

# Entrevista a António Zilhão

13/06/2008

---

**P: O que é a Filosofia da Acção?**

R: A Filosofia da Acção é aquela parte da filosofia que tem como tópico o estudo da acção humana. Há, evidentemente, um conjunto de problemas que a constituem. Estes definem, por sua vez, sub-áreas de estudo dentro da Filosofia da Acção. A título de exemplo, posso indicar os seguintes problemas e sub-áreas. Em primeiro lugar, o problema da identificação das acções. Refiro-me aqui ao problema de saber como se caracteriza uma acção – o que é uma acção? Em segundo lugar, o problema da individuação das acções. Refiro-me por meio desta expressão ao problema de determinar como é que se distinguem acções umas das outras. Em terceiro lugar, o problema da dimensão temporal da acção, isto é, o problema de determinar em que período do tempo é que é apropriado considerar-se que uma acção decorreu. Estes três problemas definem uma sub-área da Filosofia da Acção que poderíamos, talvez, designar como Ontologia da Acção. Outra sub-área desta disciplina é a da Metafísica da Acção. Aqui interessa, sobretudo, saber se as acções podem ser livres e, caso possam ser livres, em que é que consiste essa liberdade. Outra sub-área ainda é a da Epistemologia da Acção, a qual se constitui a partir da tentativa de perceber em que é que consiste explicar uma acção. Finalmente, há também problemas associados à compreensão da acção colectiva, ou seja, àquele conjunto de acções cuja execução pressupõe não apenas que o agente aja de uma determinada maneira, mas também que os outros agentes ajam igualmente de uma determinada maneira. Neste sentido, é necessário compreender como é que emerge a coordenação necessária a que a acção do agente possa ter o conteúdo que nós lhe atribuímos, quais são os princípios subjacentes a essa coordenação, etc. Creio que estes são os problemas mais importantes com que a filosofia da acção lida, e que tem que discutir, e que, por conseguinte, aquelas que eu referi são também as sub-áreas nucleares da Filosofia da Acção.

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

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

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

R: Há uma relação muito óbvia, que é a relação que se estabelece em virtude de uma das categorias essenciais subjacentes ao pensamento económico ser a categoria da racionalidade. Uma das coisas que os economistas debatem é precisamente a de saber o que é, em geral, ser racional, e como é que o agente pode agir racionalmente em certas situações específicas. Ora, a caracterização do conceito de racionalidade é também uma questão que é tratada na Filosofia da Acção. Na apresentação que há pouco fiz dos principais problemas que esta deve tratar, não me referi especificamente a este problema

porque ele aparece integrado no âmbito daqueles tópicos que eu designei como sendo os tópicos da identificação das acções e da individuação das acções. O primeiro, como disse, investiga o que é que faz com que um certo evento seja considerado uma acção (ou seja, como é que nós distinguimos acções de não-acções) e o segundo investiga como é que distinguimos umas acções das outras. Ora, se prosseguirmos a discussão em torno destes tópicos, veremos que uma das categorias fundamentais que é chamada à colação para especificar as distinções necessárias ao seu tratamento é a categoria da racionalidade. Neste sentido, a Filosofia da Acção lida com um aspecto fundador da teoria económica, que é precisamente o aspecto de determinar o que é ser racional no âmbito da acção.

**P: Comecemos então pela questão da identificação das acções. O que pode contar como uma acção?**

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

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

R: Todas as definições podem, evidentemente, ser postas em causa. Mas a tradição da Filosofia da Acção é considerar como acções aqueles comportamentos que decorrem de razões. Se remontarmos, por exemplo, a Aristóteles, o que nele encontramos é o estabelecimento de uma analogia com a razão teórica. Aristóteles argumenta reflexivamente que, no âmbito da razão teórica, a admissão de certas proposições parece compelir-nos à concordância com outras proposições. É claro que nem sempre isso acontece – por vezes eu tomo certas proposições como verdadeiras, mas, no entanto, não sinto em mim qualquer necessidade que me leva a pensar que, se eu aceitei aquelas proposições como verdadeiras, tenho que aceitar esta ou estas outras como verdadeiras. Os casos em que eu detecto essa necessidade são precisamente aqueles em que as proposições que eu aceitei como verdadeiras podem ser tomadas como premissas e solicitam que eu efectue uma inferência de carácter peculiar para uma outra proposição que será a conclusão. Estamos aqui perante a descoberta dos princípios da inferência lógica – se eu concordo com a proposição A e com a proposição B, sinto-me obrigado a concordar também com a proposição C, sendo que C decorre, como conclusão lógica, de A e B, tomadas como premissas. Isto é uma coisa que se nos impõe quando reflectimos sobre a nossa própria prática de argumentar e de pensar reflexivamente. Como disse, em Aristóteles, as considerações acerca da razão prática surgem por analogia com as considerações acerca da razão teórica. Há casos nos quais o pensamento parece implicar uma acção e há casos em que isso não acontece. Quando Aristóteles reflecte sobre qual é a diferença específica que acompanha os casos em que há esta conexão da acção com o pensamento que parece ser semelhante à conexão que conclusões têm com premissas, o que ele constata é que essa diferença específica é que os pensamentos em causa constituem *razões* para a acção. Isto é, segundo ele, a partir do momento em que eu tenho aqueles pensamentos, sinto-me compelido a agir de certa maneira, de uma forma não exactamente igual a, mas com parecenças muito relevantes com, aquela por meio da qual me sinto compelido a extrair uma conclusão a partir do momento em que aceito as premissas das quais ela se segue logicamente. A Filosofia da Acção nasce, portanto, a partir desta tradição, na qual se considera que o objecto a ser tratado por esta disciplina são aqueles comportamentos que resultam, de uma forma análoga ao modo como as conclusões resultam das premissas, de pensamentos anteriores que estão com eles numa determinada relação de consequência. Neste sentido, e dado este enquadramento, comportamentos instintivos ou reflexos, que são desempenhados de forma independente de qualquer pensamento, não contam, por definição, como acções.

**P: Para que um movimento de um corpo seja considerado uma acção é necessário que seja consciente, ou não?**

R: Aqui a pergunta não tem tanto que ver com a acção mas com o pensamento. Dada a definição anterior de o que é uma acção, acções têm que estar conectadas de determinado modo com razões, e as razões têm que ser conteúdos de pensamentos. Agora, a questão da presença à consciência ou não presença à consciência já depende do modo como se define pensamento. Se tivermos uma definição cartesiana de pensamento, não pode haver acções que não estejam associadas a razões conscientes porque, deste ponto de vista, qualquer razão tem necessariamente que ser consciente. Mas, evidentemente, podemos ter uma concepção não cartesiana de pensamento. Ora, em contextos teóricos não cartesianos, o facto de um conjunto de itens de vida mental não ser consciente não quer automaticamente dizer que ele não possa constituir

pensamento e, constituindo pensamento, que não possa estar naquela relação particular de consequência com o comportamento desempenhado pelo agente que nos permita classificá-lo como um conjunto de razões. Portanto, dentro de algumas destas concepções não cartesianas, pode ter-se pensamentos não conscientes que são razões. Nestas condições, o comportamento deles decorrente é uma acção.

**P: E a sua opinião qual é?**

R: Eu não sou um cartesiano.

**P: O movimento de um sonâmbulo conta como uma acção?**

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

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

R: Não, não me parece estranho. Há com certeza uma teoria científica acerca do sonambulismo que explica esse comportamento, e essa teoria tem que passar pela compreensão dos processos que decorrem na mente de um agente quando ele está a dormir. Portanto, há uma definição dos casos normais, e depois temos os casos de sonambulismo que são não-normais. Mas, seja como for, penso que os casos de sonambulismo têm que ser compreendidos dentro de uma teoria do sono. E eu creio que as categorias por meio das quais nós podemos formular uma teoria do sono não são categorias como acções, razões, etc. Em princípio as pessoas quando estão a dormir não agem.

**P: Mas continuam a ser agentes?**

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

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

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

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

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

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

R: De acordo com a minha experiência, pelo menos, creio que os casos de auto-engano são, com muita frequência, fictícios. De uma forma geral, creio que o que prevalece é que as pessoas sabem que não estão a agir com base nas razões com base nas quais dizem que agem. O auto-engano tem esta dimensão de ilusão, envolve mentira. Isto é uma coisa. Há outra coisa diferente, que é haver engano, no sentido de erro, do agente acerca das razões com base nas quais ele agiu. Mas isso não é auto-engano. Eu posso estar convicto que agi com base num certo número de razões quando, na realidade, não agi com base nessas razões e agi com base noutras. Mas aqui não estou a enganar-me a mim próprio; estou enganado acerca de mim próprio, o que é diferente. Claro que, neste caso, põe-se a questão de saber como é que se pode descobrir as verdadeiras razões; mas, supondo que isso é possível, pode dar-se o caso de eu estar convicto que agi com

base nas razões x e y quando, na realidade, agi com base nas razões w e z. Como disse, um caso de engano acerca das razões subjacentes à minha acção não é auto-engano porque eu não estou a enganar-me a mim próprio, estou apenas enganado a meu respeito, como quando penso que a dor que tenho nas costas é provocada por um problema na coluna vertebral quando ela nada tem a ver com a coluna vertebral mas com outra condição subjacente (uma pedra num rim, por exemplo).

**P: Então afirma que o auto-engano acontece menos do que parece.**

R: Eu acho que é uma coisa que roça a patologia... Mas não sei; se calhar acontece com frequência mesmo fora dos consultórios de psiquiatria...

**P: E nesse sentido pode ser um problema filosófico?**

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

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

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

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

R: A distinção que eu estabeleci entre intencionalidade e racionalidade parece supor que a caracterização de uma acção como racional subentende não apenas a recondução da acção às razões do agente para agir do modo como agiu mas também a avaliação da

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

**P: E quais são as origens desses padrões?**

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

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

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

**P: Por exemplo, a maximização da utilidade esperada seria uma forma possível?**

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

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

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

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



R: Para avaliar esta sua sugestão temos que ver o que está em jogo com a introdução da distinção entre competência e performance na polémica em torno da cognição. Esta distinção foi introduzida por Chomsky no âmbito da Linguística. Os princípios teóricos subjacentes a essa distinção são basicamente os seguintes. Aquilo que o linguista chomskiano tem que fazer é caracterizar a competência linguística do agente. Ora, a competência linguística do agente é, digamos assim, o software da sua faculdade da linguagem. Acontece que o software da faculdade da linguagem não pode ser completamente dado pela observação do comportamento linguístico do agente porque falar é uma acção física que envolve a mobilização de um conjunto de mecanismos de carácter físico de que o corpo humano dispõe, os quais têm limitações, também elas físicas, que canalizam de uma certa forma o desempenho linguístico dos agentes. Neste sentido, a argumentação de Chomsky é que a compreensão da competência exige não apenas a observação do comportamento real, mas também a compreensão de quais são os princípios computacionais de carácter geral subjacentes por meio dos quais esse comportamento real pode ser implementado. E esses princípios podem ir, e é natural que vão, muito para lá do que o comportamento real exhibe. Daqui segue-se que deverão existir estruturas linguísticas que nós podemos modelar por meio de construções formais, as quais, dado o seu carácter puramente formal, admitem a realização de operações tremendamente complexas que nós não vamos, todavia, encontrar no comportamento linguístico efectivo (iterações sucessivas, por exemplo). No âmbito deste último, nós podemos encontrar algumas exemplificações de baixa ou muito baixa complexidade desse género de construções, mas nada que se compare com aquilo que a posse de um mecanismo computacional de geração dessas operações nos permitiria, em princípio, fazer. Isto é, a competência instalada é algo que é representável por uma função, no sentido de uma função matemática, e a compreensão teórica da competência exige a compreensão abstracta dessa função, mesmo que, na prática, dadas as limitações físicas do aparelho fonador humano, o comportamento real apenas exhiba uma muito pequena parte do que seria possível de realizar dada a presença no aparelho cognitivo humano de um mecanismo que implementa fisicamente a função. Esta é a ideia subjacente à distinção entre competência e desempenho introduzida por Chomsky a partir das suas investigações em Linguística. Trata-se, basicamente, de uma distinção entre competência computacional, a qual tem um carácter formal, abstracto, matemático, se quiser, e desempenho prático, o qual é condicionado por outros factores, de carácter mecânico, que são independentes da competência formal. A questão então é: não se poderá aplicar o mesmo método de análise à acção humana? Quando autores como Simon falam das limitações cognitivas dos seres humanos, não poderemos defender que o que se passa é que existe uma competência instalada que vai muito para além do desempenho efectivo, mas esse desempenho efectivo tem que ser entendido em função da competência computacional instalada? A este respeito há um conjunto de coisas que podem ser ditas: primeiro, mesmo falando apenas em termos de competência instalada, há modelos de tomada de decisão que envolvem uma complexidade computacional de tal modo vasta que se justifica a dúvida sobre a possibilidade de que a competência subjacente possa encontrar-se nalgum órgão biológico, mesmo esquecendo por momentos questões associadas ao desempenho. Mas, ainda que não fosse esse o caso, esta abordagem corre, em segundo lugar, o risco de perder de vista aquilo que é essencial na acção efectiva. Porque não é a mesma coisa ter uma competência complexa capaz de tratar de modos computacionalmente muito sofisticados uma multiplicidade gigantesca de variáveis, a qual tem depois que ser canalizada através de um mecanismo que limita fortemente as suas potencialidades, e ter que agir de forma útil numa situação, na qual, seja qual for a complexidade computacional da competência

subjacente, a informação relevante disponível é necessariamente muito limitada. Na realidade, no âmbito da decisão e da acção, a questão põe-se, com frequência, nestes termos. Outra variável que também adquire, em circunstâncias práticas, uma importância fundamental é a variável temporal. E esta tão-pouco pode ser reduzida apenas ao facto de ter que se “encaixar” em menos tempo aquilo que foi previsto para mais tempo.

**P: Quais são, então, as possíveis formas de considerar uma acção irracional?**

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

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

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

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

R: É possível ter casos de *acrasia* nos quais a acção que o agente implementa é aquela que, numa análise objectiva, é a melhor para o agente e, no entanto, essa acção é acrática porque o agente está convicto de que a acção cujo desempenho seria melhor para ele seria outra, e no entanto ele desempenhou aquela. O que interessa para a consideração de uma acção como acrática é a convicção que o agente tem acerca de qual é a melhor acção a desempenhar. Outra questão, que é diferente, é a questão de saber se essa convicção está correcta ou não está correcta.

**P: E se essa convicção for um engano?**

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

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

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

ele, de facto, acha que é a melhor, então, na realidade, não há *acrasia*, só há simulação de *acrasia*.

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

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

**P: Mas o auto-engano não pressupõe convicção?**

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

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

R: Creio que sim. Como sabe, escrevi um conjunto de artigos que têm, precisamente, como objectivo mostrar como se pode dar conta do fenómeno da *acrasia* no âmbito de uma certa visão da mente humana que recorre ao apelo a heurísticas rápidas e frugais como antecedentes explicativos de um conjunto relevante de acções. Uso aqui a expressão ‘dar conta’ no sentido davidsoniano, isto é, no sentido de mostrar como é que a *acrasia* é possível definindo as suas condições de possibilidade.

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

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

não induz as pessoas em erro, ou, pelo menos, não as induz nesse erro. Mas pode induzi-las noutros ...

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

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

**P: Qual é o melhor modelo de explicação da acção, na sua opinião?**

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

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

**P: Acha que existem pessoas que agem mais de acordo com o princípio da maximização do que outras?**

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

**P: O que estaria fora do conceito de satisfazibilidade? Em que momento este não se aplicaria?**

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

**P: Esses modelos pretendem ser só descritivos ou também normativos?**

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

**P: A partir do desenvolvimento da teoria económica, e tendo como modelo o modelo da racionalidade ilimitada, tenho dúvidas que, em algum momento, algum autor do século XX tenha aceite essa ideia de racionalidade ilimitada como descritiva. Eles teriam consciência da limitação dessas teorias.**

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

**P: Mas a teoria da acção económica não explica a acção de um indivíduo, não era essa a sua intenção.**

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

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

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

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

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

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

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

**Não identifica especificamente a preferência com o desejo ou com algum resultado da dupla crença-desejo?**

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

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

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

**A ideia.**

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

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

Também, mas não só. Existem igualmente os constrangimentos reais dentro dos quais o agente se move. Aliás, o próprio agente pode também agir no sentido de auto-limitar o conjunto das suas oportunidades futuras. Esta é uma das estratégias a que Elster se



refere na sua teoria da racionalidade imperfeita. Segundo ele, esta estratégia constitui um dos modos de que os agentes humanos se socorrem para vencer o problema da acrasia. Trata-se de uma estratégia de acordo com a qual o agente faz algo agora cujo objectivo consiste em modificar as circunstâncias externas de um modo tal que o campo de oportunidades que estarão à sua disposição num momento de escolha futura significativa diminua de modo relevante. Portanto, para a delimitação do conjunto de oportunidades é importante considerar não apenas as crenças momentâneas do agente, como também os constrangimentos reais no seio dos quais ele tem que mover-se, como, ainda, a própria postura estratégica prévia do agente.

**Um agente que não ignora os custos afundados ao decidir/agir será considerado irracional sob o ponto de vista da economia. Mas como defender este mesmo ponto de vista na teoria da acção? Como fazer essa passagem? (O seu artigo “Efeito dos custos afundados, incontinência e racionalidade” é sobre este ponto, mas gostaríamos que explicitasse aqui essa ideia)**

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

sentido de se auto-instruir. Creio que, no contexto deste segundo componente, comportamentos de honrar custos afundados podem, pelo menos em alguns casos, desempenhar um papel não negligenciável na aquisição pelo agente, e na estabilização futura nele, de padrões de comportamento não desperdiçadores. É óbvio que esta ideia não surgiu do nada. Existe investigação em torno desta questão efectuada no âmbito da Psicologia Cognitiva; e existe evidência empírica, obtida nestas investigações, que parece confirmar ou, pelo menos, dar algum suporte evidencial, a esta ideia.

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

Mais uma vez, para nos entendermos, creio que há aqui que distinguir três objectos de análise distintos: o fenómeno empírico, o conceito que o captura e o rótulo que refere este último. É um facto que a maioria das pessoas não detém o segundo e não conhece o terceiro. Mas isso não significa que, na sua vida real, essas mesmas pessoas não se deparem, com frequência, com problemas que são correctamente caracterizados de acordo com esse conceito que elas não detêm. Repare-se que as situações que dão origem ao fenómeno empírico e que são normalmente ‘encenadas’ nas investigações em Psicologia Cognitiva em torno deste problema são situações triviais: um indivíduo que comprou, por antecipação, um conjunto de lições de ténis e que, depois, descobre que, afinal, até nem gosta muito de jogar ténis, mas que, uma vez que já pagou, resolve levá-las até ao fim, muitas vezes com sofrimento; ou um indivíduo que compra, no âmbito de um pacote apenas parcialmente reembolsável, um fim de semana turístico para um determinado destino que não se encontra de todo no topo da sua lista de preferências, e que, depois de já ter feito a compra, mas antes de iniciar a viagem, encontra um outro pacote para o destino da sua preferência, a um preço bastante mais barato do que o anterior, mas que acaba por escolher viajar para aquele para onde não prefere ir, não só porque já o pagou, mas também por ter sido mais caro, etc. Ora, a generalidade das pessoas já se confrontaram com situações deste género na sua vida de todos os dias. E, independentemente de não deterem o conceito, o que muitas vezes se detecta é que as pessoas se manifestam perplexas com o seu próprio comportamento, sem saberem muito bem explicar porquê. Creio que isso indica alguma coisa.

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

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

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

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

**Portanto, uma acção para ser racional tem que estar de acordo com vários princípios. No caso de Davidson, sempre pensei no princípio de continência como o princípio crucial, e portanto qualquer juízo avaliativo que cumprisse esse princípio seria considerado racional.**

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

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

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

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

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

Num certo sentido, a irracionalidade da incontinência é mais profunda, uma vez que podemos associá-la à auto-contradição. Já da irracionalidade objectiva, presente naquelas ocasiões em que um indivíduo é continente, mas em que a escolha que ele faz como sendo a melhor não é de facto a melhor, de acordo com a informação que ele tem à sua disposição e de acordo com os seus desejos, pode dizer-se que ela é menos profunda, uma vez que, precisamente, não envolve essa dimensão de auto-contradição. Existem depois diagnósticos diferentes acerca da razão pela qual as pessoas agem irracionalmente neste segundo sentido, objectivo, do termo. Por exemplo, do ponto de vista utilitarista, ou hedonista, a justificação standard é a de que isso acontece quando as

peças calcularam mal. Ora, assim como, nas ciências dedutivas, um erro numa multiplicação é considerado como sendo uma anomalia menos grave do que uma contradição lógica explícita, também no ‘ranking’ das irracionalidades práticas incorrer num erro de cálculo no processo maximizador não é considerado tão grave quanto incorrer numa auto-contradição.

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

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

**Mas mesmo assim a acrasia seria possível.**

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

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

### **Podendo ainda ser uma acção intencional?**

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

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

Não tenho a certeza que esse seja o caso – só fazendo exegese do texto de Aristóteles é que se poderia resolver essa questão – mas eu creio que Aristóteles era mais sofisticado do que isso. Mas, mesmo supondo que não o fosse, daí segue-se apenas que, na formulação do problema por Aristóteles, havia uma mistura entre os elementos específicos que são próprios do conceito e outros elementos que, por vezes, se encontram associados com ele, mas que não lhe pertencem, e esses outros elementos devem ser separados do conceito por meio de um trabalho de análise. Seja como for, e independentemente de Aristóteles, não me parece de todo que se possa dizer que o ponto de vista hedonista actual acerca da acrasia esteja preso a quaisquer considerações morais. Por exemplo, do ponto de vista de Elster e Ainslie, como referi acima, o que está em causa na acrasia é o facto de, com a simples passagem do tempo, o agente agir contra a sua escala de preferências sem que tenha havido qualquer reversão explícita das suas preferências. Como também já o disse, este fenómeno é reconduzido, em particular

por Ainslie, a um facto que ele considera ser de natureza puramente empírica, nomeadamente, o facto de, frequentemente, os seres humanos descontarem irracionalmente o futuro de uma maneira hiperbólica. Aqui perguntar-se-á: e porque é que descontar o futuro de uma maneira hiperbólica é irracional? O critério de racionalidade subjacente a este diagnóstico é o critério de racionalidade subjacente aos diferentes axiomas de racionalidade da teoria da decisão: um agente que desconte o futuro com base numa taxa hiperbólica é susceptível de ser explorado até à exaustão por um agente que desconte o futuro usando uma taxa exponencial. Para dar um exemplo clássico: o agente que desconta o futuro de uma forma hiperbólica, quando o Inverno chega ao fim, está na disposição de vender o seu casaco de Inverno por um preço baixíssimo, uma vez que já não precisa dele; depois, quando o Inverno seguinte já está à porta, o agente que tem uma taxa de desconto hiperbólica está na disposição de pagar uma quantia avultadíssima pelo mesmo casaco que vendeu na Primavera por uma quantia ridícula. Evidentemente, um agente que funcione assim não tem grandes hipóteses de sobreviver em competição com agentes que descontem o futuro exponencialmente. Por isso é que a taxa de desconto exponencial é racional enquanto que a hiperbólica é irracional. Ora, Ainslie, um autor indisputavelmente hedonista, defende que a taxa de desconto natural nos seres humanos é hiperbólica e que adoptar uma taxa de desconto exponencial exige esforço cognitivo. Daí que, com frequência, as pessoas – mesmo pessoas relativamente sofisticadas do ponto de vista cognitivo – revertam inconscientemente para a taxa de desconto hiperbólica, que, de alguma forma, teria sido deixada em nós pela evolução natural. A questão que então se coloca é a de saber como poderia explicar-se que a evolução nos tivesse deixado entregues a uma taxa de desconto hiperbólica. Ora, é de facto possível fazer algum sentido evolucionário da presença em nós de uma taxa de desconto deste género, porque é claro que, para que o próximo Inverno possa vir a ser um problema para uma criatura pertencente à biosfera terrestre, ela tem que sobreviver até lá. Seja como for, no desenho de toda esta história, não há qualquer menção a considerações de carácter moral.

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

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

questões morais. Neste sentido, creio que é uma novidade teórica relativamente recente que possa defender-se, por exemplo, que é perfeitamente possível que um indivíduo aja de acordo com os preceitos morais aos quais se sente obrigado e que, não obstante, a sua acção seja incontinente. Ora bem, se este é o caso, e supondo que existe de facto um conceito popular de acrasia, o que não é nada claro, então o facto de esse conceito popular, supondo que existe, estar mais perto do modo como os moralistas cristãos da Idade Média se referiam ao problema do que do modo como os filósofos contemporâneos se referem a ele, é absolutamente natural. Se esse não fosse o caso, não teria sido necessário um dispêndio de esforço intelectual de alguma magnitude para separar os diversos elementos que estavam misturados no conceito original. Ora, se essa energia intelectual foi despendida, e se essa separação se operou, então o putativo conceito popular actual deve, obviamente, estar mais perto do conceito que existia no âmbito da discussão teórica antes de esse esforço conceptual ter tido lugar. Qual é então a grande diferença em relação à situação posta em evidência pelo Knobe-effect? Neste último caso, o problema diagnosticado apenas ganha sentido quando considerado à luz de uma visão específica acerca do estatuto epistémico da chamada ‘psicologia popular’. Esta visão é aquela que defende que as pessoas têm em si próprias um dispositivo de carácter puramente cognitivo para se relacionarem com o comportamento das outras pessoas e, mais em particular, para o explicarem e fazerem previsões acerca dele, e que esse dispositivo corresponderia *grosso modo* à chamada ‘psicologia popular’. Deste ponto de vista, um dos elementos do *pack* cognitivo em questão seria precisamente o conceito de acção intencional. Ora, nas situações experimentais criadas por J. Knobe, a avaliação de intencionalidade que as pessoas fazem muda de acordo com variáveis de carácter moral, as quais têm basicamente que ver com a avaliação do comportamento do sujeito a ser objecto de consideração como sendo digno de louvor ou censura. Segundo Knobe, o facto de tais flutuações ocorrerem mostra que o mecanismo que nos permite relacionarmo-nos com as acções e com os comportamentos sociais dos outros não pode ser um mecanismo de carácter puramente cognitivo, cuja finalidade maior seria produzir explicações e previsões acerca do seu comportamento. Mesmo que *também* sirva esses propósitos, este mecanismo serve primeiramente outros propósitos, sobretudo de carácter avaliativo. Portanto, a conclusão que ele quer tirar é que a competência em causa não é puramente cognitiva, nem sequer predominantemente cognitiva. Ora, esta conclusão contradiz o ponto de vista mencionado acima, de acordo com o qual a psicologia popular seria uma espécie de teoria científica ingénua. Como disse, se esse fosse o caso, ela teria como seu objectivo descrever, explicar e prever o comportamento dos outros, mesmo que apenas de forma ingénua, e não avaliá-lo moralmente. A tese positiva que Knobe extrai desta discrepância é então a de que o conceito popular de acção intencional é um conceito multi-usos, e não um conceito teórico ou pré-teórico cujo objectivo seria o de fazer ciência espontânea ou pré-ciência. Ora bem, não é de todo claro como é que o conceito de acrasia poderia ser relacionado com a crítica de Knobe a esta visão da psicologia popular, uma vez que este não é um conceito psicológico tão básico quanto os conceitos de crença, desejo, intenção ou acção intencional. A própria dificuldade em fixar um rótulo para o referir atesta o seu carácter pouco ‘popular’. Neste sentido, e como já o disse há pouco, não é preciso criar situações experimentais para verificar que o conceito que é usado pelos filósofos contemporâneos é extremamente depurado. Basta ler os textos dos filósofos de há pouco tempo atrás para ver como o conceito se encontra aí misturado com elementos de outra ordem, nomeadamente de ordem moral. Deste modo, uma conclusão *à la* Knobe sobre o conceito de acrasia seria tudo menos surpreendente.



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

Acho estranha a forma como coloca a pergunta porque, precisamente, a grande expansão, a partir dos anos 50, do modelo teórico no qual a teoria da decisão está integrada faz-se em associação com o ponto de vista de acordo com o qual o conceito de racionalidade que está a ser considerado é um conceito dissociado do conceito de consciência. Uma das teses cruciais deste modelo é a de que os agentes são racionais, ou melhor, o comportamento dos agentes tende a ser racional, independentemente da realidade fenomenológica subjacente – este é, aliás, um princípio básico em Economia. E foi esta a ideia que se generalizou.

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

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

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

Isso é uma questão de exegética davidsoniana... Do meu ponto de vista, não creio que se possa dizer de Davidson que ele só conta como uma razão para a acção aqueles conteúdos que tenham estado presentes na consciência do agente nos momentos anteriores à acção. Aliás, creio até que seria bastante estranho dizer-se isso acerca do pensamento de Davidson em matéria de filosofia da acção. Mas, e independentemente do que Davidson diz ou não diz, parece-me também que uma tal afirmação contradiz a nossa própria experiência fenomenológica dos nossos processos de escolha. Não creio que seja muito frequente surpreendermo-nos a nós próprios em exercícios de ponderação das diferentes opções, em que as consideramos a todas de forma consciente e explícita. Evidentemente, há situações nas quais fazemos isso, mas não me parece de todo que sejam a maioria.

**Como responder a este tipo de objecção, típica de um discurso behaviorista?**

**“(...) representações mentais de eventos futuros são exemplos de mentalismo, vítimas de todos os problemas que discutimos (...). Onde está essa intenção interna? De que ela é feita? Como essa porta aberta fantasmagórica poderia causar meu comportamento de abrir o trinco? Como uma representação interna do Moby Dick poderia causar sua procura? Isso não é uma explicação; serve apenas para obscurecer os factos relevantes sobre o ambiente: abrir o trinco normalmente leva a uma porta aberta, e procurar uma mercadoria normalmente produz a mercadoria. Esses factos naturais explicam o comportamento sem nenhuma necessidade de introduzir intenções internas.”**

**(W. M. Baum, Compreendendo o Behaviorismo: comportamento, cultura e evolução. Porto Alegre: Artmed, 2006, p.10)**

**Esta é uma afirmação comum do behaviorismo contra a ideia de que estados mentais podem ser causalmente relevantes. Queria que comentasse.**

Eu não conheço o contexto do qual esta citação foi retirada. Tipicamente, objecções behavioristas deste género são objecções a uma investigação psicológica anterior à Segunda Guerra Mundial, na qual se considerava que os estados e processos mentais deveriam ser considerados de um ponto de vista fenomenológico, de acordo com a definição tradicional, cartesiana, de o que é um estado mental. Basicamente, tratava-se de conteúdos de consciência. Como é sabido, este ponto de vista confronta-se com um problema grave, que tem uma enorme dificuldade em resolver, que é o de perceber como é possível que os estados e processos mentais, concebidos desta forma, possam ter alguma causalidade no mundo empírico. Foi a consciência destas dificuldades que motivou o género de objecções behavioristas que trouxe agora à colação. Ora bem, a partir dos anos 60 do século XX foi possível apresentar uma interpretação do que são estados e processos mentais que é não só diferente da interpretação fenomenológica tradicional mas que tão-pouco se deixa reduzir à ideia behaviorista de que os estados e processos mentais se deixariam caracterizar completamente pela indicação dos seus *input* e *output* fisiológicos e comportamentais típicos. A nova ontologia de estados e processos mentais que então surgiu considera-os basicamente como estados computacionais do cérebro humano, ou de outros órgãos cognitivos. Neste sentido, representações mentais, crenças, desejos, intenções, etc., todos esses estados ou processos mentais de que habitualmente se fala no discurso psicológico, seriam materialmente realizados no cérebro humano, ou noutros órgãos de processamento cognitivo, de uma forma análoga àquela por meio da qual os procedimentos inferenciais, por exemplo, são realizados pelas máquinas computacionais que nós temos à nossa disposição. A partir do momento em que é introduzida esta forma de entender o problema, creio que a objecção behaviorista que referiu perde muito do seu alcance. Com efeito, se nós considerarmos que as representações mentais são, de facto, fisicamente realizadas no cérebro dos agentes, então deixa de haver qualquer obstáculo à ideia de que elas possam desempenhar um papel causalmente relevante no nosso comportamento. A ideia de que os termos para tais entidades seriam ou vazios ou inertes do ponto de vista explicativo deixa assim de exercer qualquer atracção teórica. Evidentemente, de um ponto de vista ainda mais recente, esta visão computacional da mente pode ser posta em causa como sendo algo inadequada; mas o género de objecções

que acabou de mencionar não é o género de objecções que é tipicamente apresentado no âmbito de empreendimentos contemporâneos de crítica à chamada visão computacional da mente. Portanto, se este conjunto de objecções é apresentado contra o ponto de vista fenomenológico acerca da mente, o que há a dizer é que o behaviorismo não é a única alternativa à fenomenologia. O computacionalismo, ou cognitivismo, se quisermos, é também uma alternativa à fenomenologia, e tem sobre o behaviorismo a vantagem de nos permitir continuar a falar em estados e processos mentais como tendo uma existência real e como tendo poderes causais, ao mesmo tempo que tem sobre a fenomenologia a enorme vantagem de, ao defender isso, ter também uma teoria explicativa que permite situar no mundo físico as capacidades causais desses estados e processos. Deste modo, quando, no âmbito da teoria de decisão, interpretada de um ponto de vista cognitivista, se alude a crenças e desejos como causas da acção, a ideia subjacente é que, de facto, elas se encontrariam realizadas fisicamente no cérebro do agente.

**Em que sentido pode considerar-se que razões (pares crença-desejo) explicam causalmente acções?**

**Uma explicação causal no pleno sentido da palavra não dependeria do estabelecimento de leis?**

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

**Então como entender a ideia davidsoniana de que não existem leis psico-físicas?**

Mas o pensamento de Davidson tem os dois aspectos. Quer dizer, creio que ele teria aceite tudo o que eu disse até ao momento. Só que, além do que eu disse, ele diz outras coisas. Mais em particular, ele acrescentaria ao que eu disse, por um lado, que as leis são entidades que pertencem ao domínio linguístico e, por outro lado, que o discurso psicológico e o discurso físico subjacente não são congruentes. É da conjunção destas duas teses que resulta a conclusão davidsoniana de que seria impossível estabelecer leis psico-físicas. Mas esta é uma tese linguística, se quisermos chamar-lhe assim, a qual, de acordo com o ponto de vista de Davidson, não entra em contradição com a tese, que ele também defende, de que a teoria psicológica, tal como se encontra expressa no vocabulário psicológico, refere, de facto, os eventos físicos que produzem as cadeias

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

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

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

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

Para responder a esta pergunta é preciso clarificar primeiro o ponto de vista teórico que depois a teoria da decisão, interpretada de um modo cognitivista, veio formalizar de um modo bastante preciso. Creio que é necessário caracterizar esse ponto de vista recorrendo a dois aspectos e não apenas a um. Um desses aspectos é, de facto, o da causalidade – trata-se de um ponto de vista que pretende indicar as causas das acções; mas há também um outro aspecto que convém não esquecer, que é o do individualismo. De acordo com este ponto de vista, então, a causa da acção encontra-se no interior do agente individual. Ora, estes dois aspectos não têm que estar obrigatoriamente associados. Neste sentido, existem pontos de vista que são também causalistas mas que não são individualistas. É o caso, por exemplo, do ponto de vista estruturalista, o qual não aceita que as causas reais das acções, sobretudo daquelas que são socialmente relevantes, estejam no interior do agente, mas que, no entanto, defende que existe uma explicação causal para essas acções. Para os estruturalistas, os elementos das cadeias causais relevantes são entidades como, por exemplo, as estruturas sociais, e são estas que obrigam, por assim dizer, o agente a agir de uma determinada maneira e não de outra. Este ponto de vista é um ponto de vista alternativo àquele do qual temos estado a falar, mas também é causalista. Eu creio que um tal ponto de vista se depara com escolhos teóricos dificilmente superáveis, porque o modo como o conceito de causalidade opera no interior das estruturas que se pretende ter identificado e os critérios de identidade para individualizar estas últimas são tudo menos claros, mas essa

já é outra questão. Para responder à sua pergunta, o que é importante salientar é que não podemos colocar a questão das opções teóricas disponíveis apenas em termos da oposição teorias causalistas *versus* teorias não causalistas. Há diferentes tipos de teorias causalistas. E todas elas pretendem ser explicativas, no sentido em que se fala da ciência como um empreendimento explicativo. Do mesmo modo, também há diferentes tipos de teorias não-causalistas. De facto, no seio deste outro ponto de vista, pode igualmente distinguir-se entre teorias individualistas, de uma forma geral associadas à tradição fenomenológica, e teorias holistas. Posso apresentar como exemplo destas últimas as tentativas de formalizar teoricamente algumas das ideias expressas pelo segundo Wittgenstein a propósito de jogos, regras, papéis, etc. A ideia aqui é a de que, não sendo de carácter causal, a influência da regra sobre o agente é a instância adequada para a compreensão da acção deste último no âmbito das suas interações com os outros agentes. Isto é, estas acções devem ser compreendidas como lances de jogos determinados por regras impessoais e não como resultado da operação das motivações particulares dos agentes, concebidas como sendo independentes da trama discursiva que transforma comportamentos em acções.

**Também a questão da previsão. Uma das grandes vantagens da teoria da escolha racional é que permite a previsão.**

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

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

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

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

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

**Qual é a sua opinião geral acerca da filosofia experimental?**

Diferentes coisas podem ser ditas acerca da filosofia experimental. Se pensarmos na declaração de princípios que se encontra no manifesto, aquilo de que lá se fala é de um fenómeno que é recorrente na história da filosofia. Aquelas objecções à filosofia percebida como dominante na academia americana são do género das objecções que encontramos na crítica que o jovem Marx fez à análise kantiana da prova ontológica da existência de Deus. Segundo Marx, aquilo que interessaria seria compreender porque é que as pessoas têm o conceito e não demonstrar que, do ponto de vista teórico, não haveria uma forma racional de demonstrar a existência do objecto referido por esse conceito. Portanto, as críticas do manifesto não são propriamente novas. Eu creio que a crítica do jovem Marx a Kant é bastante injusta; a análise crítica que Kant faz na *Crítica da Razão Pura* às diferentes provas da existência de Deus é não só intrinsecamente pertinente como teve sem qualquer espécie de dúvida um impacto extremamente importante na evolução da filosofia e do pensamento europeu subsequentes. Sendo a crítica injusta, isso não significa evidentemente que a questão levantada pelo jovem Marx não seja em si mesma relevante. É importante não apenas discutir a validade conceptual intrínseca dos argumentos como também compreender as razões pelas quais, independentemente da validade conceptual intrínseca dos argumentos, as pessoas aderem a certos conceitos e ideias, e deixam a sua vida ser conduzida por elas. Sem dúvida que também é importante compreender isso, o que, aliás, tem vindo a ser feito. A fundação de empreendimentos teóricos bem organizados, como, por exemplo, a antropologia cultural, teve precisamente como finalidade criar uma área disciplinar na qual se procura alcançar o género de compreensão de que o jovem Marx falava. O que, por sua vez, não quer dizer que, no interior da filosofia, não seja também relevante focar esses aspectos. Mas são os próprios fundadores do movimento da filosofia experimental

quem defende que o ponto de vista que eles apresentam não deve ser visto como alternativo ao ponto de vista que eles tomam por dominante (nos EUA), mas que tem um carácter apenas complementar. Basicamente, o que Knobe e os seus colegas defendem é a criação de um espaço institucional para eles, o que me parece ser perfeitamente legítimo. Todavia, parece-me ser importante chamar aqui a atenção para um aspecto. É claro que eu acho que é intrinsecamente relevante conhecer quais são as intuições que as pessoas de facto têm em associação com os conceitos que usam, em vez de, pura e simplesmente, como alguns filósofos têm uma certa tendência para fazer, se dizer que “a intuição obviamente é esta”. Mas convém sublinhar, mais uma vez levando em conta o que os autores do manifesto escrevem, que o objectivo que eles pretendem alcançar não se esgota nessa investigação. Esse objectivo vai mais além e consiste na tentativa de compreender o modo como a mente humana funciona, o que, evidentemente, é mais que louvável e relevante. Convém chamar a atenção para este aspecto porque, mesmo supondo que os filósofos x, y ou z dizem que a intuição associada ao uso de um termo é obviamente w quando, de facto, para a generalidade das pessoas não é, é preciso deixar claro que a razão pela qual uma certa intuição é aceitável ou razoável não é o facto de ser mais difundida ou menos difundida – não é isso que deve estar em causa. Portanto, os entusiastas deste movimento devem ter em mente, o que nem sempre parece ser o caso, que o objectivo último das investigações efectuadas na filosofia experimental para determinar quais são, na realidade, as intuições que as pessoas de facto têm não se esgota nessa determinação, mas que aquilo que, em última análise, se pretende é encontrar verdades empíricas acerca do modo como a mente humana está organizada. Não creio que seja uma objecção a este trabalho apontar-se para o facto de que esse é o objectivo de um ramo estabelecido da ciência – a Psicologia Cognitiva. Creio que não devem barrar-se caminhos para a descoberta científica com base em justificações de carácter disciplinar ou corporativo. Os programas de investigação justificam-se pelos frutos que dão e não pela sua inserção departamental. Também há filósofos que desenvolvem análises matemáticas rebuscadas no âmbito da interpretação bayesiana da causalidade, por exemplo, apesar de haver matemáticos que fazem a mesma coisa em departamentos de matemática ou estatística. A única coisa que interessa é a qualidade do trabalho.

Publicada sob o título “Acção, Decisão e Explicação da Acção” in Cadilha, S. & Miguéns, S. (coord.): *Acção e Ética – Conversas sobre Racionalidade Prática*. Lisboa: Edições Colibri, 2011, pp. 97-142.