

Os limites da racionalidade: auto-engano e akrasia

Vasco Correia

Universidade Nova de Lisboa

Abstract

In this paper, I argue that ordinary cases of self-deception and akrasia derive from the phenomenon of motivated irrationality. According to the 'motivational' account, self-deception is typically induced by the influence that desires and emotions exert upon our cognitive faculties, and thereby upon the process of belief formation. Crucially, I show that this hypothesis is consistent with the empirical research carried out by social psychologists, and that it avoids a number of paradoxes that undermine the 'intentionalist' account. But motivated irrationality also seems to account for most cases of akrasia, insofar as desires are equally liable to affect the evaluative judgments through which we assess feasible options. This analysis thus sets the premises for a unified account of irrationality.

Keywords

Akrasia, self-deception, irrationality, intentionalism, cognitive bias.

1. O problema da irracionalidade

Para um grande número de filósofos, a *akrasia* (ou «fraqueza da vontade») e o «auto-engano» (*self-deception*) constituem os dois principais desafios para qualquer teoria da racionalidade. Donald Davidson (1985: 346) sublinha a existência de uma analogia formal entre estes dois tipos de irracionalidade: o agente incontinente (*akrates*) é um agente que acredita que tem as melhores razões para fazer A e acaba no entanto por fazer B intencionalmente; e de forma similar, o agente que se engana a si próprio (*self-deceiver*) é um agente que tem as melhores razões para acreditar que *p* e acaba no entanto por se persuadir de que não-*p*.

O problema deste tipo de atitudes é que parecem pôr em causa os princípios de base que qualquer teoria da racionalidade gostaria de poder adotar, e nomeadamente: (1) que os agentes escolhem em

Disputatio, Vol. III, No. 28, May 2010

Received on 6 October 2009

princípio realizar a acção que lhes parece mais vantajosa à luz das suas próprias preferências; e, por outro lado, (2) que os agentes adoptam em princípio a crença que lhes parece mais verosímil à luz da informação de que dispõem.

Segundo Davidson, isto significa que nos fenómenos de auto-engano e de fraqueza da vontade as atitudes do agente transgridem não apenas os seus próprios juízos, mas inclusive as suas próprias normas de racionalidade:

Todas as verdadeiras inconsistências implicam um desvio relativamente às normas da própria pessoa. Isto é válido não apenas para inconsistências manifestamente lógicas mas também para a fraqueza da vontade (como Aristóteles observou), para a fraqueza da garantia e para o auto-engano. (Davidson 1986: 142)

É esse aspecto que causa tanta perplexidade e que explica o interesse e mesmo o fascínio que estes fenómenos têm vindo a suscitar na tradição analítica nas últimas décadas, impondo-se cada vez mais como uma espécie de enigma esfíngico da irracionalidade, ou, como escreve Jon Elster, «o teste crucial que uma teoria da natureza humana deve passar» (Elster 1979: 173). De facto, como compreender que se possa escolher de forma ciente e deliberada a opção menos proveitosa? E como compreender que se possa aderir de sua própria vontade, em pleno conhecimento de causa, a uma crença manifestamente falsa? Mas acima de tudo, estes fenómenos de irracionalidade parecem constituir um observatório privilegiado para o estudo das modalidades da própria *racionalidade*, quer sobre o plano cognitivo quer sobre o plano prático.

Este artigo começa por abordar o tema do auto-engano e só em seguida abordará o não menos delicado tema da acrasia. Esta ordem de argumentação não é accidental, mas assenta na hipótese segundo a qual a compreensão dos fenómenos de irracionalidade cognitiva — não apenas o auto-engano, mas também a «confusão dos desejos com a realidade» (*wishful thinking*) ou ainda os «enviesamentos do juízo» (*cognitive biases*) de que tanto falam os psicólogos sociais — encerra a chave para a compreensão dos próprios casos de irracionalidade prática. Num primeiro tempo, procuraremos desenvolver uma análise crítica da tese dita «intencionalista» sobre o auto-engano, pondo em evidência um conjunto de paradoxos que parecem comprometer a ideia segundo a qual este fenómeno resultaria de uma decisão intenci-

onal do agente. Num segundo tempo, tentaremos mostrar que cada um dos paradoxos examinados se dissipa do ponto de vista da concepção «motivacional» do auto-engano, defendida em particular por Alfred Mele 2001, que sugere que o fenómeno em questão resulta daquilo a que os psicólogos chamam «irracionalidade motivada», isto é, da influência que as emoções e os desejos por vezes exercem sobre os nossos juízos. Se esta hipótese está correcta, é de forma não intencional e inconsciente que ocorrem os casos comuns de auto-engano, assim como, de forma mais geral, os diversos tipos de irracionalidade cognitiva. Por fim, com base nos trabalhos de George Ainslie e de Jon Elster sobre o tema da acrasia, procuraremos mostrar que o fenómeno da irracionalidade cognitiva permite ainda elucidar grande parte dos casos de irracionalidade sobre o plano da acção.

2. Intencionalismo: a vontade de acreditar

No que diz respeito ao auto-engano, a tese que domina entre os filósofos contemporâneos é o chamado «intencionalismo», partilhado sob formas distintas por autores como Donald Davidson, David Pears, Dion Scott-Kakures, Sebastian Gardner e José Luis Bermudez, entre muitos outros. Para estes autores, o auto-engano não é um processo accidental de que o agente seria a simples vítima, mas, ao invés, um processo que requer uma intenção explícita e deliberada de acreditar o que ele próprio sabe ou suspeita ser uma falsidade. Como Davidson sublinha, «o agente que se engana a si mesmo *tem de ter a intenção de enganar*» (Davidson 1986: 87).

Davidson convida o leitor a examinar o caso de Carlos, um indivíduo que está prestes a passar o exame de condução e que tem razões pertinentes para acreditar que irá chumbar. Em primeiro lugar, já chumbou duas vezes no passado e, em segundo lugar, o seu monitor emitiu um parecer claramente pessimista sobre as suas capacidades. E no entanto, não obstante a forma clara como esses elementos fazem pressentir um novo fracasso, Carlos consegue convencer-se a si próprio de que vai passar o exame com sucesso.

Segundo o intencionalismo, o auto-engano que Carlos leva a cabo é perfeitamente similar a qualquer acção que se desenrole sobre o plano prático, embora o efeito dessa acção não seja um acontecimento físico mas um acontecimento mental (uma crença). Em primeiro lugar, Carlos tem *razões* aparentemente legítimas para induzir na sua

mente uma crença manifestamente inverosímil. Com efeito, explica Davidson, a simples possibilidade de chumbar de novo no exame de condução é um pensamento exasperante para Carlos, que não suporta a ideia de fracasso. E em segundo lugar, não é de todo impossível que *o próprio facto de acreditar* que vai conseguir passar o exame aumente as suas probabilidades de passar o exame, na medida em que esta nova crença é susceptível de engendrar uma auto-confiança que é de facto preciosa neste tipo de situações. Tal é pelo menos o pressuposto daquilo a que prosaicamente se chama o «pensamento positivo» (*positive thinking*).

É certo que este tipo de «razão para acreditar» não é de forma alguma apropriada do ponto de vista *epistémico*, uma vez que o desejo que *p* seja o caso não é uma boa razão para acreditar que *p* é de facto o caso. Como Davidson (1982: 297) observa, o meu desejo de possuir um belo bezerro não é uma boa razão para acreditar que possui um belo bezerro. Simplesmente, quando encarada do ponto de vista *prático*, isto é, do ponto de vista do interesse utilitário do agente, este tipo de razão para acreditar revela-se de facto compreensível, na medida em que é de certo modo legítimo, à luz da exigência de maximização do bem-estar individual, preferir uma ilusão positiva a uma verdade amarga¹. Da perspectiva puramente formal das teorias da escolha racional, em todo o caso, nada impede o agente de acreditar naquilo que lhe aprouver. No caso de Carlos, essa ilusão voluntária pode parecer um simples capricho supérfluo — ou mesmo uma ilustração daquilo a que certos autores chamam uma «cobardia epistémica» (Johnston 1989: 85; Barnes 1997: 172) — mas noutros casos compreende-se mais facilmente essa vontade de engano de si próprio, nomeadamente quando a aceitação da verdade crua e nua parece implicar um choque emocional psicologicamente devastador para o agente. Bernard Williams (1989: 148) dá um exemplo desse tipo. Trata-se de um pai que perdeu o filho num naufrágio e que persiste em acreditar que o seu filho ainda pode estar vivo, embora tudo leve a crer o contrário. Num caso deste tipo, é evidente que a crença injustificada se pode revelar extremamente útil do ponto de vista prático, evitando por exemplo que o pai em questão entre numa profunda depressão ou chegue mesmo a cometer um acto de desespero. Outro

¹ Sobre esta questão, ver em particular o livro de Shelley Taylor, *Positive Illusions: Creative Self-Deception and the Healthy Mind*, N.Y., Basic Books, 1989.

caso eloquente é o do paciente em fase terminal que se convence de que ainda tem hipóteses de vencer a doença, mau grado o diagnóstico contrário de vários médicos: ao adoptar esta crença falsa, pode pelo menos aumentar o seu bem-estar psicológico durante o pouco tempo que lhe resta viver.

Conclui-se portanto desta análise que é de certo modo racional adoptar voluntariamente uma crença, desde que o agente considere que ela lhe possa ser benéfica. O auto-engano não é seguramente racional do ponto de vista *epistémico* (ou cognitivo), mas pode revelar-se racional do ponto de vista *prático*.

Resta todavia saber como é que isso é possível, supondo que seja de facto possível. Como fazer para acreditar naquilo que se sabe (ou se suspeita) ser falso? Uma coisa é *desejar* acreditar numa falsidade, outra coisa é *conseguir* acreditar efectivamente numa falsidade em pleno conhecimento de causa. Essa questão tem alimentado um vasto debate entre os defensores do «voluntarismo doxástico», por um lado, que sugerem que é possível induzir voluntariamente a crença que se pretende, e os defensores do «evidencialismo doxástico», por outro lado, que sugerem que as crenças são essencialmente determinadas por factores causais ou epistémicos independentes da vontade do sujeito². Ora, por definição, a concepção intencionalista do auto-engano pressupõe uma certa forma de voluntarismo acerca da crença, visto que só faz sentido falar da intenção de se enganar a si mesmo caso seja efectivamente possível controlar de forma voluntária aquilo em que se acredita. E muito concretamente, os intencionalistas especificam diversas estratégias susceptíveis de permitir a indução deliberada de determinada crença:

- (1) Em primeiro lugar, é possível tentar *simular uma crença*, agindo como se fosse verdade, com a esperança de que a certo momento, por força do hábito, acabará por se tornar uma convicção sincera. Blaise Pascal (1670: 551) recomenda este método aos ateus que desejem adoptar a crença de que Deus existe: ao agirem exactamente como se tivessem fé, ao se submeterem ao ritual das missas, da água benta, das genuflexões e das rezas, sugere Pascal, acabarão por acreditar que Deus existe realmente.

² Sobre esse vasto tema, ver por exemplo Robert Audi, «Doxastic Voluntarism and the Ethics of Belief», *Facta Philosophica*, 1, 1, 1999, p. 89; Pascal Engel, «Volitionism and Voluntarism About Belief», in A. Meijers (org.), *Belief, Cognition and the Will*, Tilburg U.P., 1999, p. 9-25; e Jonathan Bennett, «Why is Believing Involuntary?», *Analysis*, 50, 1990, p. 87-107.

- (2) Em segundo lugar, como nota Williams 1989, existe também a possibilidade de recorrer a *técnicas de sugestão*, como por exemplo a hipnose. Um agente com falta de confiança nas suas competências profissionais, por exemplo, pode pedir a um psicólogo que o hipnotize e que o convença por essa via de que é realmente um profissional competente. Segundo Naylor (1985: 429) esta estratégia poderá inclusive ser executada pelo próprio indivíduo, graças à técnica de auto-hipnose.
- (3) Em terceiro lugar, é possível tentar *manipular selectivamente a informação* disponível, de modo a promover indirectamente a crença pretendida. Davidson concebe um estratagema astucioso que ilustra perfeitamente este tipo de iniciativa: «Posso tentar evitar um encontro desagradável previsto para o ano que vem, escrevendo deliberadamente uma data falsa na minha agenda e antecipando que a minha má memória fará com que já me tenha esquecido desse gesto quando o dia em questão chegar» (Davidson 1986: 145). Alfred Mele (2001: 16) concebe o caso semelhante de um homem que deturpa intencionalmente a informação que relata no seu diário íntimo, descrevendo em termos elogiosos o seu próprio comportamento, de modo a obter mais tarde uma melhor imagem de si mesmo, ao reler aquelas páginas num momento em que já terá esquecido o seu artifício. Neste tipo de casos, pode dizer-se que o «eu presente» do agente procura enganar o seu «eu futuro».
- (4) Por fim, o agente que se queira convencer a si mesmo da proposição *p* tem ainda a possibilidade de *orientar selectivamente a sua atenção*, fazendo um esforço para se focar nos dados que parecem sugerir que *p* e, reciprocamente, esforçando-se por desviar a atenção da informação que parece infirmar a realidade de *p*. Para a maioria dos autores intencionalistas, é precisamente este o método mais comum de auto-engano.

É contudo importante sublinhar que nenhum destes métodos de auto-persuasão implica a versão mais forte do voluntarismo doxástico, por vezes chamado voluntarismo «directo», segundo o qual é possível decidir acreditar qualquer coisa por simples decreto da vontade, *hic et nunc*. Os métodos acima descritos apenas requerem uma posição mais moderada: o chamando voluntarismo «indirecto», segundo o qual o controlo voluntário da crença é possível mediante artifícios indirectos, à semelhança dos mencionados, que visam uma modificação estratégica das condições causais de formação da crença.

A principal vantagem da concepção intencionalista é que possibilita uma extensão das normas éticas ao domínio epistémico. Com efeito, se o auto-engano resulta efectivamente de uma intenção consciente e deliberada, precisamente como acontece com as acções intencionais do domínio prático, nesse caso torna-se legítimo imputar ao agente a

plena responsabilidade pelas suas crenças irracionais. Desde logo, o agente que se engana a si mesmo deixa de ser encarado como a simples vítima das suas ilusões, passando a ser encarado como o próprio autor dessas ilusões; alguém que escolhe deliberadamente «enterrar a cabeça na areia» ou que «recusa encarar de frente a realidade», para usar algumas das expressões mais comuns. Ora, esse aspecto revela-se particularmente delicado se tivermos em consideração que a adopção de uma ilusão nem sempre tem efeitos benéficos. De facto, acontece por vezes que o auto-engano acabe por ter consequências nefastas, não apenas para o agente em questão mas também para os outros. Trata-se de um ponto já salientado com veemência por William Clifford 1878, num texto que inaugura o debate em torno do tema da responsabilidade epistémica e da «ética da crença». Clifford evoca o exemplo de um armador que decide negligenciar o mau estado do seu navio, evitando intencionalmente prestar atenção aos indícios que apontam para a necessidade de proceder a reparações, simplesmente porque isso aumenta o seu conforto financeiro e psicológico. O argumento de Clifford é simples: tal armador não é apenas responsável pela ilusão que daí deriva, mas também pelas vidas que eventualmente colocará em perigo ao deixar o navio largar amarras sem inspecção. A principal virtude do intencionalismo é portanto a de nos tornar responsáveis não apenas pelas nossas acções, mas também pelas nossas crenças e pela própria maneira como raciocinamos.

3. Os paradoxos do intencionalismo

Porém, o intencionalismo confronta-se a uma série de antinomias que ameaçam a sua viabilidade. As duas primeiras foram postas em evidência por Alfred Mele (1998: 38):

- (1) *Paradoxo doxástico*. Para que me possa enganar a mim mesmo, primeiro preciso de conhecer a verdade, pois de outro modo não estaria propriamente a mentir a mim mesmo, mas quando muito a cometer um erro involuntário. Sartre também sublinha este aspecto: «Enquanto [sujeito] enganador tenho de conhecer a verdade que me é dissimulada enquanto [sujeito] enganado» (Sartre 1943: 84). Mas se no início eu reconheço a realidade que acabo por conseguir negar graças a um processo de engano de mim mesmo, então é porque, a certa altura, devo ter mantido duas crenças contraditórias em simultâneo. Mas será que isso é psicológico?

gicamente possível? Será possível, por exemplo, acreditar que estou a perder cabelo e ao mesmo tempo acreditar no contrário?

- (2) *Paradoxo da estratégia*. Por um lado, o agente que se engana a si mesmo tem de estar consciente quer da sua *intenção* quer da sua *estratégia* de auto-engano, de maneira a poder levá-las a cabo com sucesso. Mas por outro lado, dado que ele é ao mesmo tempo a vítima desse engano, a própria consciência que ele tem da dita intenção e da dita estratégia de auto-engano parece comprometer a eficácia das mesmas. De facto, como pode o agente cair credulamente na sua própria armadilha se sabe à partida da sua existência?
- (3) *Paradoxo económico dos casos positivos*. Segundo o intencionalismo, o objectivo do auto-engano é obter o prazer ou a tranquilidade que proporciona a crença de que aquilo que se deseja é uma realidade. No entanto, em grande parte dos casos parece deveras contra-producente decidir adoptar uma crença falsa em nome do bem-estar psicológico, na medida em que *o próprio facto de acreditar numa falsidade* impede o agente de intervir eficazmente no mundo no intuito de maximizar esse bem-estar. Para voltar ao exemplo de Davidson, podemos observar que, ao adoptar a crença que irá conseguir passar no exame de condução, simplesmente em virtude do seu desejo de acreditar que tal será o caso, Carlos acabará por reduzir ainda mais as suas hipóteses de passar no exame, dado que essa crença falsa o impedirá de tomar medidas susceptíveis de facilitar a sua ambição (por exemplo: inscrever-se em aulas suplementares ou adiar o exame até se encontrar mais bem preparado).
- (4) *Paradoxo económico dos casos negativos*. Por fim, como compreender que o engano de si mesmo em relação à crença de que *p* possa ser motivado pelo desejo de que *p*, quando existem «casos negativos» de auto-engano nos quais o sujeito se persuade a si mesmo de coisas extremamente desagradáveis? Por que motivo, por exemplo, um ciumento compulsivo tomaria a decisão de se auto-persuadir de que a sua mulher o anda a enganar? E como explicar que o paranóico veja por todo o lado indícios que confirmam os seus receios mais angustiantes? Que benefício tiram estes agentes do processo de auto-engano, supondo que exista de facto uma intenção deliberada de acreditar na realidade ilusória?

No intuito de superar estes paradoxos, a maioria dos intencionalistas são levados a postular a hipótese da divisão da mente (Davidson, Audi, Pears, Fingarette, Gardner). Deste ponto de vista, quando se diz que o sujeito *S* se engana a si mesmo, subentende-se no fundo que *é uma parte* da sua mente — no caso das concepções freudianas, o Inconsciente ou a Censura (ou ainda o *Id*, na segunda tópica) — que engana a outra parte

ou o outro subsistema da sua mente — no caso das concepções freudianas, a Consciência (ou o Ego, na segunda tópica).

Importa de resto sublinhar que a hipótese da divisão da mente permite também explicar os casos de irracionalidade que dizem respeito ao plano da acção. O fenómeno da *acrasia* (ou «fraqueza da vontade»), em particular, será explicado como o resultado de um conflito entre as partes que compõem o indivíduo: ou entre a Razão e o Desejo (modelo platónico), ou entre a Vontade e o Desejo (modelo estóico). Cícero, por exemplo, considera que a incontinência no plano da acção se explica por uma fraqueza da Vontade ante a tentação do Desejo. Davidson descreve este ponto de vista nos seguintes termos: «Cabe à vontade decidir quem ganha a batalha. Se a Vontade é forte, dá a palma à razão; se é fraca, pode deixar que a paixão ou o prazer leve a melhor» (Davidson 1980: 35). Daí o uso de expressões como por exemplo «Fui incapaz de resistir à tentação!», ou ainda «Foi mais forte que eu!», que empregamos como se a força que é *mais forte do que nós* — em geral um simples desejo — não fizesse parte de nós exactamente ao mesmo título que a vontade ou a razão, às quais pelos vistos nos preferimos identificar.

Contudo, a hipótese divisionista é uma explicação *ad hoc* que parece levantar ainda mais problemas do que aqueles que pretende resolver. A objecção mais consequente, assinalada por Wittgenstein nas *Investigações Filosóficas*, é que essa tese implica a atribuição de atitudes como desejos, representações, memórias, sensações ou mesmo emoções a subsistemas diferenciados da mente, quando que só faz sentido atribuir essas atitudes à pessoa no seu todo ou ao *in*-divíduo enquanto tal. Wittgenstein escreve: «Só de um ser humano vivo e do que lhe é similar (no seu comportamento) se pode dizer: tem sensações, vê, é cego, ouve, está consciente ou inconsciente» (Wittgenstein 1953: 281). Anthony Kenny, por sua vez, chama «falácia do homúnculo» essa atribuição ilegítima de propriedades antropomórficas a supostas partes constituintes da mente (1984: 125).

4. A concepção motivacional

Existe no entanto uma concepção alternativa do auto-engano que permite evitar os paradoxos da irracionalidade sem obrigar a postular a hipótese da divisão da mente. Trata-se da concepção dita «motivacional», inicialmente proposta por Alfred Mele 2001, à qual aderem

também Ariela Lazar 1999 e Pierre Livet 2001. Segundo esta concepção, o auto-engano não deriva de uma intenção deliberada com raiz num raciocínio prático, mas resulta simplesmente do fenómeno de irracionalidade motivada, a que os psicólogos por vezes se referem como «irracionalidade quente» (*hot irrationality*), que consiste na influência que os nossos desejos (ou outras emoções) são susceptíveis de exercer sobre o processo de formação da crença. De certo modo, este ponto de vista conforta uma intuição do senso comum que os poetas e os provérbios populares sempre souberam veicular: «o pior cego é aquele que não quer ver»; «acreditamos facilmente naquilo que nos convém» (*quod volumus, facile credimus*); «o amor é cego»; «a mente é sempre enganada pelo coração»³, etc. À luz da concepção motivacional, os desejos e as emoções do sujeito irracional deixam portanto de aparecer como pretensas *razões* para adoptar determinadas crenças de forma intencional e passam a aparecer simplesmente como *causas* irreflectidas das crenças ilusórias do sujeito.

Esta hipótese apoia-se na profusão de estudos empíricos que a psicologia social tem vindo a realizar nas últimas décadas, sobretudo graças aos trabalhos de Daniel Kahneman e de Amos Tverski, que demonstram que as emoções deformam ou «enviesam» (*bias*) os nossos raciocínios e as nossas crenças ao afectarem as faculdades cognitivas responsáveis pelo processo de formação do juízo⁴. Por um lado, as emoções afectam a *direcção* e a própria *intensidade da atenção* que prestamos aos dados disponíveis aquando da formação do juízo. Uma pessoa demasiado ciumenta, por exemplo, tende a consagrar uma atenção excessiva às atitudes da pessoa que ama, e arrisca-se por conseguinte a interpretar uma simples suspeita como uma prova irrefutável de traição. Por outro lado, as emoções viciam o processo heurístico de *recolha de informação*. Por exemplo, o desejo de que a minha posição filosófica seja a correcta pode conduzir-me a negligenciar a consideração de dados susceptíveis de a infirmar e, reciprocamente, levar-me a sobrevalorizar os dados que parecem confirmá-la. Em terceiro lugar, as emoções deformam a própria interpretação dos

³ La Rochefoucauld, *Réflexions morales*, máxima 102.

⁴ Ver sobretudo as duas obras de referência seguintes: D. Kahneman, A. Tverski e P. Slovic, (orgs.), *Judgement under Uncertainty: Heuristics and Biases*, Cambridge, Cambridge University Press, 1978; T. Gilovich, D. Griffin e D. Kahneman (orgs.), *Heuristics and Biases: The Psychology of Intuitive Judgement*, Cambridge, Cambridge University Press, 2002.

dados que tomamos em consideração. É frequente, por exemplo, a pessoa apaixonada interpretar um simples gesto de amizade como a expressão inequívoca de um amor recíproco. E por fim, as emoções condicionam o desempenho da própria memória, o que corresponde em grande parte ao fenómeno de *memória selectiva*, uma vez que tendemos a lembrar-nos mais facilmente de acontecimentos «marcantes», associados a determinada emoção forte, do que de acontecimentos emocionalmente neutros.

Isolada ou conjuntamente, estes factores são portanto susceptíveis de nos conduzir de forma insidiosa e involuntária, sem qualquer intenção premeditada, a formar crenças em desacordo com o que os dados de que dispomos sugerem claramente. Para um observador neutro, o desfasamento entre a crença do sujeito irracional e a informação disponível pode parecer de tal maneira significativo, que é tentador pensar que o sujeito em questão está apenas a simular uma crença. Mas isso é ignorar até que ponto a nossa interpretação da realidade pode ser deturpada pela nossa afectividade, em particular quando a realidade em questão é objecto de desejos ou de receios extremamente fortes.

Em relação ao intencionalismo, a hipótese motivacional apresenta três grandes vantagens. Em primeiro lugar, como já foi referido, encontra-se em plena conformidade com o que revelam os estudos empíricos dos psicólogos sociais. Embora estes testes não permitam falsificar a tese intencionalista, permitem em todo o caso revelar que a tese motivacional se verifica num vasto número de casos. Em segundo lugar, torna-se também possível elaborar uma *teoria generalizada da irracionalidade cognitiva*, na medida em que é precisamente o mesmo fenómeno psicológico — a irracionalidade motivada — que permite explicar não apenas o auto-engano (*self-deception*), mas também a tomada dos desejos por realidades (*wishful thinking*) e mesmo toda a variedade de mecanismos de enviesamento do juízo (*cognitive bias*) que os psicólogos actualmente distinguem.

Mas acima de tudo, em terceiro lugar, a hipótese motivacional permite superar cada um dos paradoxos acima formulados — que por conseguinte não são os «paradoxos do auto-engano», como se costuma dar a entender, mas mais exactamente os *paradoxos da concepção intencionalista do auto-engano*. Com efeito, a partir do momento em que se considera que o auto-engano é um processo involuntário e não intencional, deixam desde logo de fazer sentido os dois primeiros paradoxos: por um lado, não existe coabitação na mesma mente de

duas crenças contraditórias (paradoxo doxástico), mas quando muito passagem progressiva de uma crença a outra; e, por outro lado, nem sequer se coloca o problema da consciência que o sujeito irracional supostamente teria da sua estratégia de auto-engano, visto que o sujeito é apenas vítima de uma ilusão, mas de forma alguma o seu autor. Por razões similares, os dois paradoxos económicos caem também por terra, na medida em que a causa do auto-engano é o efeito de uma emoção sobre as faculdades cognitivas do sujeito. Ora, contrariamente ao raciocínio prático e às intenções deliberadas que dele derivam, as emoções não têm por vocação maximizar o bem-estar do agente, nem a longo prazo (paradoxo dos casos positivos) nem sequer a curto prazo (paradoxo dos casos negativos)⁵. Por fim, a hipótese motivacional permite dispensar o postulado polémico da divisão da mente, e evitar assim o paradoxo do homúnculo, uma vez que o conflito interno não é suposto instaurar-se entre dois subsistemas relativamente autónomos da mente, mas somente entre atitudes mentais particulares (emoções e crenças). Não quer isso dizer que seja impossível que a mente sofra por vezes cisões internas, como parece ser de facto o caso em certas patologias de dissociação mental: trata-se apenas de observar que a tese da divisão da mente *deixa de ser indispensável* para explicar a esmagadora maioria dos casos de auto-engano.

Mas por outro lado, é importante sublinhar que a hipótese motivacional não implica uma desresponsabilização total do agente relativamente às ilusões de que é vítima. Como observa Pascal Engel, o facto de o fenómeno de crença ser essencialmente involuntário não nos deve fazer esquecer que temos um certo controlo sobre as nossas faculdades cognitivas, e por conseguinte sobre o processo de formação da crença (1999: 437). À imagem dos juízes e dos jornalistas escrupulosos, temos a possibilidade de opor um esforço de objectividade e de fiabilidade à maneira como tratamos a informação e como raciocinamos. Por exemplo, um médico que não detecta as marcas de violência no corpo de

⁵ O paradoxo dos «casos negativos» de irracionalidade cognitiva (crença irracional em algo de indesejável) pode de resto encontrar uma explicação plausível no fenómeno da «irracionalidade emocional», ou seja, no facto que nos acontece sentir determinadas emoções mesmo quando o contexto não as justifica de todo. Assim, por exemplo, não é raro que um indivíduo excessivamente ciumento se persuade a si mesmo de que a sua companheira o anda a enganar precisamente em virtude do carácter irracional do seu sentimento de ciúme e da pressão que este exerce de forma ou de outra sobre o seu juízo.

uma criança, interpretando como acidental o que manifestamente resulta de sevícias recorrentes, pode seguramente estar a ser vítima de uma ilusão cognitiva (talvez a repugnância de imaginar tamanho horror o leve a tal negligência). Contudo, esse médico permanece inteiramente responsável pelo seu juízo e pelas suas consequências, pois incumbê-lo verificar e certificar-se através de um exame minucioso que tomou em consideração todos os elementos pertinentes. Assim, se é verdade que o conteúdo das nossas crenças não depende da nossa vontade, não é menos verdade que temos a possibilidade — e por conseguinte o dever — de assegurar a fiabilidade do processo de formação das nossas crenças. A extensão da esfera da responsabilidade ao domínio epistémico permanece portanto um projecto legítimo e mesmo necessário, como disso atesta a panóplia de trabalhos recentemente consagrados ao tema da «ética da crença»⁶.

5. As raízes cognitivas da *acrasia*

Existe além do mais um outro motivo que torna ainda mais premente a questão da ética da crença, que é o problema do impacto dos juízos irracionais sobre a esfera da própria racionalidade prática. Com efeito, o fenómeno de irracionalidade cognitiva acima descrito é susceptível por sua vez de afectar os *juízos avaliativos* em função dos quais deliberamos e fazemos as nossas escolhas, isto é, os juízos que nos permitem estimar o valor relativo de cada opção executável, e por essa via determinar qual a acção mais vantajosa à luz das nossas preferências.

Ora, segundo alguns autores contemporâneos, o fenómeno da *acrasia* (ou «fraqueza da vontade») resulta precisamente do carácter irracional do juízo no momento da deliberação. George Ainslie e Jon Elster, em particular, sugerem que é uma ilusão cognitiva de «depreciação» que está na origem da acção contrária ao melhor juízo do agente. Segundo esta hipótese, o agente incontinente (*acrates*) não é um indivíduo que age deliberadamente contra o seu melhor juízo, contrariamente ao que defende Davidson, mas mais exactamente um

⁶ Ver por exemplo Linda Zagzebski, *Virtues of the Mind*, Cambridge, Cambridge University Press, 1996 ; Pascal Engel, «L'éthique de la croyance et la justification épistémique», *Carrefour*, XXIII, 2, Ottawa, 2001, p. 19-38 ; Ernest Sosa, *Knowledge in Perspective*, Cambridge, Cambridge University Press, 1991; James Montmarquet, *Epistemic Virtue and Doxastic Responsibility*, Rowman and Littlefield, Lanham MD, 1993.

agente que age de acordo com um juízo temporariamente ilusório. O problema, explicam os autores, é que somos frequentemente vítimas de uma espécie de «miopia» das preferências, que faz com que o valor das opções futuras nos pareça inferior ao que realmente é, enquanto o valor das opções imediatas, reciprocamente, nos parece superior ao que realmente é. Existe portanto uma depreciação do valor das opções disponíveis proporcionalmente ao intervalo de tempo que as separa do momento presente: quanto mais longínqua a opção, menor nos parecerá o seu valor relativo (*caeteris paribus*). Além disso, os trabalhos de Ainslie (1992, 2001) revelam que essa progressão não é *exponencial*, como seria de supor, mas *hiperbólica*, ou seja evolui em progressão geométrica até ao momento da escolha, o que significa que no momento que precede imediatamente a primeira opção se pode verificar uma inversão abrupta e temporária das preferências: «teremos tendência a preferir temporariamente benefícios menores e imediatos a benefícios maiores mas mais tardios, durante o lapso de tempo em que são iminentes» (Ainslie 2001: 38).

É por essa razão que nos acontece com frequência trair uma decisão inicial reflectida e ponderada, como por exemplo fazer uma dieta, devido a uma revisão súbita e precipitada das preferências que nos conduzirá, por um curto instante, à ilusão de que a opção imediata (comer uma sobremesa) é mais proveitosa. Embora Ainslie permaneça vago sobre esse ponto, parece legítimo supor que essa ilusão cognitiva é motivada pela influência do desejo imediato sobre o juízo avaliativo do sujeito. Elster, por seu lado, parece considerar que as emoções em geral têm um papel determinante na formação do juízo irracional: «os apetites e as emoções podem subverter a racionalidade da acção mediante uma influência negativa sobre a formação das crenças racionais», nomeadamente na medida em que «essas motivações viscerais induzem facilmente crenças motivadas» (Elster 2007: 58).

Como quer que seja, pode concluir-se desta análise que a acrasia não implica no sentido estrito uma acção contrária ao melhor juízo do agente, mas apenas uma revisão súbita das suas preferências sob a pressão e a influência de um desejo iminente. Elster sublinha este aspecto num artigo muito crítico da análise de Davidson, onde descreve precisamente o exemplo de um agente que decide fazer uma dieta por razões de saúde, mas que acaba por ceder à tentação no final de uma refeição: «À medida que a refeição avança, ocorre uma inversão das preferências no momento t^* , e quando o empregado de mesa lhe pergunta, no momento 2, se deseja uma sobremesa, ele responde

pela afirmativa (...) Antes de entrar no restaurante e, suponhamos, depois de sair, [o agente] queria evitar comer uma sobremesa. No entanto, ele não está a agir contrariamente ao seu melhor juízo *no momento em que manda vir uma sobremesa*⁷. É até provável que o próprio agente reconheça, ao deixar o restaurante, que escolheu de facto a opção errada, mas só uma ilusão retrospectiva (*hindsight bias*) pode levar esse agente a acreditar que sabia no próprio momento da escolha que aquela não era a melhor opção. Supondo que as nossas acções revelam as nossas preferências, o facto é que no preciso momento da escolha a sobremesa lhe pareceu um benefício superior.

Os únicos casos de irracionalidade prática nos quais o agente escolhe conscientemente a pior opção, ao que tudo indica, são os casos patológicos de compulsão. No caso das fobias, por exemplo, acontece com frequência que o paciente reconheça conscientemente o carácter inofensivo do objecto do seu medo, mas se sinta todavia compelido a agir em função desse medo, em detrimento do seu interesse utilitário ou do seu bem estar⁸. Mas se esta hipótese está correcta, se de facto a maioria dos casos de irracionalidade prática resulta de uma ilusão cognitiva, então a acrasia deve ser encarada como o resultado de uma «fraqueza do juízo», e não tanto como uma pretensa «fraqueza da vontade».

A nossa análise permite-nos portanto concluir que os casos mais comuns de auto-engano e de acrasia têm origem num fenómeno precisamente idêntico: a irracionalidade cognitiva motivada, isto é, a formação do juízo sob a influência de determinados afectos. Neste sentido, torna-se possível tentar elaborar uma concepção unificada da irracionalidade que seja válida quer sobre o plano da acção quer sobre o plano da crença. No que diz respeito à questão do auto-engano, vimos que o argumento mais decisivo contra o intencionalismo é o aspecto intrinsecamente paradoxal da teoria, que desemboca numa série de aporias dificilmente superáveis. A hipótese da divisão da mente permite evitar um ou outro desses paradoxos, mas trata-se de

⁷ J. Elster, "Davidson on Weakness of Will and Self-Deception", in L. Hahn (org.), *The Philosophy of Donald Davidson*, La Salle, Open Court, 1999, p. 430.

⁸ Contudo, mesmo nesse tipo de casos, parece-nos difícil admitir que é por «falta de força de vontade» que o paciente se vê impelido a agir de maneira irracional (é no entanto o que propõe J. Feinberg no texto «What is So Special About Mental Illness?», in *Doing and Deserving*, Princeton U.P., 1970). Uma hipótese mais plausível é que a origem das compulsões resida no fenómeno da irracionalidade emocional, tal como acontece nos «casos negativos» de auto-engano.

uma pressuposição *ad hoc* que por sua vez levanta novas dificuldades, como revelam em particular as reflexões de Wittgenstein e de Kenny. Além disso, parece pouco provável que os casos mais frequentes de auto-engano e de acrasia impliquem todos uma cisão na mente do sujeito. Em todo o caso, nenhum dado empírico do nosso conhecimento parece dar fundamento a essa hipótese, sendo que, na nosologia psiquiátrica do *Manual Diagnóstico e Estatístico de Desordens Mentais (DSM-IV)*, o fenómeno de «dissociação mental» permanece um caso patológico grave que pouco tem a ver com os casos comuns de irracionalidade aqui examinados. Ao invés, a concepção motivacional proposta por Mele, Livet e Lazar apresenta como vimos três grandes vantagens relativamente à hipótese rival: em primeiro lugar, permite superar cada um dos paradoxos do intencionalismo; em segundo lugar, encontra-se em conformidade com os estudos empíricos da psicologia social; e, em terceiro lugar, permite explicar a grande maioria dos casos de irracionalidade cognitiva sem recorrer à hipótese da divisão da mente. Mas supondo que é de facto de maneira involuntária e inconsciente que se formam as nossas crenças irracionais, a questão da responsabilidade epistémica recai por inteiro sobre o próprio processo de formação da crença, na medida em que podemos exercer um certo controlo sobre as condições nas quais se realiza. Se o ponto de vista que defendemos está correcto, a racionalidade das nossas crenças — e, indirectamente, a racionalidade das nossas acções — teria muito a beneficiar com a elaboração de métodos de raciocínio capazes de aumentar a fiabilidade do processo de formação do juízo em condições reais.

Vasco Correia

Instituto de Filosofia da Linguagem

Faculdade de Ciências Sociais e Humanas, Universidade Nova de Lisboa

Edifício I&D - 4º piso, Av. de Berna, 26, 1069-061 Lisboa

vasco_saragoca@hotmail.com

Referências

- Ainslie, George. 2001. *Break-down of the Will*, Cambridge: Cambridge University Press.
- Barnes, Annette. 1997. *Seeing Through Self-Deception*, Cambridge: Cambridge University Press.

- Clifford, William. 1878. The Ethics of Belief. In *Philosophy: Contemporary Perspectives on Perennial Issues*, ed. by Klemke, Kline and Holinger, New York: St. Martin's Press, 1994.
- Davidson, 1980. How is Weakness of Will Possible? In *Essays on Actions and Events*, Oxford: Oxford University Press.
- Davidson, 1982. Paradoxes of Irrationality. In *Philosophical Essays on Freud*, ed. by Wollheim and Hopkins. Cambridge: Cambridge University Press.
- Davidson, Donald. 1985. Incoherence and Irrationality. *Dialectica* 39: 345-354.
- Davidson, Donald. 1986. Deception and Division. In *Actions and Events*, ed. by Le Pore and McLaughlin., New York: Basil Blackwell.
- Elster, Jon. 1979. *Ulysses and the Sirens*. Cambridge: Cambridge University Press.
- Elster, Jon. 2007. *Agir contre soi*. Paris: Odile Jacob.
- Engel, Pascal. 2001. Sommes-nous responsables de nos croyances ?, In *Qu'est-ce que la culture?*, ed. by Michaux, Paris: Odile Jacob.
- Johnston, Mark. 1989. Self-Deception and the Nature of Mind. In *Perspectives on Self-Deception*, ed. by McLaughlin and Rorty, Berkeley, L.A., London: University of California Press.
- Kenny, Anthony. 1984. *The Legacy of Wittgenstein*, Oxford: Basil Blackwell.
- Lazar, Ariela. 1999. Deceiving Oneself or Self-Deceived? On the Formation of Beliefs *Under the Influence*, *Mind*, 108: 265-290.
- Livet, Pierre. 2001. *Self-Deception*, émotions et revisions. In *Les figures de la duperie de soi*, ed. by Giovanni, Paris: Kimé.
- Mele, Alfred. 1987. *Irrationality*, New York, Oxford: Oxford University Press.
- Mele, Alfred. 1998. Two Paradoxes of Self-Deception. In *Self-Deception and Paradoxes of Rationality*, ed. by Dupuy, Stanford: CSLI Publications.
- Mele, Alfred. 2001. *Self-Deception Unmasked*, Princeton, Oxford: Princeton University Press.
- Naylor, Margery. 1985. Voluntary Belief. *Philosophy and Phenomenological Research* 45: 427-436.
- Pascal, Blaise. 1670. In *Oeuvres complètes*. Paris: Seuil, 1963.
- Sartre, Jean-Paul. 1943. *L'être et le néant*, Paris: Gallimard.
- Williams, Bernard. 1973. Deciding to Believe. In *Problems of the Self*, ed. by Elster, Cambridge: Cambridge University Press.