resta retirar as consequências de um tal posicionamento. Ora, essa subsunção remete-nos para as 'finalidades' do tipo de ser que somos, e não podemos considerar que essas finalidades sejam simplesmente epistemológicas, muito menos escolhidas por nós. Como notava Nozick, não podemos dizer que os nossos desejos biologicamente dados, que damos por nós tendo, são por definição racionais – nós somos criaturas, começamos com determinados desejos e isso marca um limite da nossa racionalidade. Devemos então tomar esses desejos como dados brutos para uma teoria da racionalidade, considerada instrumentalmente como a mobilização dos melhores meios que conduzam ao fim que é a satisfação dos desejos? A partir de Nozick foi avançada uma resposta possível, mas a questão é obviamente problemática. A proposta de Nozick foi que não temos que ficar com os desejos como dados brutos, como de certa forma quis Hume, mas também não podemos, seguindo Kant, considerar que nos movemos a puros princípios e que apenas isso é a nossa racionalidade – não podemos transcender desse modo os nossos

² Basicamente o equilíbrio reflectido (tendo ficado em aberto saber se este deve ser considerado como 'mera' análise conceptual) e uma alternativa pragmatista.

desejos, deixá-los simplesmente para trás. Esta última ramificação do problema para as questões da motivação para a acção mostra-nos que a questão da racionalidade, para além de uma investigação sobre a natureza da justificação que traz consigo todas as questões filosóficas tradicionais acerca da natureza da mente, do pensamento e da própria realidade, deve ser, em termos pessoais (no âmbito daquela a que chamei a questão pessoal da racionalidade) colocada em termos de reflexividade, de espaço de manobra entre desejos e princípios, e portanto de uma investigação acerca do grau de voluntarismo que o nosso tipo pensamento nos oferece na formação da nossa natureza.

Até aqui falei sobretudo de indivíduos e princípios – mas como ficamos em termos da questão alargada da racionalidade (tomando agora a questão alargada como dizendo respeito sobretudo à passagem da motivação dos indivíduos para a acção à agregação social destes indivíduos)? Ficamos com o seguinte: teorias sociais e políticas não podem simplesmente presumir que os agentes humanos são 'modelizados' pela teoria da escolha racional, em completa neutralidade relativamente a suposições psicológicas e filosóficas quanto à motivação para agir, às relações entre normatividade e desejos, emoções e sentimentos morais. No entanto, considerar os agentes para além de uma modelização abstracta envolve considerar todos os factores relevantes em raciocínios e tomadas de decisão, e dada a complexidade da situação isso significa considerar muitas coisas: desde as condições dos processos cognitivos subpessoais, até desejos, emoções e sentimentos morais, até ao uso de princípios 'para além da racionalidade instrumental'. De qualquer modo, não parece justificável pensar as macro-questões sociais e políticas em termos de decisão racional desligada de conteúdo psicológico e suposições motivacionais (cabe aqui recordar a posição de S. Blackburn acerca da proposta de referência na filosofia política contemporânea, a teoria rawlsiana da justiça: se a concepção de justiça em causa não motiva, ela não moverá indivíduos à acção). Preenchida essa brecha, e dado esse passo,

novos problemas aparecerão: é possível, por exemplo, que o que parecia claro na caracterização abstracta da decisão racional tenda para o razoável e que seja uma racionalidade imperfeita que mantém indivíduos socialmente agregados (da mesma forma que são funcionamentos 'imperfeitos' que asseguram a racionalidade global de agentes individuais); interessará considerar, se as teorias sociais não podem ignorar os funcionamentos mentais internos de agentes, teorias de formação de intenções, revisão de crenças, raciocínio contrafactual e probabilístico, em contextos estratégicos, auto-engano, etc.³. Basicamente, trata-se de fazer repercutir as conclusões relativas à motivação para acção ao nível do indivíduo na teoria social e política.

As questões descritas acima dizem respeito, é claro, a várias áreas da filosofia, desde a epistemologia à filosofia moral e política. O segundo problema para uma teoria filosófica apontado na Introdução, após este primeiro problema relativo a cânones, critérios e natureza da justificação, que afinal entronca com as questões relativas à motivação para a acção, é mais especificamente epistemológico ou de filosofia da ciência cognitiva. Trata-se, recorde-se da questão da relação entre modelos formais e agentes reais. Falar de agentes reais obriga-nos a encarar a relação entre os princípios abstractamente formuláveis e a realidade (física, cognitiva) dos agentes (por exemplo, a relação entre princípios lógicos e arquiteturas e processos cognitivos de agentes concretos, submetidos a constrangimentos e limitações definidas). Os princípios são formulados a partir de dentro e de cima, i. e., a partir da nossa realidade de agentes pensantes conscientes, que se apercebem como tal e como utilizando voluntariamente princípios. O problema que foi identificado mais do que tratado diz respeito à relação desses princípios com a realidade física dos sistemas cognitivos. Se não podemos dizer – e não podemos –, por exemplo, que os nossos cérebros, quando pensamos, são 'movidos' a argumentos, então o que podemos

³ DUPUY & LIVET 1997.

dizer? O que 'é' ao nível físico, neuronal, de circuitos, isso a que provavelmente apenas ao nível mental devemos chamar 'pensamento'? (não faz muito sentido dizer que cérebros pensam, argumentam, ou agem moralmente, ou calculam. Apenas nós, pessoas, fazemos essas coisas). Embora o problema não tenha sido tratado, foi apontado um exemplo, que me parece metodologicamente significativo, do tipo de investigações aqui geradas. Refiro-me a C. Cherniak, que coloca a questão deste modo: precisamos da noção de racionalidade para fazer teoria da cognição, no entanto os sistemas cognitivos a que queremos continuar a chamar agentes racionais são apenas minimamente (e não perfeitamente) racionais (são, por exemplo, 'lógicos suficientemente bons', não lógicos perfeitos). Interessa compreender a forma dessas limitações e retirar as devidas consequências. De que tipo de limitações e de consequências se trataria? Tratar-se-ia de limitações de obrigações epistémicas para os agentes, de limites para a noção de racionalidade ela própria, da necessidade de compreender a relação entre os problemas da teoria do conhecimento e a neuroanatomia e neurofisiologia, nomeadamente investigando modelos de optimização de *wiring* (cablagem) na estrutura cerebral (este último aspecto é de resto o actual objecto de investigação de Cherniak). Quantos aos limites para a ideia de racionalidade ela própria, a situação pode ser vista da seguinte maneira: a necessidade de uma suposição de racionalidade na teoria da mente não será necessariamente acompanhada pela convicção de que algum cânone de racionalidade eventualmente exibirá a verdadeira natureza dessa racionalidade. Como é o caso com *folkconcepts* variados, a 'racionalidade' pode não ter uma verdadeira natureza – assim sendo as nossas intuições acerca de racionalidade (a partir de cima, do nosso auto-apercebimento como pensamentos conscientes) devem ser tomadas cautelosamente. É um dos resultados a esperar da ciência cognitiva que estudos empíricos conduzam a um ajustamento dos *folkconcepts* que usamos para pensar na mente – e é isso precisamente o que está em causa na discussão em torno da interpretação dos resultados da psicologia cognitiva e psicologia evolutiva a partir

da filosofia. Seremos possivelmente levados a pensar que 'racionalidade' não é uma noção muito profunda, identificável com 'valor cognitivo intrínseco', nem uma noção correspondente a algum género natural, mas um conceito relativo a um nível da realidade que envolve sistemas que são agentes, que se comportam adequadamente, do ponto de vista instrumental, relativamente aos seus mundos. A noção não é definitivamente relativa ao funcionamento, por exemplo, neurofisiológico do interior de agentes: o seu estatuto é o do 'artificial' de Herbert Simon, o interface, e o seu reconhecimento pressupõe possivelmente um ponto de vista psicológico, portanto, um (outro) agente. A propósito desta 'superficialidade', é interessante considerar uma observação feita por Nozick, a propósito do estatuto das leis lógicas, a partir de um ponto de vista quineano (i. e., a partir da ideia de continuidade entre lógica e ciência empírica, uma vez criticada a distinção analítico / sintético): Nozick nota que os físicos continuam a procurar sempre mais explicações para as leis mais profundas conhecidas num dado momento e pergunta por que razão algo de idêntico não se passa relativamente às leis da lógica. Uma vez que Nozick considera improvável que as leis lógicas mais fundamentais, ao contrário das leis físicas, já tenham sido descobertas, o que lhe parece ser o caso é que os princípios que parecem auto-evidentes para nós, para o tipo de agentes que somos, têm esse estatuto por razões de dotação evolutiva e não por serem *insights* sobre a essência da realidade⁴.

A questão da evidência dos princípios do nosso pensamento para nós próprios deve, de qualquer modo, ser abertamente tratada entre ciência cognitiva e filosofia. Esse é basicamente o que chamei o problema da suposta refutabilidade empírica da racionalidade. Saber se podemos ou não falar de refutabilidade empírica – podemos falar aqui do 'problema de Jonathan Cohen' – passa pela consideração do estatuto de *folkconcepts* e intuições pela investigação – não apenas filosófica mas de

⁴ NOZICK 1993:110.

ciência cognitiva acerca de juízo e decisão e acerca da atribuição de mente a outros sistemas cognitivos no ambiente. Em causa está o estatuto do mental na realidade, e a pertinência que tem um conceito global como racionalidade para falar de mente – de qualquer forma, todos os conceitos mentais têm a sua natureza em causa na investigação em ciência cognitiva e a abertura a investigações empíricas é a razão maior para um certo interesse novo pelo problema da mente. Estamos perante qualquer coisa como um auto-conhecimento muito pouco interno, muito pouco intuitivo. Se, como se repete desde Aristóteles, o homem é o animal racional, querendo com isso dizer que a racionalidade é o que nos define como espécie por entre as espécies e assim algo de nuclear na nossa auto-imagem, o que constatamos é que a componente 'animal' da definição nos liga mais à natureza e a evolução do que Aristóteles poderia ter pensado. De qualquer forma, mesmo se os conceitos com que nós, seres mentais, comumente falamos da mente, conceitos tais como representação e consciência, emoções, motivação, desejos, crenças, e também racionalidade, não podem ter a sua natureza definida de forma apriorista antes de a 'realidade se pronunciar', mesmo se admitimos que não é pelo facto de sermos mentais que sabemos o que é mental, que não é pelo facto de sermos suficientemente racionais que sabemos – no sentido de termos uma teoria científica da racionalidade, se é que tal teoria poderia existir – o que é racionalidade, continuamos a querer saber da nossa racionalidade, da forma como crenças, desejos e princípios vão, porque se trata de nós mesmos, de sabermos o que somos e como somos.

Recapitulei alguns dos problemas a que abordagem da racionalidade, considerada como caso prático para o problema da naturalização da epistemologia, nos trouxe. Após uma panorâmica da literatura, o princípio metodológico que declarei no início mantém-se pertinente: quando se trata de questões mentais, nem a limitação ao apriorismo nem a limitação a investigações empíricas se justifica, a forma mais razoável de abordagem

cruza as aportações das investigações empíricas com a análise conceptual, normativa e valorativa da filosofia.

Relativamente ao intuito de comparar abordagens da racionalidade na filosofia analítica e na ciência cognitiva com abordagens levadas a cabo no seio de outras tradições filosóficas, nomeadamente na tradição fenomenológica e hermenêutica, intuíto esse que constituiu uma motivação de fundo para o presente livro, quero apenas fazer notar o seguinte: os temas aqui tratados centraram-se no indivíduo e nos processos de raciocínio, decisão, emoções e sentimentos que neste se passam e que explicam e dão sentido à sua acção. Parece-me que sem um tal ponto de apoio não é possível passar a teorias mais ambiciosas ou a qualquer tipo de crítica da racionalidade.

Bibliografia

ANSCOMBE, G. E. M, 1957, *Intention*, Oxford, Blackwell (ed. utilizada: Cambridge MA, Harvard University Press, 2000).

ARISTÓTELES, 1994, *Étique à Nicomaque*, (dir. J. Tricot), Paris, Vrin, 1994.

ARMSTRONG, David, 1973, *Belief, Truth and Knowledge*, Cambridge, Cambridge University Press.

BAIER, Kurt, Egoism, in Peter Singer (ed), *A Companion to Ethics*, Oxford, Blackwell.

BARKOW, J., 1992, Beneath new culture is old psychology, in Barkow, J., Cosmides, L. & Tooby, J. 1992, *The Adapted Mind: Evolutionary Psychology and the generation of culture*, Oxford, Oxford University Press, 627-637.

BARON, J, 1988, *Thinking and Deciding*, Cambridge, Cambridge University Press.

BARON-COHEN, Simon, 1995, *Mindblindness*, Cambridge MA, MIT Press.

BARWISE, Jon , 1999, «Logic» in WILSON, Robert & KEIL, Frank, 1999, *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge MA, MIT Press, 482-484.

BLACKBURN, Simon, 1998, *Ruling Passions – A Theory of Practical Reasoning*, Oxford, Oxford University Press.

BRATMAN, 1994, "Intention", in GUTTENPLAN, Samuel, *A Companion to The Philosophy of Mind*, Oxford, Blackwell.

BROTHERS, Leslie, 1999, «Emotion and the Human Brain», in WILSON, Robert & KEIL, Frank, 1999, *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge MA, MIT Press.

BUTLER, Joseph, 1953 (1726) *Fifteen Sermons Preached at the Rolls Chapel*, London, Bell & Sons.

CAREY, S. & SPELKE, E., 1994, Domain-specific knowledge and conceptual change, in Hirschfeld, L. & Gelman, S., 1994, *Mapping the Mind*, Cambridge, Cambridge University Press.

CARROLL, Lewis, 1979 (1895), What the Tortoise Said to Achilles, in HOFSTADTER, Douglas, 1979, *Gödel, Escher, Bach*, London, Penguin, pp.43-45.

CASSCELLS, W., SCHOENBERGER, A., GRAYBOYS, T., 1978, Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine*, 299, 999-1000.

CHERNIAK, Christopher, 1986, *Minimal Rationality*, Cambridge MA, MIT Press.

CHERNIAK, Christopher, "Rationality", in GUTTENPLAN, S. (ed.) 1994, *A Companion to The Philosophy of Mind*, Oxford, Blackwell.

CHERNIAK, Christopher, "Rational Agency", in WILSON, Robert & KEIL, Frank, 1999, *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge MA, MIT Press.

COHEN, J.L, 1981, Can Human Irrationality Be Experimentally Demonstrated?, *Behavioral and Brain Sciences* 4, 317-370.

COSMIDES, L. 1989, The logic of social exchange: has natural selection shaped how humans reason? – studies with Wason selection task, *Cognition*, 31, 187-276.

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, 163-228.

COSMIDES, L. & TOOBY, J., 1996, Are Humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty, *Cognition*, 58, 1 -73.

DAMÁSIO, António, 1994, *O Erro de Descartes – Emoção, razão e cérebro humano*, Lisboa, Europa América.

DAMÁSIO, António, 1999, *O Sentimento de Si – O corpo, a emoção e a neurobiologia da consciência*, Lisboa, Europa América.

DAMÁSIO, António, 2003, *Ao Encontro de Espinosa – As emoções sociais e a neurologia do sentir*, Lisboa, Europa América.

DANCY, Jonathan, *Practical Reality*, Oxford University Press, 2002.

DAVIDSON, Donald, 1980, *Essays on Actions and Events*, Oxford, Oxford University Press.

DAVIDSON, Donald, 1963, Actions, Reasons and Causes, in DAVIDSON 1980.

DAVIDSON, Donald, 1970, How is Weakness of the Will Possible, in DAVIDSON 1980.

DAVIDSON, Donald, 1971, Agency, in DAVIDSON 1980.

DAVIDSON, Donald, 1973, Radical interpretation, in DAVIDSON, Donald 1984, *Inquiries into Truth and Interpretation*, Oxford University Press (tradução portuguesa em M. M. Carrilho (dir), *Dicionário do Pensamento Contemporâneo*, Lisboa, Dom Quixote, 1991, com o título 'Interpretação').

DAVIDSON, Donald, 1978, Intending, in DAVIDSON 1980.

DAVIDSON, Donald, 1970, Mental Events, in DAVIDSON 1980.

DAVIDSON, Donald, 1976, Hempel on Explaining Action, in DAVIDSON 1980.

DAVIDSON, Donald, 1982, Rational Animals, *Dialectica*, 36, n.º 4.

DAVIDSON, Donald, 2004, *Problems of Rationality*, Oxford, Oxford University Press.

DAWES, Robyn, 1988, *Rational Choice in an Uncertain World*, New York, Harcourt Brace.

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

DOYLE, Jon, 1999, "Rational Decision Making", in WILSON, Robert & KEIL, Frank, 1999, *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge MA, MIT Press.

DOYLE, Jon, 1999, "Bounded Rationality, in WILSON, Robert & KEIL, Frank, 1999, *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge MA, MIT Press.

DRETSKE, Fred, 1969, *Seeing and Knowing*, Chicago, University of Chicago Press.

DRETSKE, Fred, 2000, *Perception, Knowledge and Belief*, Cambridge, Cambridge University Press.

DUPUY, Jean-Pierre & LIVET, Pierre, 1997, *Les limites de la rationalité – rationalité, éthique et cognition*, Paris, La Découverte.

DUPUY, Jean-Pierre (ed.), 1998, *Self-Deception and the Paradoxes of Rationality*, Stanford, CSLI Publications.

DUPUY, Jean-Pierre, 1999, «Rational Choice Theory», in WILSON, Robert & KEIL, Frank, 1999, *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge MA, MIT Press.

ELSTER, Jon, 1979, *Ulysses and the Sirens*, Cambridge, Cambridge University Press.

EVANS, J. St. B. T, NEWSTEAD, S. & BYRNE, R. M. J. 1993, *Human reasoning: The psychology of deduction*. Hove, UK: Lawrence Erlbaum Associates.

FISCHOFF, Baruch, 1999, «Judgment heuristics» in WILSON, Robert & KEIL, Frank, 1999, *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge MA, MIT Press.

GIBBARD, Allan, 1990, *Wise Choices, Apt Feelings – A theory of normative judgment*, Oxford, Clarendon Press.

GIGERENZER, G., 1991, How to make cognitive ilusions disappear: beyond 'heuristics and biases', *European Review of Social Psychology*, 2, 83-115.

GIGERENZER, G. & HUG, K, 1992, Domain-specific reasoning: social contracts, cheating and perspective change, *Cognition* 43, 127-171.

GIGERENZER, G. 1994, Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa), in G. Wright & P. Ayton (eds), *Subjective Probability*, New York, John Wiley.

GIGERENZER, G., TODD P. & ABC Research Group, 2000, *Simple Heuristics that make us smart*, Oxford, Oxford University Press.

GIL, Fernando, 1986, *Provas*, Lisboa, INCM.

GIL, Fernando, 1993, *Traité de l'Évidence*, Paris, Jérôme Millon.

GIL, Fernando, 2000, *La Conviction*, Paris, Flammarion.

GOLDMAN, Alvin, *What is Justified Belief?* 1979, in Pappas, G. (ed), *Justification and Knowledge*, Dordrecht, Reidel.

GOLDMAN, Alvin, 1986, *Epistemology and Cognition*, Cambridge MA, Harvard University Press.

GOLDMAN, Alvin, 1999, «Justification», in WILSON, Robert & KEIL, Frank, 1999, *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge MA, MIT Press.

GOLDMAN, A., 1992a, Reliabilism, in Dancy, J & Sosa, E., *A Companion to Epistemology*, Oxford, Blackwell, 1982.

GOLDMAN, A., 1992b, *Liaisons: Philosophy Meets the Cognitive and Social Sciences*, Cambridge MA, MIT Press.

GOLDMAN, A., 1993, *Philosophical Applications of Cognitive Science*, Boulder, Westview.

GOLDMAN, A., 1994, Action (2), in GUTTENPLAN, Samuel, *A Companion to The Philosophy of Mind*, Oxford, Blackwell.

GOLDMAN, A., 1999, *Knowledge in a Social World*, Oxford, Oxford University Press.

GOLDMAN, A., 2002, *Pathways to Knowledge, Private and Public*, Oxford, Oxford University Press.

GOLDMAN, A., 2003, Epistemic Folkways and Scientific Epistemology, in P.K. Moser & A. Vander Nat, *Human Knowledge*, Oxford, Oxford University Press, 2003.

GOPNIK, Alison, 1999, «Theory of Mind», in WILSON, Robert & KEIL, Frank, 1999, *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge MA, MIT Press.

HAACK, Susan, 1993, *Evidence and Inquiry – Towards Reconstruction in Epistemology*, Oxford, Blackwell.

HACKING, Ian, 1975, *The Emergence of Probability*, Cambridge, AT the University Press.

HARMAN, Gilbert, 1999, *Reasoning, Meaning and Mind*, Oxford, Oxford University Press.

HARMAN, Gilbert, 1995, «Rationality», in SMITH, E. & OSHERSON, D., *An Invitation to Cognitive Science*, vol. III, Cambridge MA, MIT Press (também em HARMAN 1999).

HINTIKKA, J., 1962, *Knowledge and Belief*, Ithaca, NY, Cornell University Press.

HUME, David, 2001, *Tratado da Natureza Humana*, Lisboa, Fundação Calouste Gulbenkian, tradução de Serafim da Silva Fontes, prefácio e revisão técnica da tradução de João Paulo Monteiro.

HUME, David, 1975, *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, ed. L. A. Selby-Bigge, Oxford, Oxford University Press.

JEFFREY, Richard, 1983, *The Logic of Decision*, Chicago, Chicago University Press (1ª edição 1965).

KAHNEMAN, D., SLOVIC, P. & TVERSKY, A. (eds.), 1982, *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge, Cambridge University Press.

KAHNEMAN, D., & TVERSKY, A. 1972, Subjective probability: a judgment of representativeness. *Cognitive Psychology* 3, 430-454. (também em KAHNEMAN, D., SLOVIC, P. & TVERSKY, A. (eds.), 1982).

KAHNEMAN, D., & TVERSKY, A. 1979, Prospect theory: an analysis of decision under risk, *Econometrica* 47, 263-291.

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

KOORSGAARD, Christine, 1996, *Sources of Normativity*, Cambridge, Cambridge University Press.

LACEY, Alan, 2001, *Robert Nozick*, Princeton University Press, Princeton New Jersey.

LENNON, 1994, "Reasons and causes", GUTTENPLAN, Samuel, *A Companion to The Philosophy of Mind*, Oxford, Blackwell.

LESLIE, Alan, 1994, ToMM, ToBY and Agency: core architecture and domain specificity, in Hirschfeld, L. & Gelman, S., 1994, *Mapping the Mind*, Cambridge, Cambridge University Press.

LICHTENSTEIN, S., FISCHOFF, B. & PHILLIPS, L., 1982, Calibration of probabilities: The state of the art to 1980, in KAHNEMAN, D., SLOVIC, P., & TVERSKY, A. 1982.

LUCE, R.D. & RAIFFA, H., 1957, *Games and Decisions*, New York, Wiley.

McDOWELL, John, 1981, Non-Cognitivism and Rule-Following, in Holtzman S. & Leich, S., *Wittgenstein: to follow a rule*, London, Routledge, pp.141-162.

MIGUENS, Sofia 2002, *Uma Teoria Fiscalista do Conteúdo e da Consciência – D. Dennett e os debates da filosofia da mente*, Porto, Campo das Letras.

NAGEL, Thomas, 1970, *The Possibility of Altruism*, Princeton, Princeton University Press.

NICHOLS, S. & STICH, S., 2003, *Mindreading – An integrated account of Pretence, Self-Awareness, and Understanding Other Minds*, Oxford, Oxford University Press.

NISBETT, R. & ROSS, L., 1980, *Human Inference: Strategies and Shortcomings of Social Judgment*, Eaglewood Cliff, NJ, Prentice Hall.

NOZICK, Robert, 1969, Newcomb's Problem and Two Principles of Choice, in *Essays in Honor of C. G.Hempel*, N. Rescher et al. (ed.), Dordrecht, Reidl, pp. 114-146.

NOZICK, R., 1981, *Philosophical Explanations*, Cambridge MA, The belknap Press of Harvard University Press.

NOZICK, Robert, 1993, *The Nature of Rationality*, Princeton, New Jersey, Princeton University Press.

OATLEY, Keith, 1999, «Emotions», in WILSON, Robert & KEIL, Frank, 1999, *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge MA, MIT Press.

OSHERSON, Daniel, 1995, Probability Judgment, in *An Invitation to Cognitive Science*, vol. 3 Thinking, Cambridge MA. MIT Press.

PIATELLI-PALMARINI, M., 1994, *Inevitable Illusions: How Mistakes of Reason Rule our Minds*, New York, John Wiley and Sons.

PINKER, Steven, 1994, *The Language Instinct*, London, Penguin.

PRIEST, Graham, 2000, *Lógica – uma introdução concisa*, Lisboa, Temas e Debates.

PUTNAM, Hilary, 1992, *Razão, Verdade, História*, Lisboa, Dom Quixote.

QUINE, W.V, 1960, *Word and Object*, Cambridge, MA, MIT Press.

QUINE, W. V., 1969, Epistemology Naturalized, in *Ontological Relativity and Other Essays*.

RAMSEY, Frank, 1926, Truth and Probability, in H. Mellor ed, *F. R. Ramsey Philosophical Papers*, 1990, Cambridge, Cambridge University Press.

RIPS, Lance 1995, Deduction and Cognition, in *An Invitation to Cognitive Science*, vol. 3 Thinking, Cambridge MA. MIT Press.

RORTY, Richard, 1979, *Philosophy and the Mirror of Nature*, Princeton, New Jersey, Princeton University Press.

SAMUELS, R., STICH, S & TREMOULET, P., 2003, Repensando a Racionalidade: de implicações pessimistas a módulos darwinianos, *Intelectu* n.º 9 www.intelectu.com (Rethinking Rationality: from bleak implications to darwinian modules, in Lepore, E. & Pylyshyn, Z. eds, *Invitation to Cognitive Science*, 2003, tradução portuguesa de Tomás Carneiro).

- SARDO, Francisco, 2000, *Logos e Racionalidade – na gênese e estrutura da lógica de Aristóteles*, Lisboa, INCM.
- SAVAGE, Leonard, 1954, *Foundations of Statistics*, New York, Wiley.
- SCHICK, Frederic, 1997, *Making Choices – A recasting of Decision Theory*, Cambridge, Cambridge University Press.
- SCHIZGAL, Peter, 1999, «Motivation», in WILSON, Robert & KEIL, Frank, 1999, *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge MA, MIT Press.
- SCHMIDTZ, David (ed.), 2002, *Robert Nozick*, Cambridge University Press, Cambridge.
- SEARLE, John, *Rationality in Action*, MIT Press, 2001.
- SIMON, Herbert, 1998 (1969) *The Sciences of the Artificial*, Cambridge MA MIT Press.
- SIMON, H., 1982, *Models of Bounded Rationality*, Cambridge MA, MIT Press.
- SHAFIR, Eldar, 1999, «Decision Making» in WILSON, Robert & KEIL, Frank, 1999, *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge MA, MIT Press.
- SHAFIR, E. & TVERSKY, A. 1995, Decision Making, in Smith, E & Osherson, D., Eds *An Invitation to Cognitive Science*, vol. 3 Thinking, Cambridge MA, MIT Press.
- SLOVIC, P., FISCHHOF, B. & LICHTENSTEIN, S., 1976, Cognitive processes and societal risk taking, in CAROL, J.S. & PAYNE, J. W. (eds), *Cognition and Social Behavior*, Hillsdale, NJ, Erlbaum.
- SMITH, Adam, 1976, *The Theory of Moral Sentiments*, ed. D. D. Raphael e A. L. Macfie, Oxford, Oxford University Press.
- SOBER, E., 1981 «The Evolution of Rationality.» *Synthese* 46, 95-120.
- SOBER, Elliot, 1994a, Did evolution make us psychological egoists? in *From a Biological Point of View – Essays in Evolutionary Philosophy*, Cambridge, Cambridge University Press.
- SOBER, Elliot, 1994b, Why not solipsism?, In *From a Biological Point of View – Essays in Evolutionary Philosophy*, Cambridge, Cambridge University Press.
- SOBER, Elliot, 1994c, The adaptive advantage of learning and a priori prejudice, in *From a Biological Point of View – Essays in Evolutionary Philosophy*, Cambridge, Cambridge University Press.
- SOBER, Elliot, 1994d, The primacy of truth telling and the evolution of lying, in *From a Biological Point of View – Essays in Evolutionary Philosophy*, Cambridge, Cambridge University Press.

SOBER, Elliot, 1994e, Prospects for an evolutionary ethics in *From a Biological Point of View – Essays in Evolutionary Philosophy*, Cambridge, Cambridge University Press.

De SOUSA, Ronald, 1987, *The Rationality of Emotion*, Cambridge MA, MIT Press.

STALNACKER, Robert, 1999, «Logical Omniscience», in WILSON, Robert & KEIL, Frank, 1999, *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge MA, MIT Press.

STICH, Stephen, 1990, *The Fragmentation of Reason – Preface to a Pragmatic Theory of Cognitive Evolution*, Cambridge MA, MIT Press.

STICH, S., 1988, Reflective equilibrium, analytic epistemology and cognitive diversity, *Synthese* 74, 391-413.

STICH, Stephen, 1996, *Deconstructing the Mind*, Oxford, Oxford University Press.

SUTHERLAND, Stuart, *Irrationality*, 1992, New Brunswick, New Jersey, Rutgers University Press.

TOOBY, J. & COSMIDES, L. 1995, Foreword, in BARON-COHEN 1995, *Mindblindness – An Essay on Autism and Theory of Mind*, Cambridge, Cambridge University Press, xi-xviii.

TUNHAS, Paulo, 2000, Modos de Pensar (I), *Análise* 21.

TUNHAS, Paulo, no prelo, Deliberação, in *Dicionário de Filosofia Moral*, A. Marques & D. Pires Aurélio (dir).

TVERSKY, A. & KAHNEMAN D. 1971, Belief in the “law of small numbers”. *Psychological Bulletin* 76: 105-110 (também em KAHNEMAN, D., SLOVIC, P. & TVERSKY, A. (eds.), 1982).

TVERSKY, A. & KAHNEMAN D., 1981, The Framing of Decisions and the Psychology of Choice, *Science*, 211.

TVERSKY, A. & KAHNEMAN D., 1982a, “Judgment under Uncertainty: heuristics and biases”, in KAHNEMAN, D., SLOVIC, P. & TVERSKY, A. (eds.), 1982.

TVERSKY, A & KAHANEMAN, D., 1982b, *Judgments of and by representativeness*, in KAHNEMAN, D., SLOVIC, P. & TVERSKY, A. (eds.), 1982.

TVERSKY, A. & KAHNEMAN D. 1993, “Probabilistic Reasoning”, in GOLDMAN, A. (ed.), *Readings in Philosophy and Cognitive Science*, Cambridge MA, MIT Press.

VON NEUMANN, John & MORGENSTERN, Oskar, 1947 (2nd ed), *Theory of Games and Economic Behavior*, Princeton, Princeton University Press.

WILLIAMS, Bernard, 1980, *Internal and External Reasons*, in R. Harrison (ed.), *Rational Action*, Cambridge, Cambridge University Press.

WIMMER, H. & PERNER, J., 1988, Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception, *Cognition* 13, 103-128.

Índice de autores

- Anscombe, G. E. M., 7, 45, 93, 94, 99, 100, 101, 102
- Aristóteles, 7, 35, 43, 45, 47, 53, 54, 55, 56, 57, 60, 61, 76, 94, 107, 189, 197, 199
- Blackburn, S. 8, 13, 15, 37, 44, 45, 107, 114, 115, 116, 117, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 193, 199
- Butler, J. 147, 148, 154, 157, 158, 159, 199
- Cherniak, C. 22, 23, 126, 195, 200
- Cohen, J. 28, 196, 200
- Cosmides, L. 66, 80, 85, 86, 87, 88, 199, 200
- Damásio, A. 35, 111, 200
- Darwin, C. 35
- Davidson, D. 7, 29, 45, 89, 93, 94, 95, 96, 97, 98, 99, 101, 102, 103, 189, 200, 201
- Dennett, D. 29, 44, 89, 201, 204
- Descartes, R. 35, 40, 111, 133, 200
- Espinosa, B. 35, 111, 200
- Fodor, J. 81, 82, 83
- Freud, S. 35, 39, 40, 42
- Gigerenzer, G. 86, 88, 202
- Gil, F. 38, 42, 202
- Goldman, A. 8, 13, 21, 44, 45, 119, 121, 122, 123, 124, 125, 127, 128, 129, 130, 131, 132, 133, 134, 136, 137, 140, 141, 143, 144, 145, 146, 202, 203
- Hume, D. 7, 15, 36, 37, 45, 47, 104, 106, 107, 108, 109, 111, 113, 114, 115, 116, 117, 148, 153, 154, 156, 187, 188, 189, 192, 203, 204
- Jeffrey, R. 57, 58, 59, 61, 204
- Kahneman, D. 25, 26, 66, 67, 68, 69, 70, 72, 78, 79, 86, 135, 204, 208
- Kant, I. 7, 36, 37, 45, 50, 106, 107, 111, 112, 113, 114, 161, 164, 187, 189, 192, 204
- Korsgaard, C. 151
- McDowell, J. 150
- Nisbett, R. 72, 135, 144, 205
- Nozick, R. 5, 8, 41, 44, 45, 50, 64, 147, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181,

- 182, 183, 184, 185, 186, 187, 188, 191, 192, 196, 205
- Pinker, S. 80
- Quine, W. V. 29, 89, 121, 123, 205
- Simon, H. 25, 68, 196, 206
- Sober, E. 54, 207
- Stich, S. 8, 13, 44, 45, 80, 81, 82, 83, 86, 88, 90, 119, 121, 122, 133, 134, 135, 136, 137, 138,
- 139, 140, 141, 142, 145, 146, 167, 206, 207
- Tooby, J. 66, 80, 85, 86, 87, 88, 199, 200
- Tversky, A. 25, 26, 66, 67, 68, 69, 70, 72, 78, 79, 86, 135, 204, 208
- Weber, M. 39, 40
- Williams, B. 104, 105, 208

Índice temático

- Acção, 33-37, 93-99
- Autismo 30
- Auto-engano 52
- Akrasia* 52
- Cânones de racionalidade 19
- Coerentismo 51
- Cognição 38
- Cognitivismo 150
- Competência 77
- Construtivismo 151
- Crítério de correcção 21, 128
- Decisão 56
- Deliberação 56
- Desejos, segundo Blackburn 161-162
- Desejos, segundo Nozick 185-188
- Dilema do prisioneiro, 64, 166, 176, 178, 179
- Educação sentimental 155, 162
- Egoísmo psicológico/Egoísmo ético 54, 153-154
- Emoções 34
- Epistémica 124
- Epistémica primária, 124
- Epistémica secundária 124
- Epistemologia analítica 133
- Equilíbrio reflectido 129-130
- Expressivismo 148
- Fiabilismo 51
- Framing* 27, 67
- Fundacionalismo 51
- Heurísticas 26, 67
- Intenção 100-101
- Irrracionalidade 52
- Justificar 21
- Lógica 17, 19
- Maximização da utilidade esperada 20
- Modelos de racionalidade 27
- Modularidade 79-84
- Motivação 34, 149
- Nilismo epistémico 140
- Pensamento 38
- Performance* 77
- Princípios 166-168, 169-179
- Problema de Newcomb 166, 176
- Psicologia evolutiva 79-80

Raciocínio e verdade	140	Racionalizar uma acção, segundo Davidson	95
Racionalidade, definição instrumental	48	Razões para a acção	105
Racionalidade, interpretação normativa	19	Regras-J	127, 128, 132
Racionalidade, interpretação descritiva	19	Relativismo	140, 153, 163-164
Racionalidade limitada	25-26	<i>Satisficing</i>	26
Racionalidade mínima	24	Sentimentalismo	110, 148
Racionalidade prática	11, 91	Sentimentos morais	115, 154-155
Racionalidade teórica	11, 119	Tarefa das crenças falsas	30
Racionalização, sentido weberiano	39-40	Teoria da decisão	20, 57-65
Racionalização, sentido freudiano	39-40	Teoria da mente	30
		Teoria das probabilidades	19
		Utilidade	58
		Utilidade simbólica	174

Índice

Prólogo	9
Origens	15
<i>Introdução – Estudos da racionalidade e a ideia de uma teoria filosófica da racionalidade.</i>	
<i>Questões para uma teoria filosófica da racionalidade</i>	17
1. Cânones de racionalidade e critérios de correcção	19
2. Teorias normativas e agentes reais	22
3. Modelos de racionalidade e sua suposta refutabilidade empírica	25
4. Suposição de racionalidade na teoria da mente	29
5. Estatuto do mental na realidade e estatuto dos <i>folkconcepts</i>	31
6. Desejos e estrutura motivacional dos agentes. Acção e racionalidade na acção	33
7. A questão alargada da racionalidade	37
8. A questão fenomenológica da racionalidade	41
9. A questão pessoal da racionalidade.	43
<i>Capítulo 1 – O que é ser racional?</i>	
<i>Investigações sobre racionalidade: história e organização da literatura</i>	47
Racionalidade: racionalidade instrumental, racionalidade teórica, racionalidade prática	47

Aristóteles, a racionalidade teórica e a racionalidade prática	53
Teoria da decisão racional	57
Poderá a racionalidade ser empiricamente refutada?	
Psicologia cognitiva: juízo e tomada de decisão	65
Como interpretar os resultados da psicologia cognitiva?	76
Psicologia evolutiva e racionalidade – resultados e interpretações	79
 <i>Capítulo 2 – Filosofia e racionalidade prática: o que devemos fazer?</i>	91
Acção, modelo crença-desejo, intenção na acção, razões e causas. E. Anscombe e D. Davidson	93
Racionalidade na acção: a teoria instrumental	104
Filosofia moral: D. Hume <i>versus</i> I. Kant	106
Hume: motivação para a acção	108
Kant: motivação para a acção	111
Hume, Kant e a deliberação	113
Da filosofia moral à filosofia política. Da motivação individual para a acção a concepções de sociabilidade e coordenação das acções de indivíduos – a tradição sentimentalista (D. Hume e A. Smith)	114
 <i>Capítulo 3 – Filosofia e racionalidade teórica: em que devemos acreditar?</i>	119
Quem deve estudar a racionalidade, o cientista cognitivo ou o filósofo?	
A. Goldman e S. Stich acerca de questões meta-epistemológicas	119
A. Goldman – a natureza da justificação	122
S. Stich – contra a 'epistemologia analítica': justificação e pragmatismo	133

 <i>Capítulo 4 – Duas teorias filosóficas da racionalidade</i>	147
Simon Blackburn – a mente e a natureza da racionalidade prática	147
Motivação para a acção: <i>sentimentalismo versus racionalismo</i> . Natureza das pretensões normativas éticas: expressivismo.	148
Os desejos, o eu e os outros: sentimentos morais e crítica ao egoísmo psicológico	153
Importância e problemas da proposta de Blackburn	163
Robert Nozick – uma teoria da racionalidade simbólica	165
Princípios	166
Da teoria da decisão à ética: princípios e significado simbólico	169
Regras práticas para a racionalidade teórica. A verdade como fim cognitivo	179
Limites da racionalidade instrumental: natureza e valor da racionalidade. Desejos e princípios	185
 <i>Conclusão</i>	189
 <i>Bibliografia</i>	199
 <i>Índice de autores</i>	209
 <i>Índice temático</i>	211