"MLAG is an innovative research group within academic philosophy in Portugal. Its members work in areas of philosophy related to cognitive science, areas in which philosophical work can be of great interest to people outside philosophy, such as philosophy of mind, philosophy of language and philosophy of action. I am very happy to have helped start their work, sharing with people from the University of Porto the way we do things in Rutgers - judging by the results, it proved to be more than inspiring"

O MLAG é um grupo de investigação inovador no seio da filosofia académica em Portugal. Os membros do grupo trabalham em áreas da filosofia relacionadas com a ciência cognitiva, áreas nas quais o trabalho filosófico pode ser de grande interesse para pessoas fora da filosofia, tais como a filosofia da mente, a filosofia da linguagem e a filosofia da acção. Sinto-me muito feliz por ter ajudado o grupo a começar, partilhando com as pessoas da Universidade do Porto a forma como fazemos as coisas em Rutgers – avaliando pelos resultados, a partilha foi mais do que inspiradora

Ernest Lepore
Professor, Rutgers University, Philosophy Department
Director, Rutgers Center for Cognitive Science

Sofia Miguens
Department of Philosophy
Mind Language and Action Group (MLAG) – Institute of Philosophy
University of Porto

Carlos E. E. Mauro
Mind Language and Action Group (MLAG) – Institute of Philosophy
University of Porto

U.PORTO

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA E DO ENSINO SUPERIOR Portugal

Faculdade de Letras da Universidade do Porto

Instituto de Filosofia da Faculdade de Letras da Universidade do Porto

SOFIA MIGUENS & CARLOS E. E. MAURO

PERSPECTIVES ON RATIONALITY

Perspectives on Rationality

Edited by

SOFIA MIGUENS
CARLOS E. E. MAURO

MIND LANGUAGE AND ACTION DISCUSSION PAPERS
RATIONALITY BELIEF AND DESIRE SERIES

Perspectives on Rationality

Perspectives on Rationality

Edited by / Coordenação

Sofia Miguens
Carlos E. E. Mauro

Mind, Language and Action Discussion Papers
Rationality, Belief, Desire series **(volume I)**

Mind Language and Action Group - MLAG
Gabinete de Filosofia Moderna e Contemporânea
Instituto de Filosofia
Faculdade de Letras
Universidade do Porto

Apoios: FCT, FLUP, MLAG

Perpectives on Rationality

Perspectives on Rationality; coord.: Sofia Miguens e Carlos E. E. Mauro
ISBN 972-8932-21-9

I – Sofia Miguens, 1968 –
II – Carlos E. E. Mauro, 1972 –

Mind, Language, Action Discussion Papers Collection

Colecção *Mind, Language, Action Discussion Papers*
Coord.: Sofia Miguens e Carlos E. E. Mauro
ISSN 1646-6527

Depósito Legal nº

Edição
Faculdade de Letras da Universidade do Porto
Porto, Dezembro de 2006

Prefácio

Iniciamos, com esta publicação, uma série destinada a trazer ao público os trabalhos em curso no MLAG (Mind, Language and Action Group), o grupo de investigação em filosofia analítica do Gabinete de Filosofia Moderna e Contemporânea do Instituto de Filosofia da Faculdade de Letras da Universidade do Porto. Na medida em que os membros do MLAG estão envolvidos em trabalhos de ordem muito diferente – desde artigos e comunicações a apresentar em conferências nacionais e internacionais até à elaboração de dissertações, à produção de antologias e manuais, etc –, as publicações da responsabilidade do grupo serão de diferente teor. Esperamos que essa variedade possa satisfazer as diversas necessidades de leitores interessados nas questões da filosofia da mente, filosofia da linguagem e filosofia da acção.

Procuramos neste volume disponibilizar alguns resultados do Projecto *Rationality, Belief Desire II – from cognitive science to philosophy* (POCI/FIL/55555/2004) *(2005-2007)*, que sucedeu a *Rationality Belief Desire I – motivation for action from the viewpoint of the theory of mind (2003-2005)*. Em geral os projectos de investigação do MLAG situam-se na área da filosofia contemporânea de orientação analítica, com ênfase nas áreas da filosofia da mente, da linguagem e da acção, e tendo em consideração um contexto de ciência cognitiva. Os artigos que incluímos neste primeiro volume da série provêm na quase totalidade de um Colóquio que organizámos na FLUP em Junho de 2006 no âmbito do Projecto *Rationality, Belief, Desire II*, o Primeiro Colóquio-MLAG. Alguns outros textos são da autoria de membros do MLAG que na altura não puderam participar no Colóquio, mas que têm tido o seu trabalho em discussão no âmbito do grupo.

Agradecemos à Fundação para a Ciência e a Tecnologia, que tem sido o principal patrocinador das nossas actividades, quer, neste caso, através do Projecto POCI/FIL/55555/2004, quer das várias bolsas de Doutoramento atribuídas a membros do grupo. Agradecemos também à Professora Maria José Cantista, coordenadora do Gabinete de Filosofia Moderna e Contemporânea, estrutura na qual o MLAG se insere, bem

como à FLUP, e a todos aqueles que de uma forma ou outra têm acompanhado as actividades do MLAG. A origem do próprio MLAG deve muito ao incentivo, ao apoio e à amizade de Ernest Lepore (Universidade de Rutgers); em geral tem sido muito importante para nós o contacto e o intercâmbio com outros grupos de investigação dedicados a áreas próximas das nossas – devemos aqui destacar o Departamento de Lógica e Filosofia Moral da Universidade de Santiago de Compostela e o Instituto de Filosofia da Linguagem da Universidade Nova de Lisboa. Agradecemos também a Charles Travis (King's College, London), que tem estado permanentemente disponível para colaborar connosco enquanto consultor do Projecto.

Aproveitamos esta ocasião para agradecer ao Professor Eduardo Rego, do Departamento de Matemática da FCUP, a sua incansável e muito útil participação no Colóquio que deu origem ao volume.

Pelas revisões linguísticas e sugestões estruturais de vária ordem não podemos deixar de agradecer a Susana Cadilha (FLUP), David Davies (FLUP) e João Alberto Pinto (FLUP).

Porto, Dezembro de 2006

Sofia Miguens
Carlos E. E. Mauro

Preface

This is the first volume of a new series, the *Mind, Language and Action discussion papers*. Through that series we intend to render available to a wider audience the research carried on within the Mind, Language and Action Group (MLAG). MLAG is the analytically oriented research group of the Gabinete de Filosofia Moderna e Contemporânea (Institute of Philosophy – FLUP). Since the members of MLAG are involved in quite different kinds of works, ranging from articles and talks, to dissertations, translations, anthologies, textbooks, etc, we expect the following volumes to be of different kinds. Anyway we hope that they will meet the needs of readers interested in issues concerning philosophy of mind, philosophy of language and philosophy of action.

Most articles included in this volume were originally presented as talks given at the *Rationality, Belief, Desire II* Workshop we organized at Faculdade de Letras in June 2006. Some others are written by members of MLAG who at that time could not be present, but whose work is being discussed within the group.

We would like to thank our main sponsor, Fundação para a Ciência e a Tecnologia, for making our work possible through research grants (in this case POCI/FIL/55555/2004) and doctoral scholarships awarded to several members of the group. We also want to thank the coordinator of Gabinete de Filosofia Moderna e Contemporânea, Professor Maria José Cantista, for her support. We are also thankful to the institution where most of our work takes place, Faculdade de Letras da Universidade do Porto, which has provided us with material conditions without which such work would not be possible. We are thankful to many people who in way or another have helped MLAG's activities. The group's existence owes much to Ernest Lepore (Rutgers University) – his support, his example and his friendship were immensely important. Contact and exchange with research groups which are close to us, both geographically and thematically, has also been essential – that is the case of Departamento de Lógica y Filosofia

Moral (Universidad de Santiago de Compostela) and Instituto de Filosofia da Linguagem (Faculdade de Ciências Sociais e Humanas – Universidade Nova de Lisboa). We are very grateful to people in both institutions. We would also like to thank Charles Travis (King's College, London), who has been permanently available and most helpful as consultant of the Project.

We also thank Professor Eduardo Rego, from the Department of Mathematics – University of Porto for his generous and helpful participation in the RBD2 Workshop from which most of the material published here comes.

Finally, for their many structural and linguistic suggestions regarding this volume, we are especially thankful to Susana Cadilha (FLUP), João Alberto Pinto (FLUP) and David Davies (FLUP).

Porto, December, 2006

Sofia Miguens
Carlos E. E. Mauro

Índice

Racionalidade, Crença, Desejo: um programa de investigação

Sofia Miguens[1]

> «Alguns animais pensam e raciocinam; ponderam, testam e rejeitam hipóteses, agem em função de razões, por vezes após deliberarem, imaginarem consequências e pesarem probabilidades; eles têm desejos, esperanças e ódios, às vezes por boas razões. Também cometem erros no cálculo, agem contra o seu próprio melhor juízo, ou aceitam doutrinas com base em evidência não adequada. Qualquer um destes feitos, actividades, acções ou erros é suficiente para mostrar que um tal animal é um animal racional, pois ser um animal racional é simplesmente isso, ter atitudes proposicionais, não importa quão confusas, injustificadas, ou erróneas estas atitudes possam ser. Proponho que esta é a resposta. A pergunta é: que animais são racionais?» Donald Davidson, *Rational Animals*.

1. Natureza e motivações do Projecto *Rationality, Belief, Desire II – from cognitive science to philosophy*. Atribuições de racionalidade e de irracionalidade.

Os artigos aqui reunidos resultam do trabalho em curso no âmbito do Projecto de investigação *Rationality, Belief, Desire II – from cognitive science to philosophy* (POCI/FIL/55555/2004) e devem ser vistos como explorações dos temas pelos quais o Projecto se ramifica[2]. No centro dos nossos interesses estão os vários aspectos da *racionalidade.* A racionalidade é tomada como característica de *agentes cognitivos*. Numa primeira descrição, agentes cognitivos são sistemas guiados por representações que se comportam de forma adaptada ao ambiente em função de finalidades[3]. Entre os agentes

[1] Investigadora Responsável do *Rationality, Belief, Desire II – from cognitive science to philosophy* (POCI/FIL/55555/2004)

[2] O livro introdutório *Racionalidade* (Miguens 2004) serviu com proposta programática para o desenvolvimento do presente Projecto (cf. a lista de questões para uma teoria filosófica da racionalidade aí apresentada, pp. 19-45).

[3] Esta é uma definição operacional, utilizada como ponto de partida – obviamente passa por cima de toda a discussão em ciência cognitiva acerca do que poderão ser representações e finalidades em sistemas cognitivos físicos. A nossa única justificação é o nível, num agente cognitivo, dos processos que sobretudo nos interessam (a que poderíamos chamar processos cognitivos

cognitivos estão incluídos os humanos, mas não apenas estes. Consideramos que para falar de agentes cognitivos não é estritamente necessário evocar desde logo a consciência. Esta consideração provém dos estudos da cognição. Em contrapartida, na filosofia, a análise da racionalidade de agentes conduz a questões específicas de *racionalidade prática* (racionalidade na decisão e na acção) e *racionalidade teórica* (racionalidade no raciocínio e no processo de fixação ou revisão de crenças) e frequentemente pressupõe a consciência dos agentes. De qualquer forma, e deixando de momento em suspenso a questão da consciência, para chegar às questões práticas e teóricas da racionalidade é necessário, pelo menos em princípio, falar de agentes. Ora, considerar certas partes do mundo como agentes supõe considerá-las dotadas de crenças e desejos e utilizar uma linguagem mentalista para as descrever. Do ponto de vista da filosofia da mente – embora estas questões se prolonguem pela epistemologia e pela metafísica – a natureza de tais estados e de tal linguagem está em aberto. Por essas razões o quadro de referência do Projecto é a filosofia da mente – é isso que é sintetizado no título pelo propósito de abordar conjuntamente a natureza da *racionalidade* e a natureza das *crenças e* dos *desejos*.

Pensamos que é fundamental, para compreender a racionalidade, procurar compreender os fenómenos de *irracionalidade*, quer teórica, quer prática. O desejo de avaliar de forma fundamentada as acusações de irracionalidade feitas a agentes, raciocínios, decisões e instituições nas mais diversas circunstâncias foi uma motivação fundamental para o Projecto. Interessou-nos sobretudo o facto de as acusações de irracionalidade serem supostamente superadas por proclamações ou apelos à racionalidade (como quando numa discussão as pessoas não se entendem, mas também quando se procura sustentar o estatuto de leis jurídicas e morais, quando se fala do estatuto da ciência, quando se fala de progresso de sociedades ou da qualidade de uma decisão política). Na verdade, é bastante frequente proclamar a racionalidade ou a irracionalidade sem aprofundar o que com isso se quer dizer. Mas devemos ter bem presente que é muito mais simples dizer o que é um argumento válido do que dizer o que é uma inferência justificada, uma decisão apropriada ou uma criatura racional.

Olhando para nós próprios, aparentemente, todos queremos ser

superiores). No artigo incluído neste volume, no entanto, J. P. MAÇORANO explora o problema da natureza da representação como dizendo respeito às relações entre informação e aquilo a que de um ponto de vista mentalista chamamos crença.

racionais, no sentido em que nenhum de nós quer ser considerado irracional: não queremos ter crenças não fundamentadas, não queremos raciocinar mal, não queremos decidir mal, não queremos agir contra o nosso melhor juízo. Mas porquê? O que é que isso tem a ver com a nossa forma de sermos mentais e humanos? O facto é que muito embora aparentemente desejemos ser racionais, muito frequentemente admitimos não o ser – os humanos parecem ser capazes de agir contra o seu melhor juízo[4], de acreditar coisas que não têm razões para acreditar, de não acreditar naquilo que têm razão para acreditar, de acreditar em contradições, etc. Por que será assim? Será que quando acontece acreditarmos no que não devíamos acreditar, ou fazermos o que não devíamos fazer, se trata apenas de um desvio por ignorância, como quando desconhecemos uma regra que se aplicaria a dado momento na resolução de um determinado problema, ou será que a explicação é mais complicada, e procurá-la envolve saber mais sobre a forma como as nossas mentes são?

Para desenvolver uma abordagem dos problemas da racionalidade e da irracionalidade propusemo-nos explorar um conjunto de questões em torno do aspecto volitivo da nossa subjectividade, questões que de alguma forma nos decompõem conceptualmente enquanto agentes. As questões que se seguem foram constante objecto de atenção no Projecto: O que são desejos? O que são intenções? O que são emoções? O que é que finalmente nos motiva a agir? Seremos egoístas psicológicos, sempre em última análise motivados pelo interesse próprio? Será que apenas essa forma de agir pode conduzir a alguma forma de satisfação ou felicidade? Qual é a origem de tal egoísmo? Poderá (e deverá) ele ser de alguma forma superado na acção moral e racional? Será possível sabermos o que queremos sem sabermos o que somos? Como se relaciona a estrutura do nosso querer com aquilo que pensamos que somos, i.e. com as nossas auto-representações enquanto agentes? De onde virá a motivação para em algumas ocasiões considerarmos menos os nossos próprios desejos do que as necessidades de outras pessoas, ou da sociedade (por exemplo a necessidade de justiça)? Será a acção moral, realmente, ou em última análise, não egoísta ou será que para haver sequer motivação para a acção tem que haver sempre egoísmo? Como se relacionam egoísmo e emoções? Serão as emoções puramente irracionais? Será que em

[4] Esta é obviamente uma afirmação polémica, e não mereceria sequer o acordo de todos os membros do MLAG envolvidos no Projecto RBD2 (cf. Mauro & Cadilha, no presente volume).

agentes como nós razões podem realmente ser causas de acções? Em função de tudo isto, e para além das definições disponíveis e comummente utilizadas[5], o que é afinal racionalidade na acção?

De facto, estas são questões com as quais o nosso primeiro projecto sobre racionalidade, (*Rationality, Belief, Desire – motivation for action from the viewpoint of the theory of mind, 2003-2005*) já lidava. O projecto RBD1 centrou-se na questão da motivação para a acção e nele foram consideradas questões tais como a concepção instrumental de racionalidade[6], o modelo crença-desejo de explicação da acção[7], a natureza das razões para agir, a relação entre razão e paixões (ou desejos) em concepções mais ou menos racionalistas da estrutura do querer[8], a especificidade, face a crenças e desejos, de estados mentais tais como intenções[9], teorias filosóficas das emoções[10] (aqui veio a interessar-nos sobretudo o debate cognitivismo-não cognitivismo), o egoísmo psicológico enquanto núcleo da teoria da escolha racional, os fundamentos psicológico-flosóficos da economia tanto quanto estes envolvem a ideia de escolha racional de agentes[11], etc. Desta forma, o nosso primeiro projecto assumiu inevitavelmente uma forte componente de filosofia da acção e filosofia moral[12]. É ainda

[5] Por exemplo a definição instrumental, de acordo com a qual é racional o agente que mobiliza os meios adequados à prossecução dos fins que tem em vista (evidentemente para isso os fins devem estar já dados – e de onde virão?) e a ideia de 'maximização da utilidade esperada' (para que o agente possa agir de forma a maximizar a utilidade esperada, as suas preferências devem estar de alguma forma definidas – e como é que isso acontece?). Isto para não falar de presunção de que os agentes têm as crenças e desejos que lhes permitem pensar nas coisas do mundo em termos de meios e fins, utilidade e probabilidade.

[6] Cf. por exemplo Madeira 2003ª, Madeira 2004

[7] Madeira 2003a.

[8] Miguens 2004, Capítulo 2 (Filosofia e racionalidade prática – o que devemos fazer?).

[9] Madeira 2003b.

[10] Mendonça 2004.

[11] Cf. a investigação de doutoramento de C. E. E. Mauro – embora esta tenha vindo a assumir contornos mais complexos, começou por visar precisamente os fundamentos filosófico-psicológicos da economia e especialmente a noção de egoísmo psicológico enquanto essência, em última instância, da teoria da escolha racional. O egoísmo psicológico é a concepção segundo a qual as pessoas são sempre motivadas pelo interesse pessoal; no limite, a ideia é que o agente racional age sempre de forma auto-interessada, maximizando a utilidade esperada de sua acção.

[12] Esta foi uma consequência do desenvolvimento do Projecto RBD1. À partida procurou-se simplesmente mapear os estudos da racionalidade, teórica e prática. Muitos dos trabalhos acabaram por centrar-se na resposta a duas questões: i) *O que move um agente à acção?* e ii) *Que forma tem uma teoria da racionalidade na filosofia? (*aqui foram tomadas como referência as obras de S. Stich, A. Goldman e R. Nozick e S. Blackburn).

isso que se passa, de resto, com o projecto RBD2.

Uma outra motivação fundamental para os nossos dois Projectos sobre racionalidade, além do interesse pelos fenómenos de irracionalidade, era de natureza mais técnica mas dizia também ela respeito a uma necessidade específica de fazer atribuições de racionalidade[13]. Estávamos interessados nos problemas que as teorias da interpretação na filosofia da mente enfrentam. Em teorias da interpretação, ou teorias interpretativas, como as de W. V. Quine, D. Davidson e D. Dennett, o ponto de partida da teoria da mente é uma atribuição de racionalidade[14]. Quanto a este aspecto, a origem mais longínqua dos projectos sobre racionalidade foi o meu trabalho sobre a Teoria dos Sistemas Intencionais de Daniel Dennett[15]. A Teoria dos Sistemas Intencionais – estou a chamar assim o conjunto de posições defendidas por D. Dennett em filosofia da mente – é uma teoria quineana da interpretação, e está comprometida com a atribuição aos agentes de crenças na sua maioria verdadeiras e de inferências na sua maior parte racionais. Na ausência de tal suposição seria pura e simplesmente impossível considerar outras partes do mundo como mentais. No caso específico da Teoria dos Sistemas Intencionais, a atribuição de racionalidade é condição para as teorias da representação, da consciência, da acção e da pessoalidade. Do esclarecimento do estatuto de tal suposição depende, assim, todo o edifício. Por esta razão, no trabalho referido questionava-se já quais seriam as condições e as implicações de tal suposição. Perguntava-se, em particular, se se trataria de uma suposição apriorista, como parece ser o caso numa outra teoria interpretativa do mental, a de Donald Davidson. A resposta era negativa. Antes de dizer porquê, e porque a posição de Davidson constitui para nós uma referência, começo por recordá-la. Em *Could*

Uma vez levado a cabo o mapeamento acima referido, as referências ramificaram-se, e as questões tratadas tornaram-se mais específicas; neste momento alguns membros do MLAG estão interessados sobretudo em teoria da acção e filosofia moral, outros em questão relativas à natureza da lógica, outros em filosofia da mente, tratando questões que vão desde as teorias da interpretação, teoria das emoções, causação mental, teorias da identidade, etc.

[13]Aqui deveria dizer-se 'imputações de racionalidade *e argumentos a favor da impossibilidade de irracionalidade'*.

[14] Considero para todos os efeitos que a origem deste posicionamento se encontra na forma como W. V. Quine considera o princípio da caridade no contexto da tradução radical (cf. *Word and Object*, 1960, p.59).

[15] A avaliação da Teoria dos Sistemas Intencionais enquanto teoria da mente é o principal objectivo de Miguens 2002.

There Be a Science of Rationality[16], Davidson defende o seguinte acerca da sua teoria do pensamento, linguagem e acção: «Toda a teoria está construída sobre as normas da racionalidade; foram estas normas que sugeriram a teoria e que lhe deram a estrutura que tem. Mas tudo isto está já nas partes formais e axiomatizáveis da teoria da decisão e da teoria da verdade, e estas são tão precisas e claras como qualquer teoria formal da física. No entanto, normas ou considerações de racionalidade também entram com a aplicação da teoria a agentes reais, no momento em que o intérprete atribui as suas próprias frases para capturar os conteúdos dos pensamentos e enunciações de um outro agente. O processo necessariamente envolve decidir que padrão de atribuições torna o outro inteligível (não inteligente, evidentemente!) e isto é uma questão de usar os nossos próprios standards de racionalidade para calibrar os pensamentos de um outro agente. De alguma forma é como ajustar uma curva a um conjunto de pontos, o que se faz na melhor das ciências. Mas há um elemento adicional no caso psicológico: na física temos uma mente a trabalhar para fazer tanto sentido quanto possível de um objecto que está a ser tratado como sendo sem-cérebro (*brainless*): no caso psicológico, existe um cérebro em cada ponta. Normas estão a ser empregues como standard de normas.»[17].

A ideia de Davidson é portanto que as características daquilo que é atribuído na interpretação são determinadas por teorias formais, previamente ao processo de interpretação aplicado a agentes reais. Podemos ver aqui circularidade, podemos também considerar que Davidson apresenta um argumento apriorista a favor da impossibilidade de irracionalidade daquilo que conta como mental. Antes de recusarmos esta forma de pensar, de a vermos como um círculo vicioso, vale a pena parar para considerar que talvez as coisas não possam ser muito diferentes se de facto os constrangimentos da racionalidade estão inscritos na nossa forma de sermos seres mentais e linguísticos. E nós damos por nós sendo mentais e linguísticos – não podemos propriamente dar um passo atrás, sair dessa condição para a ver a partir de fora, e só então descrevê-la. Como afirma John Searle, «Podemos debater de forma inteligível teorias da racionalidade, mas não a racionalidade[18]».

De qualquer forma podemos, em seguida, perguntar por alternativas. Qual seria a alternativa à ideia apriorista davidsoniana?

[16] Davidson 2004a.
[17] Davidson 2004a: 130
[18] Searle 2001: xiv.

Uma alternativa seria por exemplo considerar que é o design real de agentes que sustenta a atribuição de racionalidade, sendo o design ele próprio resultante de evolução por selecção natural. A racionalidade estaria assim para a cognição como a adaptação para a vida. Quer no caso da vida quer no caso da cognição estaríamos perante fenómenos de função e adaptação. No caso específico de Dennett, a rejeição da etiqueta de 'instrumentalismo' para a sua teoria da interpretação tem a ver com isto mesmo – ele considera que a Teoria dos Sistemas Intencionais não é instrumentalista na medida em que aquilo que o intérprete faz não é projectar racionalidade mas sim reconhecer padrões existentes, apoiado na realidade do design de agentes[19]. A explicação do design é posteriormente remetida para a teoria da evolução por selecção natural, o que, no caso do design mental, significa remeter a teoria da mente para considerações sub-pessoais. Desta linha de pensamento resulta um segundo argumento a favor da impossibilidade de irracionalidade de agentes reais, além do argumento apriorista davidsoniano[20].

Em geral, se os argumentos a favor da impossibilidade de irracionalidade se sustentam, isso significa que devemos conceber a irracionalidade como um fenómeno no seio da racionalidade, um fenómeno para o qual é preciso encontrar lugar e explicação[21]. Como diz Davidson, «o tipo de irracionalidade que cria problemas conceptuais não é a falha de uma outra pessoa em acreditar ou sentir ou fazer o que nós consideramos racional, mas a falha, no interior de uma única pessoa, na coerência ou consistência no padrão de crenças, atitudes, emoções, intenções e acções.»[22]. Talvez esta seja a forma correcta de ver as coisas. No entanto, a tentação de considerar como irracional no raciocínio ou na decisão das outras pessoas aquilo que *nós* consideramos irracional não é facilmente evitável – na verdade muita da literatura empírica sobre racionalidade enfrenta este problema. Em última análise, trata-se de saber se a racionalidade e a irracionalidade – para usar a expressão de J. Cohen[23] – podem ser empiricamente demonstradas. Será que faz sentido considerar que aborda agentes reais

[19] Para uma apresentação resumida desta ideia, cf. Miguens 2006b.
[20] Curiosamente o tema da impossibilidade da irracionalidade é retomado nos trabalhos mais recentes do projecto relativamente à racionalidade prática. Cf. nomeadamente Mauro & Cadilha, no presente volume, onde é defendida a tese da impossibilidade de irracionalidade no momento da acção.
[21] Cf. Amen, no presente volume.
[22] Davidson 2004 a: 170.
[23] Cohen 1981.

sem qualquer pressuposição acerca da sua racionalidade ou irracionalidade, e que se vem a descobrir pela experiência se este é racional ou irracional? Ou haverá algo de errado nesta forma de ver as coisas?

Alguns estudos empíricos do raciocínio e da decisão apresentam-se como tendo provado a irracionalidade das pessoas. A pretensão seria substanciada pelo facto de tais estudos mostrarem que os juízos e decisões das pessoas não são conformes a alguma visão ideal da racionalidade dada pela lógica, teoria das probabilidades ou teoria da decisão: não é em função de tais princípios que as pessoas julgam e decidem. As pessoas pensam e decidem apoiadas antes em princípios heurísticos que simplificam os processos e que são em geral eficazes mas que conduzem também a distorções e enviesamentos persistentes (pense-se por exemplo nos clássicos estudos reunidos por A. Tversky, P. Slovic e D. Kahneman[24]).

Mais recentemente alguns autores procuraram contestar tais conclusões acerca de irracionalidade persistente[25] – é esse por exemplo o caso dos estudos na psicologia evolucionista que reinterpretam 'casos de irracionalidade' bem conhecidos na literatura, como o caso de Linda ou a performance na tarefa de Wason. As distorções persistentes seriam função de características adaptativas dos mecanismos cognitivos dos humanos, não devendo por isso ser consideradas irracionais.

Será então que nenhuma prática, mecanismo ou componente das mentes humanas pode, quando considerado de um ponto de visto evolutivo, ser irracional? Esta seria uma conclusão demasiado radical.[26]

---

[24] Kahneman, Slovic & Tversky 1982.

[25] Cosmides e Tooby 1996, Barkow, Cosmides & Tooby 1992. Cf., para uma descrição do problema, Miguens 2004, pp. 84-88.

[26] Um contra-exemplo analisado em Stich & Sripada 2005 diz respeito às emoções. Mesmo sendo estas resultantes de um processo evolutivo, pode dar-se o caso de existirem actualmente emoções inadaptadas. Procurando explicitar o que se faz quando se utiliza standards de racionalidade como modelos para a investigação empírica, é considerada em Samuels, Stich & Tremoulet 2003 a possibilidade de se tratar da caracterização de uma competência, à maneira do Conhecimento de Língua chomskyano. A ser esse o caso, torna-se necessário considerar a arquitectura cognitiva dos agentes e questões de modularidade. Entre as teses demasiado pessimistas (de acordo com as quais seríamos constitutivamente irracionais) e respostas demasiado optimistas (tais performances são justificadas pela história evolutiva da espécie), Samuels, Stich & Tremoulet procuram, em *Rethinking Rationality*, um artigo que foi muito discutido nas actividades do Projecto, defender uma hipótese intermédia. Em Stich & Sripada procura-se ainda mostrar como, mesmo a partir de um ponto de vista evolucionista, o qual é evocado por alguns autores para afastar as acusações de irracionalidade, é possível compreender a persistência em agentes de dispositivos que são *actualmente* irracionais. Stich e Sripada procuram evitar o

Penso que existe aqui uma questão prévia à interpretação dos dados resultantes de investigação empírica, e que é aqui que faz sentido a tese de Cohen de acordo com a qual a irracionalidade não pode ser empiricamente demonstrada. O ponto de Cohen era que princípios normativos tais como os da lógica ou da teoria das probabilidades não devem ser considerados como hipóteses de ciência natural, hipóteses que podem ser testadas, confirmadas ou refutadas. Eles constituem antes o quadro de referência da abordagem. Esta tese vai de encontro às posições em filosofia da mente a que me referi acima, de acordo com as quais não podemos sequer considerar qualquer parte do mundo como irracional sem uma suposição prévia de racionalidade – a irracionalidade é um fenómeno no interior da racionalidade. Claro que fica em aberto aqui saber que racionalidade é esta, que não pode ser identificada com os cânones formais usuais. Não é também nada claro quão racionais têm que ser os agentes a quem é imputada mente – o que é certo é que constrangimentos demasiado exigentes (tais como capacidade perfeita de cálculo, consistência de crenças, etc) parecem impossíveis de sustentar.

No trabalho a que tenho vindo a referir-me, a partir de Dennett e por razões relacionadas com teoria da cognição, defendi que a racionalidade aqui evocada não pode ser uma noção muito profunda, nem sequer definível com precisão[27]. Trata-se de uma noção do nível do agente, comportamentalmente ancorada, ligada a relações instrumentais entre meios e fins. Ela aplica-se ao agente como um todo e não tem sequer por base géneros naturais que seriam representações (isto seria um mecanismo da racionalidade à la Fodor). Uma tal noção mínima de racionalidade pode ser essencial para fazer teoria da mente, mas não é de todo defensável dar o passo seguinte que consistiria em considerar que a sua natureza é explicitada por 'cânones de racionalidade' tais como a lógica, a teoria das probabilidades ou a teoria da decisão. É neste sentido que uma tal 'racionalidade' é por princípio insusceptível de uma caracterização precisa: é uma noção pragmática, que não deve ser considerada como o nome para qualquer valor cognitivo intrínseco. Ao defender esta posição[28] não pretendi de forma alguma desfazer, ou

'excesso' da oscilação, quando se considera as emoções, do clássico veredicto de irracionalidade para a ideia de que as emoções 'são racionais como tudo o que resulta de evolução'. Agradeço especialmente a Tomás Carneiro o estudo, exposição e discussão destas questões, bem como a tradução dos dois artigos referidos.

[27] Miguens 2002: 510.

[28] Miguens 2002.

deflacionar, os problemas da racionalidade teórica e da racionalidade prática (o que devemos acreditar? o que devemos fazer?), mas apenas fazer notar que não estamos justificados, ao abordá-los, em colocar a racionalidade como a plataforma segura, evocada como uma última palavra. Os problemas da racionalidade teórica e prática – o que é racionalidade na acção? O que é racionalidade no raciocínio? – continuam em aberto, e são tão pertinentes como antes, apenas menos susceptíveis de uma resposta óbvia. A racionalidade evocada nas teorias da interpretação não é, portanto, perfeita. Esta tese tem, obviamente, que ser elaborada e desenvolvida. De qualquer forma o importante aqui é constatar que não parece ser possível enumerar a priori um conjunto de crenças verdadeiras e de princípios de inferência sem os quais não chamaríamos a um agente racional. S. Stich formula isto no âmbito da sua teoria pragmatista da racionalidade dizendo que não apenas 'racionalidade' não é o nome para algum valor cognitivo intrínseco como não há possibilidade de formular a priori constrangimentos para todos os agentes racionais possíveis[29].

As teorias da interpretação na filosofia da mente interessaram-nos particularmente devido ao papel que nelas a racionalidade desempenha. Mas, naturalmente, a pergunta pelo estatuto da suposição de racionalidade que acompanha as teorias da interpretação é apenas um dos aspectos pelos quais é possível questionar se tais teorias serão sustentáveis. As teorias da interpretação são, como é sabido, teorias anti-reducionistas[30], e constitui um problema de fundo saber se o anti-reducionismo quanto ao mental é em última análise coerente. Um dos grandes problemas aqui é certamente esse 'resto inexplicado' das teorias interpretativas do mental que ao mesmo tempo se pretendem monistas ou fisicalistas que é o intérprete: o que é o intérprete? de onde vem a racionalidade por ele atribuída? J. Fodor[31], que é ele próprio um anti-reducionista, mas que defende um tipo de anti-reducionismo totalmente diferente, acusa a abordagem de Dennett de ser 'transcendental', e para Fodor isso não é um elogio, antes se relaciona com ausência de explicação.

Da perspectiva de Fodor, admitir a existência de representações reais é o primeiro passo para poder falar da racionalidade de agentes: as

[29] Stich 1990.

[30] Davidson é particularmente claro quanto às 'razões' para o anti-reducionismo: estas são a normatividade da interpretação, o carácter causal de conceitos mentais tais como 'acção' e o externalismo.

[31] Cf. Miguens 2005b.

representações são mais fundamentais do que a racionalidade e são o princípio para a explicação desta última de um ponto de vista cognitivo. Ora, aí onde o realismo intencional fodoriano coloca representações mentais reais, possibilitando uma explicação sub-pessoal da racionalidade de agentes reportada às computações sobre representações, as teorias interpretativas, enquanto teorias transcendentais, não colocariam nada. O problema não é simples, e em última análise não diz respeito apenas a mecanismos cognitivos: trata-se de saber como é possível conjugar de forma coerente (se é que é de todo possível) naturalismo e fisicalismo com normatividade e subjectividade.

Além da questão geral acerca do lugar do mental no mundo físico, a abordagem da questão da racionalidade do ponto de vista da filosofia da mente, conduziu-nos ainda a questões específicas acerca do tipo de mentes que são as mentes humanas e do tipo de acções que são as acções humanas – não sendo as mentes humanas o único tipo de mente, nem as acções humanas o único tipo de comportamento de agentes no mundo, o facto é que a teoria da mente e da acção que se defende se reflecte inevitavelmente em algo que nos interessa bastante enquanto humanos: uma determinada concepção do humano. No trabalho que tenho vindo a referir[32], quando se considerava a aplicação das teorias da representação e da consciência à forma de pensar na natureza de pessoas e acções, objectava-se ao intelectualismo das noções de consciência e de pessoalidade defendidas por Dennett. Uma discussão em curso no Projecto é aquela pela qual se pretende objectar a um análogo intelectualismo que se encontra na teoria davidsoniana: no artigo de onde foi retirada a citação que abre o presente texto, Davidson começa por considerar que nem crianças com uma semana nem caracóis podem ser justificadamente consideradas criaturas racionais[33] – apenas criaturas na posse do conceitos de crença e de verdade são capazes de pensamento objectivo e merecem ser chamadas racionais[34]. Entre outras coisas, interessou-nos avaliar a razoabilidade desta pretensão, bem como explorar as alternativas.

---

[32] Miguens 2002, Capítulo 4.

[33] Davidson 2001: 95.

[34] Creio que há algo de errado com o intelectualismo de Dennett e Davidson acerca destas questões – e parte do que há de errado tem a ver com a ausência de consideração da percepção na teoria da mente. Em termos de desenvolvimento do Projecto – cf. entrevistas com Charles Travis – esse tem sido um caminho explorado.

2. Questões metodológicas: ciência cognitiva e filosofia.

As duas motivações de partida dos projectos sobre racionalidade, ambas relacionadas com atribuições de racionalidade, conduziram-nos à filosofia da acção, filosofia moral e filosofia da mente. Tornam-se necessários agora alguns esclarecimentos metodológicos para compreender a extensão dos temas do projecto a outros campos. Na medida em que o projecto RBD2 foi concebido a partir de um foco metodológico estas considerações têm também relevância teórica por si.

Se o nosso primeiro projecto se centrou na questão da motivação para a acção, o segundo projecto tem no seu horizonte, precisamente, questões metodológicas relativas à relação entre filosofia e ciência cognitiva – é essa a justificação do seu sub-título «da ciência cognitiva à filosofia». O campo da ciência cognitiva é muito diverso e os debates filosóficos da ciência cognitiva vão desde questões tais como o inatismo, a modularidade e a natureza das representações, que poderão ter relação com os nossos interesses de investigação, até questões relativas à modelização da cognição, como as relativas ao conexionismo, que pelo menos neste momento estão mais afastadas. Para delimitar um território para o nosso trabalho deixamo-nos guiar pelos problemas da mente, linguagem e acção que neste momento nos ocupam. O nosso interesse metodológico genérico pelas relações entre filosofia e ciência cognitiva concretizou-se assim no objectivo de procurar clarificar o estatuto da teoria da mente no âmbito de uma concepção de epistemologia naturalizada, e entretanto clarificar a própria concepção de uma epistemologia naturalizada. É obviamente possível entender a ideia de epistemologia naturalizada como envolvendo a expulsão de questões normativas. Por isso nos pareceu especialmente interessante cruzar esta questão com a nossa abordagem dos fenómenos de racionalidade e irracionalidade. Será razoável defender que os estudos da racionalidade devem ser simplesmente entregues à ciência cognitiva, devendo a filosofia retirar-se do campo?

Tal objectivo conduziu-nos naturalmente a procurar explicitar as diferenças entre ciência cognitiva e filosofia, especialmente filosofia da mente, na abordagem da natureza do mental. Fomos desde logo obrigados a constatar que a filosofia da mente é entendida de formas muito diferentes pelos filósofos da mente contemporâneos. Na verdade, o simples facto de tentarmos identificar e comparar usos da ideia de epistemologia naturalizada conduziu-nos inevitavelmente às controvérsias que a concepção de filosofia da mente como disciplina

gera[35]. Assim, procurar compreender o que se faz quando se faz filosofia da mente foi uma (meta) questão constantemente em aberto, tratada inicialmente a partir dos três filósofos que escolhemos como orientação (Davidson, Fodor, Dennett), mas não se restringindo a eles (aliás, se algum destes filósofos continua a fazer convergir os interesses de todos os elementos do grupo, é neste momento unicamente Davidson[36]).

Mas a pretensão mais específica do Projecto acerca das relações entre filosofia e ciência cognitiva ainda não foi mencionada. Ela foi a seguinte: partimos da suposição de que existe um problema filosófico da racionalidade para além dos problemas cognitivos relevantes, tais como aqueles relativos a raciocínio e decisão. A ideia é que investigações, na ciência cognitiva, acerca de questões como raciocínio, decisão, emoções ou teorias da mente, são necessárias mas não suficientes para se ter uma teoria filosófica da racionalidade, tal como esta foi definida no objectivo programático de fundo que envolveu ambos os projectos. De acordo com esta definição, uma teoria filosófica da racionalidade envolveria (i) uma *descrição ou caracterização dos factores em jogo nas ocasiões em que agentes passam de determinadas crenças para outras crenças, adicionam ou eliminam crenças do seu corpo de crenças, ou optam, a partir de um conjunto de crenças e desejos, por um curso de acção por entre várias alternativas*, (ii) *um conjunto de hipóteses acerca da forma como decidimos entre critérios de correcção quando falamos da justificação ou racionalidade de crenças e acções*, (iii) *um conjunto de hipóteses acerca das razões por que queremos saber (se de facto queremos) se as nossas crenças são verdadeiras e os nossos raciocínios e acções racionais.*

Não se encontrará tudo isso apenas na ciência cognitiva. Na verdade na definição acima encontram-se algumas pistas para a diferença entre filosofia e ciência cognitiva que nos interessa, e que se relaciona com ideias como estas: (i) não basta evocar cânones de racionalidade como a lógica ou a teoria da decisão como modelos no estudo empírico do raciocínio e decisão, é preciso explicitar as razões por que se os evoca, e justificar a legitimidade de tal evocação, (ii) não basta descrever processos da racionalidade teórica e prática, é preciso compreender a natureza da prescrição envolvida quando afirmamos que se deve raciocinar ou agir de uma determinada maneira, (iii) descrições

[35] Miguens 2006c.
[36] Cf. Cadilha, neste volume.

da aplicação de regras e princípios não dizem ainda nada acerca da ligação entre subjectividade e normatividade.

Uma formulação do problema filosófico da racionalidade seria então a seguinte. Se tomarmos agentes humanos conscientes, que em determinadas ocasiões pensam e agem de acordo com determinados princípios, o problema é: por que é que nessas circunstâncias devem ser utilizados esses princípios? E por quê exactamente esses princípios? No projecto tratamos a questão nos termos de um 'critério de correcção' (tomei o termo de A. Goldman[37]): o que é necessário para responder a estas questões é um critério de correcção. Critérios de correcção explicitam as razões por que consideramos determinados cânones, ou standards, de racionalidade (regras, normas, princípios) como tal. Encontramos aqui alternativas tão diferentes como a evocação da natureza supostamente a priori do conhecimento lógico, a sustentação pública da normatividade em jogos de linguagem, ou o valor de sobrevivência de processos psicológicos que maximizam o número de crenças verdadeiras dos agentes (caracterizado por exemplo de uma perspectiva fiabilista). Podemos, na verdade, imaginar critérios de correcção alternativos para os standards de racionalidade. Como decidiremos então entre critérios de correcção alternativos? O próprio Goldman considera todas estas questões no âmbito de uma investigação conceptual da natureza da justificação e evoca o equilíbrio reflectido para responder à última questão acima. Mas mesmo que não concordemos com o carácter puramente conceptual de tal investigação, nem com a particular resposta de Goldman, somos levados a constatar que a investigação sobre a natureza da racionalidade, que poderíamos pensar ser relativa a processos cognitivos específicos, conduz a questões fundacionais acerca da natureza do pensamento e da linguagem e acerca da relação do pensamento como o mundo.

Encontramos em Robert Nozick[38] a segunda formulação do problema filosófico da racionalidade que nos guiou. Para Nozick, o problema filosófico da racionalidade diz respeito ao facto de certos agentes, nós próprios, não apenas usarem princípios para pensar e agir como terem que decidir que princípios *devem* utilizar para pensar e agir[39]. Formular o problema em termos de princípios conduz-nos a perguntar: Qual é a natureza desses princípios? Em que reside a sua força? O que nos compele a segui-los, se é que algo o faz?, e sobretudo

[37] Goldman 1986.
[38] Nozick 1993.
[39] Cf. Nozick 1993 e Bizarro 2003.

Que princípio de decisão deve ser utilizado para escolher os princípios a usar?

À primeira vista, pelo menos, configurar o problema filosófico da racionalidade como um problema acerca de princípios faz-nos ter em mente normatividade explícita. Nada nos impede, no entanto, de admitirmos a existência de normatividade prévia a esse estado explícito: podemos supor que numa acepção mínima de racionalidade instrumental as descrições de racionalidade valem para qualquer agente cognitivo, consciente ou não, e radicam na existência de razões-funções no mundo natural (Dennett chamar-lhes-ia *free floating rationales)*. O que entendemos por racionalidade deveria assim começar a ser considerado para compreendermos o seu lugar no mundo 'a partir de baixo', a partir da consideração dos benefícios, do ponto de vista evolutivo, do facto de agentes serem racionais, e não com a evocação de standards formais. Apenas posteriormente, em alguns agentes cognitivos, em alguns tipos de mentes, vieram a existir razões explícitas para a acção e a procura de razões independente de finalidade imediata – apareceu assim um cuidado com razões, por exemplo com a qualidade de raciocínios e decisões, uma procura de standards explícitos de racionalidade, que, na expressão de Nozick[40] '*now floats free*'. Procurar compreender o que está em causa na passagem de normatividade não explícita à normatividade explícita é outra das questões que nos ocupa.

Voltando à questão geral da diferença entre a filosofia e a ciência cognitiva quando se trata de questões de racionalidade, para além de termos procurado especificar as formulações do problema filosófico da racionalidade em termos de critérios de correcção e de justificação dos princípios, fomos levados ainda a concluir que não fazia sentido procurar compreender fenómenos de racionalidade e irracionalidade independentemente de uma teoria geral do pensamento, linguagem e acção. Essa pareceu-nos também uma incumbência da filosofia, sobretudo tanto quanto uma tal teoria deveria considerar questões relativas à perspectiva da primeira pessoa (chamemos-lhe eu, vontade livre, subjectividade, o que desejarmos[41]), e ao entendimento. Afirmei anteriormente que começámos por tomar como referências as obras de Davidson, Dennett e Fodor, e a verdade é que o trabalho em torno de Davidosn se tornou particularmente importante para nós aqui, à medida que procurávamos lidar com (i) a natureza da perspectiva de primeira

[40] Nozick 1993.

[41] No contexto da racionalidade prática John Searle fala, de forma expressiva, em '*the gap*' (Searle 2001, Capítulo 3).

pessoa, (ii) a subjectividade como entendimento, (iii) a ligação da subjetividade com a normatividade[42].

Este direcionamento relaciona-se em parte com a importância da consciência no que entendemos por pensamento, e pode ser defendido independentemente da ideia segundo a qual investigações filosóficas são exclusivamente aprioristas. Na verdade, o tipo de trabalhos a que nos propusemos nos Projectos assumem implicitamente que as investigações filosóficas não devem limitar-se a uma metodologia apriorista. Em contrapartida, é certo que nem toda a investigação em ciência cognitiva tem que ser, nem é, filosoficamente interessante. No entanto, não há como evitar constatar que certas investigações em curso em áreas da ciência cognitiva tais como por exemplo a psicologia do desenvolvimento e a psicologia evolutiva são, pelo seu próprio teor, enormemente relevantes para filósofos interessados em certo tipo de problemas, nomeadamente aqueles que dizem respeito à natureza da mente, da linguagem e da acção. É por exemplo esse o caso, pensando no presente projecto, da investigação sobre decisão, emoções, atenção conjunta, etc.

Torna-se imperativo compreender a importância, do ponto de vista filosófico, de tal investigação, e esta foi outra questão teórico-metodológica para nós. Penso que a referida importância tem a ver com o seguinte[43]: aqueles que vêm a ser, no tipo de agente racional que nós somos, os mecanismos da mentalidade, resultam da contingência histórica, evolutiva, de uma determinada interacção com o mundo. Assim sendo, como nota Charles Travis[44] esses mecanismos são *species specific*, i.e. não são característicos do pensador qualquer mas sim de pensadores humanos, resultantes de um processo de evolução específico. Os nossos mecanismos mentais são, para usar a expressão de Travis, 'paroquiais'; os seus produtos não são produtos do 'pensador qualquer', mas de um tipo específico de pensador. Os mecanismos

[42] Miguens 2005c e Miguens, artigo no presente volume.

[43] Aqui devo agradecer a Charles Travis as ideias iluminadoras que ao longo das entrevistas que têm acompanhado o Projecto RBD2 provocou em mim. O teor das entrevistas (cf. Miguens 2005ª e Miguens, no prelo) mostra que a questão foi abordada a partir da discussão de Wittgenstein e de Frege, com frequentes evocações (em grande medida críticas) de Davidson, McDowell e Fodor. Na prática, este foi o gérmen para o desenvolvimento, no Projecto, de uma dimensão de história da filosofia contemporânea que excede os filósofos da mente e da linguagem que começamos por apontar como guias, por se ter tornado evidente que as questões que tínhamos colocado acerca de agentes racionais se prolongam em questões acerca da relação entre o pensamento e o mundo e acerca natureza da experiência.

[44] Miguens 2005a.

mentais eles próprios, tal como são neste mundo, são o objecto das ciências da cognição; a singularidade, o facto de os mecanismos mentais serem, para usar a expressão de Travis, 'paroquiais'; o facto de os seus produtos não serem pensamento do 'pensador qualquer', mas de um tipo específico de pensador, tem implicações epistemológicas, metafísicas e éticas importantes e são estas que devem ser consideradas pela filosofia. Esta é uma pista essencial para a compreensão da diferença das estratégias de abordagem da filosofia e da ciência cognitiva sobre objectos que são aparentemente os mesmos, tais como a mente e a racionalidade. À filosofia cabe procurar compreender tal singularidade como singularidade, por oposição à suposta 'racionalidade do pensamento qualquer e do pensador qualquer'.

3. Esboço de algumas respostas.

Ao longo dos trabalhos dos Projectos fomos esboçando várias respostas para o conjunto de problemas de uma teoria filosófica da racionalidade. Algumas delas foram sendo avançadas no que até aqui afirmei. Recapitulo, para terminar, apenas alguns pontos, razoavelmente consensuais naquilo que fizemos.

Partimos do princípio de que para enfrentar o problema filosófico da racionalidade não basta identificar e evocar standards de racionalidade – é preciso procurar critérios de correcção e avançar razões para escolher por entre as alternativas. Isso leva-nos em última análise até questões gerais acerca da relação entre pensamento e realidade.

As questões 'Por que é que queremos ter crenças verdadeiras?' e 'Por que é que queremos agir e pensar de forma racional?' aparecem de forma muito diferente quando são formuladas genericamente acerca de agentes cognitivos – aparecerão aí como plausíveis hipóteses relativas a evolução e sobrevivência – e quando são formuladas 'a partir de dentro'. 'A partir de dentro' não se tratará, pelo menos num certo plano, exactamente de algo que o agente *quer*, mas antes do facto de o agente se encontrar constituído, 'desenhado', para ser de uma certa maneira (instrumentalmente racional, maximizador das crenças verdadeiras, etc). Isto justifica uma repetição da questão: será que de facto queremos ter crenças verdadeiras e pensar e agir de forma racional? S. Stich, do seu ponto de vista pragmatista, diria aqui que enquanto agentes cognitivos não queremos saber de verdade e racionalidade, na medida

em que essas não são as finalidades da acção[45], e é das finalidades da acção que nós queremos saber. Temos portanto que procurar saber o que determina tais finalidades – o que em última análise nos faz querer, e também que conceber a nossa relação, enquanto agentes que pensam em si próprios de uma determinada maneira, com tal querer.

É aqui que entram as questões acerca do aspecto volitivo da nossa subjectividade, que nos ocupam já desde o primeiro projecto. Qual será então a natureza dos nossos desejos? Serão determinações brutas do tipo de ser que somos? Poderemos de alguma forma apoderar-nos e dominar aquilo que desejamos? O que nos faz querer? O que nos move à acção? Seremos obrigados a escolher, ao procurar identificar o que nos move à acção, entre desejos humeanos e princípios kantianos? E como devemos conceber o papel das emoções aqui? Como é que alguma coisa vem a ter valor para alguém?

Podemos até certo ponto de alguma forma apoderar-nos e dominar aquilo que desejamos, e o mero facto de podermos defender que existem razões para acções e que as possuímos e usamos para causar acções vai nesse sentido. É duvidoso no entanto que a nossa situação seja a de um desaparecimento kantiano dos desejos dando lugar ao exercício da vontade racional pura.

Mas poderemos mesmo saber aquilo que de facto desejamos? Na medida em que aquilo que desejamos não apenas tem que ser (re) conhecido por nós como se relaciona com aquilo que pensamos que somos, a questão recai sobre o tema do auto-conhecimento. Não podemos contorná-lo, e diversas posições têm que ser consideradas. O problema envolve entre outras coisas a unidade das nossas mentes, e recai sobre a forma como pensamos na relação entre crenças, desejos, emoções, etc (que unidade existe, se é que existe, quem manda em quê? Por exemplo, eu não quero querer aquilo que quero – qual destes 'quereres' sou eu?)[46]

Mas será que crenças, desejos e intenções são os únicos motores da acção? Não deveremos considerar que as emoções e outros processos não cognitivos, tais como as protoemoções, produzem comportamento e participam no processo de tomada de decisão? Alguns dos trabalhos do projecto assentaram na tese de que as explicações cognitivistas das emoções não dão conta da importância destes processos não cognitivos, e como tal não explicam convenientemente o

[45] No presente volume Tomás Magalhães Carneiro analisa a aplicação desta ideia a uma teoria das emoções.
[46] O problema é em parte tratado em Amen, no presente volume.

modo como agimos e por que decidimos agir[47].

Será que a atribuição de racionalidade e irracionalidade está totalmente em nosso poder, ou o próprio acto de considerar algo como mental tem como condição uma atribuição de racionalidade? De acordo com as teorias interpretativas do mental, o caso é este último, o que assumirá a forma de algum argumento a favor da impossibilidade de irracionalidade, tornando a irracionalidade um fenómeno dentro da racionalidade e esvaziando de sentido a ideia de 'refutar empiricamente a racionalidade'.

Uma das ideias básicas do projecto foi reportar a racionalidade a agentes reais e às finalidades e afazeres destes no mundo. Nestas circunstâncias, o que é que princípios formais nos dizem acerca de processos de racionalidade tais como raciocínios e decisões? Qual é, por exemplo, a relação entre o raciocínio como processo de transformação de representações num agente e os cânones formais que supostamente se lhe aplicam, e cuja proveniência é a lógica? Qual é a relação entre a acção e decisão, e os princípios da teoria da decisão? Princípios de teorias formais fornecem-nos modelos para pensar sobre esses processos – a necessidade de modelos para considerar um âmbito de fenómenos é a norma em ciência –, mas avançar mais do que isso é arriscado. É aqui que se tornam pertinentes por exemplo discussões em história e filosofia da lógica – e estas podem inclusivamente conduzir-nos a notar que a própria concepção do que é a lógica e do que se faz em lógica permite imagens bastante diferentes do raciocínio e do pensamento[48].

4. Os artigos

Passo a descrever brevemente os artigos incluídos no presente volume e que constituem alguns dos resultados dos trabalhos de investigação dos membros do MLAG no último ano. Os artigos resultam dos trabalhos que os seus autores têm em curso, em muitos casos teses de mestrado e doutoramento. Como se verificará, eles dão voz a diferentes posições perante os problemas até aqui referidos, manifestando inclusivamente divergências fundamentais.

No primeiro artigo, *Por que não pode existir uma acção irracional,* Carlos Mauro e Susana Cadilha, defendem a tese segundo a

[47] Cf. Carneiro, no presente volume.
[48] Cf. Pinto, no presente volume.

qual não pode haver contradição no momento da acção, ou o agente não agirá de todo. A tese é, naturalmente, polémica, e por ocasião do colóquio, conduziu a uma discussão em torno de noções tais como 'desejo revelado', 'crença revelada', diferença entre intenção prévia e intenção na acção, crenças ociosas, e 'racionalidade como conceito aplicado a um instante'. Em geral, a tese deflaciona a (suposta) importância da racionalidade por exemplo na produção de normas morais.

José P. Maçorano, em *L. Floridi e a filosofia da informação,* expõe os conceitos nucleares e as principais consequências epistemológicas e metafísicas da filosofia da informação desenvolvida por L. Floridi (Universidade de Oxford).

Compara, nomeadamente, os conceitos de informação e de dados (*data*), sublinhando as diferenças ontológicas entre eles. Dadas tais diferenças, defende que as propostas de Floridi não são compatíveis com as concepções de conhecimento que se baseiam em representações mentais. De resto, como o autor explica, Floridi defende a ideia de um conhecimento construído pela informação, entendendo a informação enquanto dados estruturados segundo uma sintaxe e uma semântica bem definidas. De um ponto de vista filosófico mais geral, são exploradas as consequências relativistas e pragmatistas das teses de Floridi.

Em *Crença, triangulações e atenção conjunta,* Sofia Miguens continua a exploração do pensamento de D. Davidson com vista a uma teoria geral do pensamento, da linguagem e da acção que possa fundamentar e sistematizar o tratamento dos temas específicos do Projecto. Está em causa a última fase da obra de Davidson, na qual este desenvolve um conjunto de teses em torno da ideia de triangulação, que trazem modificações à anterior concepção de interpretação radical. O foco geral é desta vez a intersubjectividade e a importância desta na possibilidade de pensamento objectivo. A questão específica é a aplicabilidade da noção de crença a outros agentes que não os humanos, ou, mais em geral, a aplicabilidade do conceito de crença a mentes pré-conceptuais e pré-linguísticas. O artigo pretende ainda explorar as relações entre ciência cognitiva e filosofia no que respeita à intersubjectividade, já que aquilo que é tratado por Davidson sob o título de triangulação é também objecto de estudos empíricos, sob o título de atenção conjunta.

Continuando com o estudo do pensamento de Davidson, segue-se o texto de Susana Cadilha, *A teoria da acção de Donald Davidson e o problema da causação mental*. O artigo é uma análise crítica de alguns

aspectos da filosofia de Donald Davidson, nomeadamente da sua teoria da acção e da proposta ontológica com ela intimamente relacionada. O principal problema é o da causação mental – a autora discute a possibilidade de defendê-la no interior do esquema davidsoniano.

Em *Emoções e racionalidade derivada* Tomás Magalhães Carneiro procura investigar o estatuto a atribuir ao background não cognitivo numa teoria filosófica da racionalidade. Procura ainda saber como será possível encontrar um critério normativo de racionalidade. O autor discute nomeadamente as consequências dos resultados da psicologia evolucionista para o campo das teorias da racionalidade, sobretudo os trabalhos sobre emoções e racionalidade das emoções. Discute ainda propostas específicas de S. Stich e J. Searle quanto ao estatuto da racionalidade. A sua ideia fundamental é que a racionalidade das proto-emoções é derivada de formas de intencionalidade superiores.

Em *Davidson on Irrationality and Division*, Miguel Amen enfrenta directamente o tratamento que Davidson faz da questão da irracionalidade, defendendo-o de críticas de J. Heil (Divided Minds). Davidson defende que para compreendermos a irracionalidade devemos postular uma mente dividida, enquanto Heil apresenta objecções a essa tese, aparentemente simples e directa. A ideia de Heil é que mesmo que uma mente dividida fosse suficiente para explicar a irracionalidade, ela não é necessária (é mesmo supérflua). Miguel Ámen defende Davidson das críticas que lhe são dirigidas, procurando corrigir a interpretação que Heil dele faz.

A lógica é um dos possíveis cânones da racionalidade. Mas de forma alguma é legítimo considerar a lógica como oferecendo-nos sem mais as regras que o raciocínio deve seguir – antes de o fazermos conviria procurar esclarecer a natureza da lógica. O trabalho em história e filosofia da lógica constitui uma importante via para um tal esclarecimento. O artigo de João Alberto Pinto, *Boole e Frege: matematização da lógica vs. logificação* situa-se aí. Em termos gerais, estão em discussão no seu trabalho concepções de lógica como linguagem e concepções de lógica como cálculo, articuláveis com diferentes concepções sobre o mental que lhes podem estar subjacentes.

Resta-me esperar que a leitura dos artigos, e a consideração dos temas aqui propostos, bastante diversos, como perspectivas sobre a questão filosófica da racionalidade seja esclarecedora e frutuosa.

Rationality, Belief, Desire: a research programme

Sofia Miguens[1]

> «Some animals think and reason; they consider, test and reject hypothesis; they act on reasons, sometimes after deliberating, imagining consequences and weighing probabilities, they have desires, hopes and hates, sometimes for good reasons. They also make errors in calculation, act against their own best judgment, or accept doctrines on inadequate evidence. Any one of these accomplishments, activities, actions, or errors is enough to show that such an animal is a rational animal, for to be a rational animal just is to have propositional attitudes, no matter how confused, contradictory, absurd, unjustified, or erroneous these attitudes may be. This, I propose is the answer. The question is: what animals are rational?» Donald Davidson, *Rational Animals*[2]

1. The nature of Project *Rationality, Belief, Desire II – from cognitive science to philosophy* and its motivations. Assigning rationality and irrationality.

The articles collected here result from the research project *Rationality, Belief, Desire II – from cognitive science to philosophy* (POCI/FIL/55555/2004) and should be regarded as explorations of the issues into which the Project branches[3]. At the centre of our interests lie the various aspects of rationality. We take rationality to be a trait of cognitive agents. Cognitive agents are representation-guided systems, characterized by some goal-structure, which behave in an adapted way in their environment[4]. We assume that in order to consider them as such

---

[1]Principal Investigator of *Rationality, Belief, Desire II – from cognitive science to philosophy* (POCI/FIL/55555/2004)

[2] Davidson 2001: 95.

[3] The introductory book *Racionalidade* (Miguens 2004) was used as a guiding plan for the development of the current project (cf. questions for a philosophical theory of rationality, pp. 19-45)

[4] This definition is a starting point, and we are very much aware of the fact that it glosses over cognitive science discussions about the status of representations and goals. Our sole justification for this starting point is the level of the processes we are mostly interested in here, which are higher cognitive processes. J. P. Maçorano's article in this volume, though, goes some way into exploring the problem of the nature of representation, as well as the issue of relations between information and what we, from a mentalistic point of view, call belief.

it is not strictly necessary to evoke conscious awareness from the start. This way of approaching agents is common in cognitive science. In philosophy, by contrast, approaches to specific issues of practical rationality (rationality of decisions, rationality in action) and theoretical rationality (rationality of reasoning, rationality in the process of fixing and revising beliefs), often take conscious awareness of agents for granted. Leaving consciousness aside for the moment, it is important to notice that (i) talk of *agents* is necessary for formulating any questions concerning rationality, and (ii) taking certain parts of the world to be agents means taking them to have beliefs and desires, and thus describing them by means of *mentalistic language*. From the viewpoint of philosophy of mind – although such questions naturally extend to epistemology and metaphysics – the nature of such states and such language is an open question. Thus the Project's philosophy of mind background, which is expressed in its title by the purpose of jointly dealing with the nature of *rationality* and with the nature of *beliefs and desires*.

We believe it is very important, in order to understand rationality, to try to understand phenomena of both practical and theoretical *irrationality*. In fact, the wish to weigh accusations of irrationality addressed to agents, pieces of reasoning, decisions, institutions, under various circumstances, was a fundamental motivation for the Project. What is even more interesting with these accusations is the fact that they are very often followed by an appeal to Reason, which is supposed to make it possible to overcome the former flawed situation (as for instance when, in an argument, people don't understand each other, also when one tries to ground the status of laws, in juridical and moral contexts, or when one considers the progress of societies, or the quality of political decisions). It is in fact all too frequent to proclaim rationality or irrationality without feeling the need to know what is involved therein. But the truth is, it is easier to say what a valid argument is then to say what a justified belief, an appropriate decision or a rational creature are. It seems that we all want to be rational, in the sense that no one wants to be regarded as irrational: we do not want to hold unjustified beliefs, we do not want to be bad at reasoning and deciding, we do not want to act counter to our own best judgment. But why is that so? What does it have to do with our way of being human and with the kind of minds we are? Because the fact is, although we apparently wish to be rational, we admit that very often that does not seem to be the case – humans are certainly capable of acting against

their own best judgment[5], of believing things they have no reason to believe, of not believing that which they have reason to believe, of believing in contradictions, and so on. Again, why is that so? Is it that when it happens that we believe what we should not believe, or do what we think we should not do, this happens just as a matter of ignorance, as when we are unaware of a particular rule which would apply at a certain point in solving a given problem, or do such facts tell us more about the way our minds work?

In order to face these problems we have to have a picture of the volitional aspect of our nature. In the Project, the following questions were intended to conceptually take us apart as agents for that purpose, and have been the object of constant attention: What are desires? What are intentions? What are emotions? What is it that ultimately motivates us into acting? Are we psychological egoists, always motivated by self-interest? Is it that only self-interested action can lead to any form of satisfaction or happiness? What is the origin of selfishness? Can it, or should it, be in any way overcome in moral and rational action? Is it possible to know what we want without knowing what we, ourselves, are? How is our will, or our willing, structured, and how does such a structure relate to our self-representations as agents? Where does motivation come from on those occasions where humans seem to have less regard for self-interest than for other people's, or for society's, needs (need for justice, for instance)? Is a moral action necessarily and ultimately non-selfish or is it the case that for there to be motivation there simply has to be selfishness? How do selfishness and the emotions relate? Are emotions simply irrational? Is it really the case that in agents such as ourselves reasons can cause actions? Beyond those definitions available and commonly used[6], what is, after all, rationality in action?

These are in fact questions our first rationality Project (*Rationality, Belief, Desire – motivation for action from the viewpoint of the theory of mind*, 2003-2005) already dealt with. Project RBD1 had

[5] This is, of course, objectionable, and not every member of our group believes it is even possible (cf. Mauro & Cadilha).

[6] Namely (i) the instrumental definition (a rational agent is capable of recruiting the means appropriate to achieve the ends she pursues – of course, for that, ends have to somehow already be there, in the agent, prior to any action) and (ii) the idea of acting so as to 'maximize expected utility' (again, agents preferences, considered in deliberation, should be somehow previously defined and stable). In both cases, agents have beliefs and desires which enable them to consider things in terms of means/ends, utility, probability, etc.

the question of motivation for action as its main focus. Within that project we dealt with issues such as instrumental conceptions of rationality[7], the belief-desire model for the explanation of action[8], the nature of reasons for acting, the relation between reason and passions in more or less rationalistic conceptions of the will and of rational action[9], the specific nature of mental states such as intentions in contrast with beliefs and desires[10], philosophical theories of emotions (focusing especially on the cognitivism / non-cognitivism debate)[11], psychological egoism as, ultimately, the core of rational choice theory, psychological-philosophical foundations of economics, inasmuch as these involve rational choices of agents[12], etc. The first Project had thus a strong component of theory of action and of moral philosophy[13] and that is still the case in project RBD2.

The other basic motivation for the rationality Projects, besides the general interest in irrationality phenomena and in accusations of irrationality also concerned a specific theoretical need for assigning rationality[14]. It was a motivation of a more technical nature: we were interested in the problems faced by interpretation theories in the philosophy of mind. In fact, theories of mind such as those developed

[7] Madeira 2003ª.
[8] Madeira 2003a.
[9] Miguens 2003.
[10] Madeira 2003b.
[11] Mendonça 2004.
[12] Cf. Carlos Mauro, PhD dissertation. Although this work came to extend itself to other questions, it started by considering the philosophical-psychological foundations of economics, especially the concept of psychological egoism as the core of rational choice theory. Psychological egoism is the idea according to which people are always motivated by personal interest. Ultimately, this means that the rational agent acts always in virtue of self-interest, aiming at the maximization of expected utility.
[13] This was a consequence of the development of the project. From the start, we intended to have a broad perspective of the extensive literature on theoretical and practical rationality. Another objective was to find answers to the following questions, in ways which would orient future research. The questions were: (i) what motivates an agent into acting? (ii) what does a philosophical theory of rationality look like? what kinds of issues does it deal with? Here we have taken as references the works of S. Stich, A. Goldman, R. Nozick and S. Blackburn. Once these initial steps were taken, the research interest of the members of the group have naturally become more specific: some members of MLAG are currently interested mostly in philosophy of action and moral philosophy, others in questions concerning the nature of logic, still others in philosophy of mind (in topics ranging from interpretation theories, to theories of mind, emotions, mental causation, identity theories, etc.)
[14] We should speak not only of rationality assignments but also about arguments in favour of the impossibility of irrationality.

by W. V. Quine, D. Davidson and D. Dennett, have as their starting point a rationality assignment[15]. Here, the origin of the rationality Projects goes all the way back to my work on D. Dennett's Intentional Systems Theory[16]. Intentional Systems Theory – I'm using it as a general label for the various theses D. Dennett's theory of mind includes – is a Quinean theory of interpretation, and is committed to assigning to an agent, by default, beliefs which are mostly true, and inferences which are mostly rational. In the absence of such an assignment, it is simply not possible to take certain parts of the world as minds. In the specific case of Intentional Systems Theory, assigning rationality is a condition for the theories of representation, of consciousness, of action, and of personhood. So the whole structure stands or falls depending on the legitimacy and coherence of such a starting point. In the above mentioned work, I tried to explore the conditions and the implications of such an assignment of rationality. In particular, I was then interested in understanding whether that was done aprioristically, as seems to be the case in another, more well known, interpretation theory of mind, that of Donald Davidson. The answer was negative. Before saying why, and since Davidson's position is of central interest to us, I will consider it first. In his article *Could There Be a Science of Rationality*[17], Davidson defends the following thesis about the status of his theory of thought, language and action: «The entire theory is built on the norms of rationality; it is these norms that suggested the theory and give it the structure it has. But this much is built into the formal, axiomatizable parts of decision theory and truth theory, and they are as precise and clear as any formal theory of physics. However, norms or considerations of rationality also enter with the application of the theory to actual agents, at the stage where an interpreter assigns his own sentences to capture the contents of another's thoughts and utterances. The process necessarily involves deciding which pattern of assignments makes the other intelligible (not intelligent, of course!) and this is a matter of using one's own standards of rationality to calibrate the thoughts of the other. In some ways this is like fitting a curve to a set of points, which is done in the best of sciences. But there is an additional element in the psychological case: in

[15] The origins of this idea can be found in the way W. V. Quine considers the charity principle within radical interpretation (cf. *Word and Object*, 1960).
[16] Globally assessing Intentional Systems Theory as a set of positions on philosophy of mind issues is the purpose of Miguens 2002.
[17] Davidson 2004a.

physics there is a mind at work making as much sense as possible of a subject matter that is being treated as brainless, in the psychological case, there is a brain at each end. Norms are being employed as the standard of norms.»[18].

Davidson's idea is thus that traits assigned in interpretation are determined by formal theories, and that happens prior to any actual interpretation of another being. We can see circularity here, we can also consider that Davidson presents an argument for the impossibility of irrationality of what is to count as a mind. Before we shun such circularity, maybe we should stop to consider that things could not look very different if rationality constraints are built into the structure of mind and language – it is not as if we could step back and look at that condition from the outside and then describe it. As John Searle puts it, «we may intelligibly debate theories of rationality, not rationality.»[19]

Anyway, what could be an alternative to this aprioristic view? An alternative would be, for instance, to consider that it is the design of agents that gives the rationality assignment grounding, and that such design is a result of evolution by natural selection. Rationality simply is to cognition what adaptation is to life; both are cases of function and adaptation. In fact this is the idea behind Dennett's rejection of any characterization of Intentional Systems Theory as instrumentalist. He doesn't see his interpretation theory as instrumentalist because he thinks that what the interpreter does is not to project rationality but rather to recognize existing patterns, resulting from real design of agents[20]. Explaining design is a task for the theory of evolution by natural selection, which, in case we are considering mind-design, means reporting theory of mind to sub-personal considerations about agents. This line of thought involves a second argument for the impossibility of irrationality of agents, besides the davidsonian aprioristic argument[21].

In general, if arguments for the impossibility of irrationality stand, we should conceive of irrationality as a phenomenon within rationality, a phenomenon for which one should find a place.[22] As Davidson puts it, «The sort of irrationality that makes conceptual trouble is not the failure

---

[18] Davidson 2004a: 130.

[19] Searle 2001: xiv.

[20] For an overview, cf. Miguens 2006b.

[21] Interestingly, the issue of impossibility of irrationality comes back under another guise in the most recent work of Project members. Cf. Mauro & Cadilha, present volume, defending the thesis of impossibility of irrationality in the moment of action.

[22] Cf. Amen, in present volume.

of someone else to believe or feel or do what we deem to be reasonable, but rather the failure, within a single person, of coherence or consistency in the pattern of beliefs, attitudes, emotions, intentions and actions»[23]. Yet, it is not easy to avoid the temptation to regard as irrationality in reasoning or in the decision-making of others what *we* ourselves take to be irrational – in fact a great deal of empirical literature on the subject confronts this problem. Ultimately, the question is whether rationality and irrationality – to use J. Cohen's formulation[24] – can be empirically demonstrated. In other words, does it make any sense to assume that in empirical studies of rationality we start with no presuppositions at all about the rationality or irrationality of agents, and then find out through experience whether specific agents are rational or irrational? Or there is something wrong with this way of looking at things?

Some empirical studies of reasoning and decision have been followed by the conclusion that irrationality has been demonstrated. This would be supported by finding out that actual reasoning and decision do not conform to ideal standards, such as those of logics, probability theory, or decision theory. People simply do not think and decide by following such principles. They tend rather to use heuristic principles which simplify situations and are in general effective but also lead to persistent biases (this was one of the main points of the classic book by A. Tversky, Slovic & D. Kahneman[25]).

More recently, some authors have tried to avoid such conclusions about persistent irrationality[26] – that is namely the case in evolutionary psychology studies in which the results of irrationality test-cases well known in the literature, such as Linda the bankteller and the Wason selection task, are reinterpreted. Persisting biases are seen as resulting from adaptive characteristics of cognitive devices, which because they are adaptive should not be considered irrational.

Are we to conclude then that no skill, mechanism or component of human minds can, inasmuch as it is looked upon from the point of view of evolution, be considered irrational? This would be too strong a conclusion[27]. I think there is a question prior to the interpretation of the

[23] Davidson 2004 a: 170.
[24] Cohen 1981.
[25] Kahneman, Slovic & Tversky 1982.
[26] Cosmides & Tooby 1996, Barkow, Cosmides & Tooby 1992. Cf. for a summary Miguens 2004, pp. 84-88.
[27] Trying to make clear what is at stake when standards of rationality are evoked as models for empirical research, Samuels, Stich & Tremoulet 2003 consider the possibility

results of empirical research, and this is in fact where Cohen's thesis makes sense. According to this thesis, irrationality cannot be empirically demonstrated because normative principles such as those of logic or probability theory must not be considered as natural science hypotheses, which may be tested, confirmed or rejected. They are rather the very framework for the approach. In fact, this thesis rejoins the positions in philosophy of mind I have mentioned above, positions according to which we cannot consider anything as irrational without presupposing rationality – if there is such a thing as irrationality is should be regarded as a phenomenon within rationality. The problem is, it is obviously not clear which rationality we are talking about here, since it cannot be identified with the usual formal standards. It is not clear either how rational an agent must be to be taken to be a mind – too demanding constraints, such as having a perfect ability for calculation or a consistent web of beliefs, seem impossible to sustain.

I have claimed, based on Dennett's work and on reasons related to theory of cognition, that the concept of rationality as it is used in interpretation theories cannot be a deep or precisely defined one[28]. It is rather an agent-level notion, behaviourally based, tied to instrumental means-ends relations. It applies to the agent as a whole and is not even based on real representations, taken to be natural kinds (this would be a Fodorian view of what makes for the rationality of real agents). Such a concept of rationality may be indispensable to the theory of mind, but it is certainly not possible to simply identify it with standards of rationality such as those logics, probability theory and decision theory provide us with. That is why rationality in this sense is not liable to any precise characterization: It is a pragmatic notion, which should, thus, not be considered a label for some kind of intrinsic cognitive value. With this view[29], I did not intend to refuse or in any way deflate the problems of theoretical and practical rationality (problems concerning

of a Chomskyan style competence. If that is the case, questions of cognitive architecture and modularity should be considered. In between pessimistic theses (we are irrational) and overoptimistic ones (supposedly irrational performances are justified by the evolutionary history of the species), Samuels, Stich & Tremoulet try, in *Rethinking Rationality*, a text which was quite often discussed in Project's meetings, to defend an intermediate way. Stich & Sripada try to show how from an evolutionist point of view, evoked by some to keep accusations of irrationality away, it is still possible to make sense of the persistency in agents of currently irrational devices. I am especially thankful to Tomás Carneiro for studying and discussing these questions, introducing them to the members of the group, as well as for the translations of the texts.

[28] Miguens 2002: 510

[29] Miguens 2002.

what we should believe and what we should do), but rather point out that we are not justified in evoking rationality as some kind of secure and well-known ground, the kind of thing we hope for when we give it the last word in questions of thought and action. The minimal conception of rationality which the theory of mind needs and which is good for dealing with any cognitive agent is thus not to be identified with much more sophisticated and specific notions such as belief consistency, deductive closure or perfect inferential capacity. The problems of practical and theoretical rationality remain untouched by such a thesis, which is of a different level. So, rationality as a concept in use in interpretation theory of mind is not perfect rationality. This thesis must be explored, but anyway, it does not seem possible to enumerate aprioristically a set of true beliefs and inference principles without which we would not call an agent 'rational'. In his pragmatist theory of rationality S. Stich formulates this by saying that *there is no way to formulate a priori constraints for every possible rational agent*[30].

Interpretation theories are, within the field of philosophy of mind, and given the role rationality plays in them, those we have been most interested in. Naturally, the status of rationality assignments is only one of the aspects which might make one doubt whether it is possible to sustain such theories. Interpretation theories are anti-reductionist[31], and there is a general question whether anti-reductionism is ultimately coherent. In theories which take themselves to be physicalist (that is at least the case with Dennett's) the unexplained residue which is the interpreter is a big problem: what is the interpreter? Where does the rationality assigned by the interpreter come from? J. Fodor[32], who is himself an anti-reductionist, but one whose anti-reductionism has a totally different form, blames Dennett's approach for being 'transcendental' (Davidson is often accused of the same sin). For Fodor that is definitely no compliment; it is rather related with an absence of explanation.

For Fodor, admitting of real representations is the first step which makes talk of rationality of agents possible: representations are more fundamental than rationality and are in fact the ground for explaining

[30] Stich 1993.

[31] Davidson is very direct when identifying reasons for anti-reductionism: these are (i) normativity of interpretation, (ii) causal character of mental concepts such as action, (iii) externalism.

[32] Cf. Miguens 2005, J. Fodor e os problemas da filosofia da mente. Fodor, Davidson and Dennett were our first references where it came to the question How should one go about doing theory of mind?

rationality from a cognitive point of view. There where Fodorian intentional realism places real mental representations, making a subpersonal explanation of the rationality of agents possible, taking it to concern computations of representations, a transcendental theory of the mind places nothing. This is not an easy problem: under the guise of a discussion about the starting point for the theory of mind what is at stake here is ultimately how naturalism or physicalism on the one hand, and normativity and subjectivity on the other could possibly stand together[33].

Besides the general question regarding the place of mind in a physical world, the question of rationality also led us to a question about the specific type of minds which are human minds and the specific kind of doings which are human actions – the approach to the nature of mind and action one defends is inevitably reflected in something which we, as humans, should care a lot about: a conception of what it is to be human. In the work I have been referring to[34], I raised objections to the underlying intellectualism of Dennett's theories of consciousness and personhood. An ongoing discussion in the current Project concerns a similar intellectualism in Davidson: Davidson bluntly states in «Rational Animals»[35] that small babies, like snails, cannot justifiably be considered rational creatures – only creatures capable of having concepts of belief and truth are capable of objective thought and thus deserve such title[36].

2. Methodological issues: cognitive science and philosophy.

The two motivations for the Rationality Projects identified above, both related to the status of rationality assignments, led us into the fields of philosophy of action, moral philosophy and philosophy of mind. A few methodological considerations are now needed, in order to understand how the Project extended to other fields. Inasmuch as RBD2 Project was itself conceived as having a focus on methodology, they have theoretical relevance as well. While the first Project centred on the

[33] The fact that this is the strating point of McDowell's Mind and World is one of the reasons why we were led to this author.
[34] Miguens 2002, Capítulo 4.
[35] Davidson 2001.
[36] I believe there is something wrong with Dennett's and Davidson's intellectualism concerning these issues – part of what is wrong has to do with not considering perception in theory of mind. Within the project – cf. interviews with Charles Travis – that's what some of us have been working on.

question of motivation for action, the second Project was conceived as having a focus on methodological questions regarding the relations of philosophy with cognitive science – that is the justification for the sub-heading «from cognitive science to philosophy». The field of cognitive science is very diverse, and debates in the philosophy of cognitive science range from issues such as nativism, modularity and the nature of representations, which do relate to our current interests, to others which are not so directly related to them, for instance those concerning connectionism. We used our philosophy of mind framework to delimit our concerns and decided that our general purpose, in considering the relations of philosophy and cognitive science, should be to understand the status of theory of mind in a framework of naturalized epistemology. Along the way, we intended to try to make clear what one means by naturalized epistemology. A possible understanding of naturalized epistemology is, of course, that epistemology should simply drop all normative questions. That's why we thought it would be especially interesting, since our interests were focused on the normative phenomena of rationality and irrationality, to try and see where naturalized epistemology leads. Is it the case that the study of rationality should simply be handed over to cognitive science?

To deal with such an issue we also had to try to make clear what makes cognitive science and philosophy, especially philosophy of mind, approaches to the mind different from one another, if indeed they are. We were fully aware from the start of the fact that contemporary philosophers understand the field of philosophy of mind in very different ways. Actually, simply trying to identify and compare uses of the idea of naturalized epistemology in philosophy of mind led us into controversies which the very nature of philosophy of mind as a discipline provokes. We took as starting point and as guidance the philosophies of Dennett, Fodor and Davidson, but did not restrict ourselves to their works. What we tried to do was to take them as offering concrete answers (and different ones) to the question 'How does one go about doing theory of mind'. Anyway, at present maybe only Davidson is still common ground and common interest to the members of the group[37].

I haven't yet mentioned our main contention about the relations between philosophy and cognitive science concerning rationality. We assume there is a philosophical problem of rationality, beyond the

[37] Cf. Cadilha, in the present volume.

related cognitive problems (such as those concerning reasoning, decision-making, etc). We think that cognitive science research about questions such as reasoning, decision, emotions, theories of mind, is an essential contribution for a theory of rationality. Still, we think cognitive science is not sufficient to answer all questions we identified in the guidelines of the project. We think that a theory of rationality should include (i) a description or characterization of the factors at play on occasions when agents move from certain beliefs to others, add or eliminate beliefs from their corpus of beliefs, or opt for a course of action from several alternatives, based on a set of beliefs and desires; (ii) a set of hypotheses about the way we decide about rightness criteria when we talk of justifiedness or rationality of beliefs and actions; (iii) a set of hypotheses about the reasons why we want to know (if indeed we do) if our beliefs are true and our reasoning and actions rational. And those are things that won't be found in cognitive science alone.

Some clues to the difference between philosophy's and cognitive science's approaches would then be, for instance, that (i) standards of rationality such as those of logic and decision theory may provide us with models in the study of processes of reasoning and decision – but it is still necessary to say why they apply, if indeed they do, (ii) simply describing processes of reasoning and decision-making is not enough to understand the nature of the prescription involved, (iii) descriptions of application of rules and principles do not yet say anything about the connection between subjectivity and normativity.

One formulation of the philosophical problem of rationality would thus be the following. If we take consciously aware human agents as they think and act, in certain occasions, according to certain principles, the problem is: why should such principles be used in such circumstances? And why those principles exactly? In the project we dealt with this question in terms of criteria of rightness (a term we took from A. Goldman[38]), and we took it that what was needed here to answer such questions was a criterium of rightness. Criteria of rightness should make the reasons for which we take certain rules, norms or principles, to be standards of rationality explicit. Here we were led to alternatives as different as the supposedly apriori nature of logical knowledge, the public nature of language-games based normativity, or the survival value of psychological processes which maximize the number of true beliefs of agents (characterized, for instance, within a

[38] Goldman, 1986.

reliabilist theory of epistemic justification). But if there are several candidates to criteria of rightness how are we to decide for one? And what is it that we are doing when we are involved in such decisions? Goldman himself deals with these questions in the context of a conceptual analysis of justification – but even if we do not agree with that approach, what is important here is that we clearly see that research on rationality inevitably touches foundational questions about the nature of thought and language, and, ultimately, the relation of thought to the world.

Another formulation of the philosophical problem of rationality we took as reference is due to Robert Nozick[39]: there is a philosophical problem of rationality because there are certain agents – ourselves – who not only use principles to think and to act, but also should decide which principles they should use to think and to act[40]. This formulation forces us to ask questions such as: what is the nature of such principles? Where lies their power? What makes us follow them? And above all, what decision principle should be used to decide about which principles to use?

Questions regarding principles are often formulated having explicit normativity and conscious agents in mind. Yet, Nozick himself considers the existence of a kind of normativity prior to that state; inasmuch as there are descriptions of rationality which are good for any agent, consciously aware or not, descriptions related to *free floating rationales,* there are ways things are supposed to be. If we turn our eyes in that direction, if we decide that what we take to be rationality should be considered from this 'bottom-up perspective', in order to understand its place in the world, and not by evoking formal standards, we will end up looking upon explicit normativity as something which only later, in some cognitive agents, in some kinds of minds, came to exist. And once it was there, it involved a search for reasons independent of an immediate aim – a care for reasons, for the quality of reasoning and decision-making that, in Nozick's term (1993), '*now floats free*' and should be explained as such.

Going back to the general question of the difference between philosophy and cognitive science when it comes to dealing with rationality, besides having tried to formulate it in terms of the relation between standards of rationality and criteria of rightness, and in terms

[39] Nozick 1993.
[40] Cf. Nozick 1993 and Bizarro 2003.

of a decision principle for the use of principles, it also soon became clear that we could not go about trying to understand phenomena of rationality and of irrationality independently of a general theory of thought, language and action. We saw that as a specifically philosophical task, which involved considering first person perspective (whether we call it self, subject, will[41]) and understanding. I said before that we initially took as guidance the work of three authors, J. Fodor, D. Dennet, e D. Davidson. Working on Davidson became quite important for us here, as we tried to deal with questions such as (i) the nature of first person perspective, (ii) subjectivity as understanding, as well as the linguistic nature of such an understanding, and (iii) the connection of subjectivity and normativity[42].

One thing should be clear: the need for this focus on the nature of first person perspective can be defended independently of the idea according to which philosophical investigations should remain exclusively aprioristic. In fact, the kind of approach we favoured in the rationality projects implicitly states that we do not think philosophical investigations should stick to an aprioristic methodology. It is certainly the case that not all cognitive science research is philosophically interesting, nor should it be. Yet, one can not fail to notice that certain kinds of research, such as those of development psychology and evolutionary psychology, are, by their very nature, quite relevant for philosophers who have an interest in certain kinds of problems, namely problems concerning the nature of mind, language and action. In this Project, that is the case of research on decision, emotions, joint attention, etc. It is important to try to formulate what constitutes such relevance, from a philosophical point of view and I believe it has to do with the following[43]: those that come to be, in the kind of rational agent that we are, the mechanisms of mind and rationality, result from evolution, and thus from the historical contingency of a certain kind of

[41] J. Searle calls it in a very expressive way, in the context of his theory of practical rationality, 'the gap' (Searle 2001).

[42] Cf. Miguens 2005 and Miguens in the present volume.

[43] I must thank Charles Travis for the illuminating ideas which he has provoked in me through the interviews which have accompanied the development of the project. In these interviews it becomes clear that this has been done starting from the discussion of Wittgenstein and Frege, and frequently considering and criticizing Davidson, McDowell and Fodor. As a result, a new dimension of 'history of contemporary philosophy' of thought mind and language has grown, largely exceeding the authors we started with; this happened in fact as it became clear that the questions about rational agents that we were posing could not be dealt with while avoiding other more general questions about thought and world and the nature of experience.

interaction with a world. They are, as Charles Travis[44] likes to put it, *species specific*, that is, they are not characteristic of any thinker whatsoever but rather of a certain type of thinkers, the humans, as result of a specific process of evolution. Our 'mental ways' are thus, in that sense, parochial and their products should not be looked upon as the product of just any thinker but as the product of a specific type of thinker. This singular character has epistemological, metaphysical and ethical implications, and provides us with yet another clue for understanding the different strategies of approach of philosophy and cognitive science to mind and rationality: working it out as such is a philosophical task.

3. Sketching some answers.

Throughout the Project's activities we tried to develop answers to the questions for a philosophical theory of rationality that were mentioned above. I have already said something about that – now I will just recapitulate some points.

We took it that facing the philosophical problem of rationality means not only identifying standards of rationality but also going beyond them, looking for criteria of rightness, or for the justification of the principles, and working out reasons to choose between alternatives there. In doing that, we kept in mind that the questions 'Why is it that we want to have true beliefs?' and 'Why is it that we want to think and act on a rational way?' may lead to different answers when they are (i) posed about cognitive agents in general (this is where hypotheses about evolution and survival are plausible) and (ii) when they are formulated from within the agent. From within we cannot exactly say about agents that they want to have true beliefs; it is rather that they find themselves having been endowed with a certain cognitive design and acting in a certain way (instrumentally rational, maybe maximizing true beliefs, etc…). One should then repeat the question: is it really the case that we want to have true beliefs and to think and act rationally? From his pragmatist point of view S. Stich would say that as cognitive agents we do not care for the truth of our beliefs or the rationality of processes, in that those are not our aims, our ends[45], and that – our aims, our ends, what we want – is what we care about.

[44] Miguens 2005.

[45] In his contribution to this volume Tomás Magalhães Carneiro analyses the way this idea applies to a theory of emotions.

This is where questions about the volitional aspect of our nature enter – where do our aims come from? What is the nature of our desires? Are they brute determinations of the kind of beings we are? Is it in any way possible to take over and control that which we desire? What is it that makes us desire or want? Must we choose, when identifying that which moves us into action, between humean desires and kantian principles? How are we to conceive the role of emotions here? How is exactly that something comes to be of any value to us? We are capable, to a certain extent, of taking over and controlling that which we find ourselves desiring and the mere fact that it is possible to claim that there are reasons for actions, and that we are in possession of them, and that they cause actions, goes in that direction. It is doubtful though that it is ever the case that desires step away, in a kantian way, to make it possible for rationality in action to be 'pure reason'.

But is it really possible to know what we do in fact desire? In order to answer such question, we have to understand not only how is it that what we desire can be known (or recognized) by us, but also how it relates to what we think we are, to our self-representations as agents. This brings in the question of self-knowledge, and here we must consider different views – anyway the problem concerns the unity of the mind of an agent, and this bears on the way we think about how beliefs, desires, intentions and emotions stand together in a mind (what kind of unity is there (if indeed there is unity)?, what rules upon what? when I, for instance, want something that I do not want to want – is this wanting still me, or not?).[46]

Should we ever think that beliefs, desires and intentions are the sole intervening factors in action? Shouldn't we consider that emotions and other non cognitive processes, such as proto-emotions, have effect upon behaviour and mind, and have a role in the decision processes? At least some lines of work within the project were based on the conviction that cognitivist theories of emotions do not account for the importance of such non-cognitive processes, and so do not appropriately account for the way we decide and act[47].

The answers to the questions above must make sense within a general view of mind and thought, and there some decisions must be taken about how explicit a level we are dealing with. Is the attribution of rationality and irrationality totally within our power, or the very act

[46] The problem is partly dealt with in the article by Amen, in this volume.
[47] Carneiro, in the present volume.

of taking a mind as a mind – which we, being the kind of being we are – do anyway – has as its condition rationality assignment? According to interpretation theories, this is the case, and so irrationality is a phenomenon within rationality, and it simply does not make sense to try to empirically refute rationality.

Throughout the whole project, our basic idea was to think of rationality as a characteristic of real cognitive agents and their doings in the world, not taking for granted that we are already in possession of standards of rationality, but rather asking what, in such circumstances, do formal principles tell us about rationality processes. Our idea was that they provide us with models to think about such processes – and the need for models of phenomena is certainly common in science – but any extra step here, trying to claim more, would be risky. That is why discussions in the history and philosophy of logic could become very interesting – especially because they show us that logics may be conceived in different ways which correspond to different places and status for reasoning and thought.[48].

4. The articles.

I will now briefly summarize the contents of the articles included in this book, which result from ongoing research of the members of the Mind, Language and Action Group (MLAG) in the last year. As may be realized, by reading the articles, they stem from different positions regarding the problems identified above. One might even say that they express fundamental divergences among the members of the group, and indeed we take that to be a good thing.

In the first article, the contribution of Carlos Mauro e Susana Cadilha, *Why there cannot be an irrational action*, the authors defend that there cannot be contradiction in the moment of action, or the agent will not act at all. This is naturally, quite controversial, and much discussion followed the presentation of the talk, centering on notions such as revealed desire, revealed belief, differences between previous intention and intention in action, otiose beliefs, and the restriction of the application of the concept of rationality to an instant of action. Also, the authors general intention of deflating the importance of rationality in the production of moral norms was much debated and opposed by some

[48] Cf. Pinto, in this volume.

of the people present.

In his article, *L. Floridi and the philosophy of information,* J.P. Maçorano analyses the core concepts and main epistemological and metaphysical implications of the approach to the philosophy of information developed by Oxford philosopher Luciano Floridi. More specifically, he contrasts the concepts of information and data, pointing out the ontological differences between the two. Given such differences, Floridi's proposals are not compatible with a representation-based conception of knowledge. As J.P. Maçorano explains, Floridi defends an information-based concept of knowledge, taking information as data structured according to defined syntax and semantics. From a more general point of view, the relativist and pragmatist consequences of this position are explored.

In *Crença, triangulações e atenção conjunta,* Sofia Miguens continues previous work on Davidson's philosophy. As before it is assumed that only a general theory of the nature of thought, language and action may ground, and render systematic, the treatment of specific issues of the project. In Davidson's work it is possible to find such a theory. In this article, Davidson's views on triangulation, expressed in his last writings, are considered and assessed. These views change the former conception of radical interpretation, and bring intersubjectivity to bear on objective thought. The specific problem dealt with is the use of the concept of belief when considering non-linguistic agents. The article is also an attempt to explore the relations between philosophy and cognitive science concerning a specific problem, since Davidson's 'triangulation' is the object of empirical studies under the name of 'joint attention'.

Still in the context of the studies of the work of Davidson, Susana Cadilha, in *A teoria da acção de Donald Davidson e o problema da causação mental*, critically analyses some aspects of his philosophy, namely his theory of action and the ontology connected with it. Given the project's focus in agents and action, she focuses on mental causation, trying to make the implications of Davidson's position clear.

In his article, *Emoções e racionalidade derivada* Tomás Magalhães Carneiro considers the status of the non-cognitive background of agents in a philosophical theory of rationality, and also the possibility of finding a normative criterion of rationality for dealing with such issues. He discusses the implications of evolutionary psychology results – especially the work on emotions and on the rationality of emotion – for rationality theories. He considers the

specific proposals of S. Stich on the status of rationality, and also J. Searle's work on intentionality and the background. His main contention is that the rationality of proto-emotions is derived from higher forms of intentionality (involving conscious awareness).

In his paper *Davidson on Irrationality and Division*, Miguel Amen deals directly with Davidson's approach of irrationality. He defends Davidson from J. Heil's criticisms in Divided Minds. Davidson claims that in order to understand irrationality we should postulate a divided mind, while Heil poses objections to such a prima facie simple and direct claim. Even if a divided mind were sufficient to explain irrationality, it is not necessary (in fact is even superfluous). Miguel Amen defends Davidson from criticisms, while he also tries to correct Heil's interpretation of his theory.

Logics provides us with candidates to standards of rationality, namely for rationality in reasoning. Still it is not in any way legitimate to think of logics as simply revealing the rules any reasoning should follow (in fact, it is not even that simple to put forward a definition of reasoning, as something in contrast with transformations of information in a cognitive system). Before any claims are made about the relations between logics and reasoning, it would be helpful to make clear issues regarding the nature of logics itself. Work in the history and philosophy of logics is an important way in for that. That's what João Alberto Pinto's article, *Boole e Frege: matematização da lógica vs. logificação* is concerned with. He contrasts conceptions of logics as language and conceptions of logics as calculus, which can be articulated with different places for mind and reasoning .

I hope that reading these articles, which offer quite different perspectives and approaches to the philosophical problems of rationality, will prove to be illuminating and enriching.

Referências / References:

Ámen, Miguel, "Davidson on Irrationality and division", in Miguens S. & Mauro (coords), *Perspectives on Rationality*, Porto, Faculdade de Letras da Universidade do Porto.

Barkow, J., Cosmides, L. & Tooby, J. 1992, *The Adapted Mind: Evolutionary*

Psychology and the generation of culture, Oxford, Oxford University Press.

Bizarro, Sara, 2003, "Robert Nozick e a natureza da racionalidade", *Intelectu* nº 9.

Cadilha, Susana., "A teoria da acção de Donald Davidson e o problema da causação mental" in Miguens S. & Mauro (coords), *Perspectives on Rationality*, Porto, Faculdade de Letras da Universidade do Porto.

Caló, Susana, 2006, *A natureza histórica da cognição: debates filosóficos na teoria dos sistemas dinâmicos na ciência cognitiva*, Dissertação de Mestrado, Porto, Faculdade de Letras da Universidade do Porto.

Caló, Susana, 2004, "Do corpo e das crenças à acção: o mundo autista", *Trólei* nº 4

Carneiro, Tomás, 2005, "Para acabar de vez com o cognitivismo", *Intelectu* nº 11

Carneiro, Tomás, "Emoções e racionalidade derivada", in Miguens S. & Mauro (coords), *Perspectives on Rationality*, Porto, Faculdade de Letras da Universidade do Porto.

Cohen, Jonathan, 1981, "Can irrationality be empirically demonstrated?" *Behavioral and Brain Sciences*, 4, 317-370.

Cosmides, L. & Tooby, J., 1992, "Cognitive adaptations for social Exchange", in Barkow, J., Cosmides, L. & Tooby, J. 1992, *The Adapted Mind: Evolutionary Psychology and the generation of culture*, Oxford, Oxford University Press, 163-228.

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

Davidson, Donald, 2001, "Rational Animals", in *Subjective, Intersubjective, Objective*, Oxford, Oxford University Press.

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

Davidson, Donald, 2004, "Could There Be a Science of Rationality", in *Problems of Rationality*, Oxford, Oxford University Press.

Galvão, Pedro, 2004, "Teoria da decisão, racionalidade e ética: o utilitarismo de Harsanyi",*Trólei* nº 4

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

Goldman, Alvin, 1986, *Epistemology and Cognition*, Cambridge MA, Harvard

University Press.

Kahneman, D., Slovic, P. & Tversky, A. (eds.), 1982, *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge, Cambridge University Press.

Maçorano, José Pedro, "L. Floridi e a filosofia da informação" in Miguens S. & Mauro (coords), *Perspectives on Rationality*, Porto, Faculdade de Letras da Universidade do Porto.

Madeira, Pedro, 2003, "O que é a teoria instrumental da razão prática?", *Intelectu* nº 9.

Madeira, Pedro, 2003, "O que é o modelo crença-desejo?" *Intelectu* nº 9.

Madeira, Pedro, 2003, "A objecção de Nagel ao modelo crença desejo", *Intelectu* nº 9.

Madeira, Pedro, 2004, "Razão instrumental e razões para agir", *Trólei* nº 4

Mauro, Carlos, *A impossibilidade de irracionalidade na acção,* Dissertação de Doutoramento, em curso.

Mauro, Carlos, 2005, Somos ou não inevitavelmente racionais? *Intelectu* nº 11

Mauro, C. & Cadilha, S. "Por que não pode existir uma acção irracional" Miguens S. & Mauro (coords), *Perspectives on Rationality*, Porto, Faculdade de Letras da Universidade do Porto.

Mendonça, Dina, 2004, "Experience as reason: emotions and values in the construction of rationality", *Trólei* nº4.

Miguens, Sofia, 2002, *Uma Teoria Fisicalista do Conteúdo e da Consciência*, Porto, Campo das Letras.

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

Miguens, Sofia, 2005a, "O que tem a filosofia a dizer à psicologia?" Entrevista a Charles Travis, *Intelectu* nº 11.

Miguens, Sofia, 2005b, "J. Fodor e os problemas da filosofia da mente", *Intelectu* nº 11.

Miguens, Sofia, 2005c, "O problema do auto-conhecimento", *Intelectu* nº 11.

Miguens, Sofia, 2006a, "Why Can't There Be a Science of Rationality – D. Davidson and Cognitive Science", in Miguens, S., Pinto, J.A. & Mauro, C.E., 2006, *Análises / Analyses*, Porto, Faculdade de Letras da Universidade do Porto.

Miguens, Sofia, 2006b, "D. Dennett's brand of anti-representationalism", *Protosociology* 22.

Miguens, Sofia, 2006c, "Os problemas da filosofia da mente", *Diacrítica*.

Miguens, Sofia, no prelo, Charles Travis – entrevista nº 2 (Wittgenstein e Frege, pensamento e linguagem).

Miguens, Sofia, "Conceito de crença, triangulações e atenção conjunta", Miguens S. & Mauro (coords), *Perspectives on Rationality*, Porto, Faculdade de Letras da Universidade do Porto.

Morando, Clara, 2005, "Teorias da mentalidade: uma aproximação filosófica", *Intelectu* nº 11.

Nozick, Robert, 1993, *The Nature of Rationality*, Princeton NJ, Princeton University Press.

Pinto, J.A, "Boole e Frege: matematização da lógica *vs.* logificação", in Miguens S. & Mauro (coords), *Perspectives on Rationality*, Porto, Faculdade de Letras da Universidade do Porto.

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

Samuels, R., Stich, S. e Tremoulet, P., 2003, "Repensando a racionalidade: de implicações pessimistas a módulos darwinianos", tradução portuguesa de Tomás Magalhães Carneiro, *Intelectu* nº 9.

Searle, John, 2001, *Rationality in Action*, Cambridge MA, MIT Press.

Stich, Stephen, 1990, *The Fragmentation of Reason*, Cambridge MA, MIT Press.

Stich, S. & Sripada, C., 2005, "Racionalidade, evolução e emoções", tradução portuguesa de Tomás Magalhães Carneiro, *Intelectu* nº 11.

Travis, Charles, 2006, *Thought's Footing*, Oxford, Oxford University Press.

Por que não pode existir uma acção irracional[*]

Carlos E. E. Mauro[1]
Susana Cadilha[2]

Resumo: O âmbito geral deste artigo é o problema da explicação da acção. Nesse contexto, a nossa proposta é a de que não é possível atribuir irracionalidade a nenhuma acção de nenhum agente real. Os argumentos centrais são: 1º) argumento de carácter prático-descritivo – a acção tem uma razão/causa específica formada por uma crença e um desejo apenas – isto significa que existem apenas *uma crença na acção* e *um desejo na acção*; 2º) um desejo ou uma crença só poderão ser considerados determinantes se não estiverem ociosos; 3º) a intenção prévia não se confunde com a intenção na acção – estão dissociadas; 4º) não é possível definir *a priori* a racionalidade de uma acção.

Abstract: The general scope of this paper is the problem of the explanation of action. In this context, we defend that is impossible to attribute irrationality to any action of any real agent. The steps are: first, a descriptive argument – there is only one determinant belief and one determinant desire to the agent in the moment of action; *action* differs from *deliberation; s*econd, any belief or desire can only be considered determinant to the agent if they are not idle; third, we dissociate *prior intention* and *intention in action;* fourth, the rationality of an action cannot be a priori determined.

[*] Agradecemos os comentários dos amigos do MLAG: Sofia Miguens, Tomás Carneiro, José P. Maçorano, Miguel Ámen; de Juan Vázquez Sanchez da Universidade de Santiago de Compostela; e de André Barata da Universidade da Beira Interior. É importante ressaltar que uma versão anterior e diferente foi lida no colóquio Estatuto do Singular (RICI – Universidade Nova de Lisboa) em Maio de 2006 e será publicada nas actas do referido colóquio em meados de 2007.

[1] Membro e investigador do *Mind Language and Action Group* – MLAG – do Instituto de Filosofia da Universidade do Porto. Bolseiro de doutoramento da FCT.

[2] Membro e investigadora do *Mind Language and Action Group* – MLAG – do Instituto de Filosofia da Universidade do Porto. Bolseira de doutoramento da FCT.

I

A teoria da acção debate-se, classicamente, com dois tipos de problemas. O primeiro de entre eles é de carácter metafísico e prende-se com a natureza da acção. Consideremos alguns eventos que ocorrem no mundo – o que distingue uma acção de um mero movimento físico ou de algo que simplesmente nos acontece? O segundo problema diz respeito à explicação da acção. É neste ponto que a nossa análise incidirá.

A tese aqui proposta é a de que não é possível imputar irracionalidade a qualquer acção de qualquer agente real. Os argumentos centrais são: 1º) argumento de carácter prático-descritivo – a acção tem uma razão/causa específica formada por uma crença e um desejo apenas – isto significa que existem apenas *uma crença na acção* e *um desejo na acção*; 2º) um desejo ou uma crença só poderão ser considerados determinantes se não estiverem ociosos[3]; 3º) a intenção prévia não se confunde com a intenção na acção – estão dissociadas; 4º) não é possível definir *a priori* a racionalidade de uma acção.

II

Tendo em vista o nosso problema – a questão da explicação da acção – assumimos a pertinência do chamado modelo crença-desejo, e é desse mesmo pressuposto que partimos em direcção aos nossos argumentos.

O modelo crença-desejo sustenta que qualquer acção de qualquer agente real é necessariamente produzida por crenças e desejos, (isto é, por um desejo com o qual estaria relacionada uma crença[4]). Para explicar uma acção de acordo com o modelo crença-desejo, atribuem-se, portanto, certos estados mentais ao agente, tais como crenças e desejos específicos. São eles que constituem a razão pela qual o agente levou a cabo uma determinada acção. Dizer que alguém faz alguma

---

[3] Isto parece óbvio; no entanto, uma grande confusão conceptual pode ocorrer quando não distinguimos o processo de deliberação/decisão da acção propriamente dita – mais à frente, na secção III, trataremos desta distinção. O conceito "desejo ocioso" aparece aqui no sentido de Anscombe § 36 – "a característica principal de um desejo ocioso consiste no facto de o homem não fazer nada, possa ou não, para o cumprimento do seu desejo". G. E. Anscombe, *Intention*. 2nd. Edtion. Cambridge, Harvard Press, 2000.

[4] Há discussão quanto ao peso que é devido a cada um destes elementos na questão da explicação da acção – se o desejo deve ou não ser considerado o factor decisivo (cf. Pedro Madeira, "A objecção de Nagel ao modelo crença-desejo (e o realismo moral)", in *Intelectu* nº9 – www.intelectu.com)

coisa por alguma razão é dizer que estão aí implícitos algum tipo de desejo e algum tipo de crença do indivíduo – quaisquer que tenham sido as causas da formação de tais desejos e crenças. A esse par crença-desejo Davidson denomina de "razão primária"[5]. Trata-se, então, de uma teoria psicologista da acção. [6]

Desse modelo costuma também fazer parte a premissa segundo a qual o par crença-desejo que explica a acção constitui a causa dessa acção. Aqui se encerra o aceso debate entre wittgensteinianos e seus oponentes – em jogo está a questão de discernir se razões podem ou não ser causas[7]. Ainda que possamos concordar com essa premissa, ela não vai estar aqui em discussão. Assumiremos então, seguindo Davidson, que: a) uma acção é um evento intencional porque tem uma descrição intencional (o que distingue uma acção é o facto de poder ser descrita recorrendo a crenças e desejos); b) uma acção é explicada por uma crença e por um desejo; c) razões são causas.

III

Tendo como base este modelo explicativo, a tese que pretendemos defender é a de que (*a*) não é possível definir como irracional qualquer acção de qualquer agente real. Como consequência, parece que teríamos também que defender que (*b*) não é possível definir como racional qualquer acção de qualquer agente real. Facto que nos levaria a algum tipo de (*c*) relativismo acerca da racionalidade. Tentaremos demonstrar que (*a*) é verdadeiro, que (*b*) é falso e que (*c*) não faz sentido.

De acordo com a definição usual, uma acção irracional – caso da akrasia, por exemplo – seria aquela que manifestasse inconsistência relativamente a certas crenças ou desejos do agente (uma inconsistência interna relativamente às crenças/desejos do agente – *inner inconsistency* – e não uma inconsistência que se estabelece a partir de fora, relativamente a padrões externos).

Mais concretamente, seria um caso em que o agente age contrariamente às razões que ele próprio julga serem as mais relevantes no que toca à acção em questão. Ele opta por um determinado curso de

---

[5] Davidson é, pois, um dos mais conhecidos defensores do modelo crença-desejo na teoria da acção.

[6] Por oposição às teorias não-psicologistas, que sustêm que não são as nossas crenças e desejos que impulsionam as nossas acções.

[7] O que remete para a questão da causação mental.

acção, por lhe parecer o mais razoável, mas acaba por agir num sentido diferente, que até pode ser o oposto.

Num tal caso, estaria a ser violado o chamado princípio de continência, segundo o qual o agente deve agir de acordo com o seu melhor juízo, depois de considerados e avaliados todos os aspectos em jogo[8]. Há irracionalidade porque a razão que o leva a agir em sentido oposto suplanta o próprio princípio de continência – essa razão não é uma razão contra o princípio em si (não o põe em causa), mas é usada como tal; há irracionalidade porque o agente não atribui às considerações relevantes que estão em jogo o peso que elas de facto teriam (ou esse peso não é suficiente para o incitar a agir).

Com o intuito de tentar mostrar como uma tal acusação de irracionalidade é inadequada, consideramos ser necessário tornar claros todos os passos que conduzem o agente até ao momento da acção propriamente dito. Explicar uma acção é uma tarefa que implica a explicitação de todo um processo que envolve, pensamos nós, diferentes momentos que devem ser cuidadosamente distinguidos.

De que modo surge uma acção? No movimento contínuo que conduz até à acção, sustentamos ser fundamental a distinção entre o processo de deliberação (seja essa deliberação declarada ou não – isto é, quer o agente declare verbal e explicitamente qual será a sua acção, quer seja uma tomada de decisão "interior", "silenciosa", não formulada) e o momento em que a acção é realizada – ou seja, a acção propriamente dita. Se no *processo de deliberação* que conduz à decisão estão em jogo muitas crenças e desejos, até mesmo desejos e crenças contraditórios, no momento em que a *acção* é despoletada, pelo seu lado, apenas um par particular composto por uma determinada crença e um determinado desejo são relevantes, pelo que a acção não pode ser considerada irracional. Isto porque esse tal par forma-se sempre no próprio agente. É importante ressaltar que a acção ocorre porque há um agente e porque há uma intenção específica, seja ela consciente ou não. O agente é um elemento cuja coerência poderia ser avaliada, provavelmente, num nível muito complexo, onde as crenças e os desejos fariam sentido quando introduzíssemos num modelo explicativo variáveis como: propensões inatas, possíveis conteúdos inatos, cultura, sociedade, história de vida, entre outras. [9]

Concretizando, por meio de um exemplo banal – um atleta julga

---

[8] Cf. Donald Davidson, "Paradoxes of Irrationality", in *Problems of Rationality*, Oxford, Oxford University Press, 2004.

que, dadas as suas circunstâncias, o melhor que tem a fazer no feriado que se avizinha é treinar para a próxima prova. Contudo, acaba por passar o dia com os amigos a passear. Agiu, portanto, contrariamente ao seu melhor juízo, e a sua acção, por definição, é tida como irracional, ou acrática. Isto porque parece existir uma incoerência entre a acção realizada e as crenças ou desejos do agente (nomeadamente a crença de que tem melhores razões para treinar do que para sair com os amigos, e a própria crença de que deve seguir aquela que lhe parece a melhor alternativa de acção possível, conjuntamente com o desejo de ganhar a prova). O que sustentamos é que não é possível dizer que existe tal incoerência, porque no exacto momento em que ele age, só uma crença e desejo são determinantes – o desejo de se divertir[10] e a crença de que tal será possível escolhendo esse curso de acção. Não há qualquer contradição no próprio momento da acção, senão ele não agiria de todo. Nesse preciso momento, foram esse desejo e essa crença específicos que se sobrepuseram aos outros e despoletaram a acção – o agente não pode ser acusado de irracionalidade, nem tão pouco de "fraqueza de vontade", só porque esse par crença-desejo não coincide com o que anteriormente havia julgado ser o mais razoável. É garantidamente possível (e natural) que, no decorrer do processo de deliberação, diferentes desejos e crenças, mesmo que contraditórios entre si, estivessem presentes à consideração do agente; no entanto, se aquela determinada acção teve lugar, foi porque apenas um par crença-desejo específico se revelou importante para ele no exacto momento em que agiu.

Não acreditamos, portanto, que seja adequado declarar que uma tal acção é inconsistente, dadas as suas crenças e desejos.

Alguém poderia contra-argumentar sustentando que para que uma acção se concretize têm de estar disponíveis mais do que uma crença e um desejo; serão precisas, nomeadamente, crenças de suporte que tornem essa acção viável. Por exemplo, quando fazemos algo como carregar num interruptor para acender a luz, há um largo conjunto de crenças que dá sentido a essa acção – a crença, por exemplo, de que a casa tem electricidade e que ela está a funcionar, entre outras. O que nós sugerimos é que apesar de tais crenças latentes (de suporte) serem

[9] Na parte final do artigo, procuraremos explicar melhor esta ideia.

[10] Desejo que poderia ser realizado por meio de outras acções como, possivelmente: beber, ler, contemplar, conversar, beijar, ter relações sexuais. É importante entender que são outras acções e que cada uma terá o seu par crença-desejo.

necessárias para que o nosso acto de ligar a luz tenha sentido, não é a elas que a nossa acção se fica a dever. Não é o facto de eu acreditar que existe electricidade que me leva a ligar a luz. Se assim fosse, eu passaria a carregar em todos os interruptores que encontrasse apenas por possuir a crença de que a casa tem electricidade e que ela está a funcionar.

Essa é a pergunta fundamental – o que conduz à acção? E é a ela que pretendemos responder dizendo que só um par crença–desejo despoleta a acção propriamente dita, separando assim o momento da decisão do momento da acção. Nós ficamos a conhecer (ou não) qual a crença e o desejo relevantes para o agente no momento em que este actua, e não no momento em que declara a sua intenção. Isto porque o nosso julgamento é naturalmente carregado de considerações de vária ordem, de carácter moral, por exemplo – frequentemente, nós julgamos como melhor aquilo que é socialmente aceitável ou moralmente tido por correcto, daí que as nossas intenções (declaradas ou não) apontem nesse sentido, mas sem que sempre consideremos que seja esse de facto o caminho mais acertado. Nem sempre aquilo que dizemos ser o melhor a fazer é o que de facto queremos fazer, e mesmo quando nos parece que queremos de facto agir num certo sentido, nem sempre o fazemos necessariamente. E nem por isso podemos ser considerados agentes irracionais.

Tal ideia aparece directamente ligada ao segundo argumento que pretendemos trazer para a discussão – um desejo ou uma crença só poderão ser considerados determinantes se não estiverem ociosos. Podemos pautar-nos por inúmeros desejos ociosos – o desejo de ganhar a prova do nosso atleta, por exemplo – e ter à disposição várias crenças latentes (de suporte), mas que nenhuma relevância desempenham no momento da acção. Isto é, não adianta desejar obter uma boa colocação numa prova e ter a crença de que passando o dia a treinar conseguiremos isso, para que o indivíduo se decida pelo curso da acção coerente com tal desejo declarado. O indivíduo revelará no momento da acção o desejo e a crença determinantes, aos quais podemos chamar *desejo revelado* e *crença revelada*.[11]

[11] Os conceitos de *desejo e crença revelados* não significam que os desejos e crenças foram perfeita e completamente revelados. Esses conceitos determinam a diferença entre: a) os *desejos e as crenças presentes no processo de decisão* (que podem ter sido, ou não, declarados pelo agente); b) a *crença e o desejo na acção* (que só poderíamos conhecer se tivéssemos uma teoria perfeita e completa da acção – que não é o caso); c) a *crença e o desejo revelados* (aqueles aos quais podemos basear-nos para investigar a intenção do

Antes da acção ter lugar, tudo está a um nível teórico-abstracto, um nível em que "tudo" é possível, até crenças opostas/ con- traditórias; mas, na acção, o agente contará apenas com uma intenção, ou seja, apenas uma crença e um desejo, caso contrário não ocorrerá um evento intencional e sim um evento involuntário. Estamos aqui a distinguir, portanto, entre intenção prévia e intenção na acção.[12] Sustentamos que a intenção prévia não é determinante para a acção e não é algo que comprometa o curso de acção que escolhemos. Seja qual for a intenção prévia, é a intenção na acção que conta, pois é nesse ponto que as crenças e desejos se revelam.

O que faz, então, com que alguém se decida por um determinado curso de acção? Quais são os elementos relevantes? O facto de legítima e conscientemente acreditarmos que a nossa melhor opção é, por exemplo, treinar, não tem qualquer influência no resultado da nossa acção? Se só a crença e desejo revelados no momento da acção importam, isso significa que muitas das nossas crenças são ociosas, mesmo as que parecem mais evidentes[13]. Defenderemos que as evidências a que o agente pensa ter chegado são relevantes, mas não são suficientes para transformar uma crença (latente) ociosa numa crença na acção. Parece-nos mais plausível imputar desejabilidade às crenças de tal forma que uma crença será a *crença na acção* se for mais desejada do que as outras crenças em jogo. O caso do auto-engano é um exemplo flagrante disto mesmo que acabámos de defender – nesse caso, todas as evidências apontam num determinado sentido, mas o desejo do agente sobrepõe-se de tal forma que tais evidências são pura e simplesmente ignoradas.

Até porque, convém acrescentar, o processo de racionaliza- ção[14]

agente, seja ela consciente ou inconsciente).

[12] Cf. Donald Davidson, "Intending", in *Essays on Actions and Events*, Oxford, Oxford University Press, 1980.

[13] Para muitas pessoas as tais crenças mais evidentes são também crenças "melhores" porque levariam o agente a obter os melhores resultados. Discordamos desta ideia. Isto porque teríamos que definir o que entendemos por "os melhores resultados". Teríamos que afirmar que existem critérios naturais válidos para avaliar os resultados a partir das crenças e dos desejos de um agente específico. Ora, julgamos isto inadequado e impossível.

[14] Processo através do qual se encontra a razão pela qual o agente leva a cabo uma determinada acção. Cf. Donald Davidson, "Actions, Reasons and Causes", in *Essays on Actions and Events*, Oxford, Oxford University Press, 1980.

nem sempre é suficiente para que o agente conheça as razões das suas acções. Na sua formulação ideal, pela racionalização davidsoniana o agente é capaz de saber quais são as crenças e desejos que estão na base da sua acção, mas, de facto, nem sempre é esse o caso, na medida em que entram muitas vezes em jogo desejos inconscientes aos quais não temos acesso cognitivo/introspectivo. Esta tese vem apenas reforçar a ideia de que nem sempre os nossos desejos declarados são os nossos desejos mais prementes – estes só se revelam no momento da acção.

IV

O que pretendemos demonstrar com este artigo é que não é sustentável acusar um agente real de irracionalidade, apelando para uma possível inconsistência/contradição entre as suas crenças e desejos e a sua acção. Essa possibilidade de contradição existe em termos lógicos, mas não nos parece possível quando o que está em questão é um agente real. Em termos concretos, o que se passa é que, a um dado ponto, outras crenças (mesmo que erradas) e outros desejos (mesmo que imorais ou inadequados) adquiriram maior peso e tornaram-se os propulsores da acção. Quando um agente real (por oposição a uma entidade abstracta, que se regesse por princípios formais inquebrantáveis) vai contra o seu melhor juízo não se pode falar de contradição – o que pode haver é um revelar daquelas que poderiam ser as verdadeiras intenções do agente, ou uma mera mudança de planos. No exacto momento da acção, foi um particular desejo acompanhado por uma particular crença que se revelaram decisivos para o agente. Com efeito, se dois desejos contraditórios estivessem presentes no momento da acção, sem que a vontade do agente pendesse para um dos lados, então este não agiria de todo, ou tomaria uma decisão ao acaso, circunstância na qual não nos poderíamos referir a uma acção intencional e voluntária.

No instante em que eu opto por um curso de acção que não coincide com aquele que foi determinado pelo meu melhor julgamento, eu posso estar a agir em sentido contrário ao que é socialmente aceite, posso estar a ser irresponsável ou a revelar um comportamento imoral, mas não a ser irracional. O facto é que, para que a nossa acção seja considerada racional, não temos que nos pautar por aqueles desejos que, avaliados todos os dados em jogo numa determinada situação, concluímos serem os mais acertados mediante qualquer referência externa, seja ela moral ou não. Até porque podemos estar simplesmente a seguir aquilo que social e contextualmente é tido como o melhor, o

que não tem nada a ver com racionalidade. Ser racional não é seguir normas previamente instituídas, por nós ou pela sociedade – apelando para uma definição mais lata, tudo pode ser racional se tivermos razões para tal, e se agimos intencionalmente é porque tivemos essas razões. O que é importante sublinhar é que essas razões podem ser várias, mas podem não seguir nenhum princípio de coerência ou de constância, tal como o chamado princípio de continência. Não se pode normatizar/normalizar a acção humana, porque não se pode regularizar o ser humano de uma forma apriorística e infalível.

Talvez isto não nos permita afirmar que as acções são sempre racionais, sejam elas quais forem. Talvez possamos seguir um outro caminho – as acções não seriam nem irracionais nem racionais, pois é uma lógica diferente que rege a racionalidade prática, que não uma lógica bivalente. Ser racional no sentido em que nos temos vindo a referir é inerente ao ser humano. Não que ele tome sempre as decisões consideradas mais acertadas ou que representem o melhor resultado para o próprio ou para outrem, mas porque ele age com base em crenças e desejos muito próprios, específicos, quase que podemos dizer pessoais e intransmissíveis. E nesse sentido nunca poderemos dizer que as suas acções são irracionais – elas seguem uma lógica específica, envolvem uma análise custo-benefício que para o agente se revelou pertinente no momento em que age, dadas as suas particulares vontades e ambições. Mas dizer que uma acção não pode ser irracional, não é igual a afirmar que todas as acções são racionais (no sentido instrumental) – é, antes, sustentar que uma acção racional não pode ser definida à partida, com base em princípios ou normas abstractas, que a racionalidade deve ser considerada a um nível distinto, não como uma característica adjectiva de certas acções, mas como uma propriedade/qualidade substantiva da realidade humana. Analogamente, podemos dizer que a racionalidade é o "sangue" da acção intencional; isto é, da mesma maneira que se não houver sangue não haverá vida, se não houver racionalidade não haverá acção intencional.

Cada indivíduo age de acordo com razões/causas produzidas dentro de um sistema único, específico e pessoal. Isto significa que não conseguimos agir irracionalmente – podemos sim agir imoralmente, podemos produzir o pior resultado e não o melhor (individual ou colectivo), podemos agir ineficazmente, mas sem determinar com isto a irracionalidade na acção.

No entanto, é natural que se observe certa coerência entre acções (razões) ao longo do tempo. Isto ocorre graças: a) à habilidade/

necessidade adaptativa do Homem; b) pela força psicológica da cultura; c) a certos elementos inatos. Por que razão tendemos geralmente a seguir certos desejos ou impulsos, a procurar determinadas coisas em detrimento de outras, e a agir em consonância, manifestando uma certa regularidade nas nossas acções? Ou porque isso se revelou útil no decorrer do nosso processo evolutivo, e/ou porque somos marcados por certos desejos inconscientes que caracterizam a própria espécie em toda a sua extensão, e/ou porque socialmente nos sentimos pressionados a tal. O que quisemos salientar, porém, é que se estes factores não forem suficientes para determinar a nossa acção e dar-lhe essa coerência esperada, não temos, ainda assim, motivos para ser considerados agentes irracionais.

A Filosofia da Informação de Luciano Floridi : Pressupostos Epistemológicos[1]

José Pedro Maçorano[2]

Resumo: Neste artigo analiso os conceitos nucleares e as principais implicações epistemológicas e metafísicas da abordagem à filosofia da informação desenvolvida pelo filósofo de Oxford Luciano Floridi. Mais especificamente, ele compara os conceitos de informação e de dados, sublinhando as diferenças entre os dois. Dadas essas diferenças, as propostas de Floridi não são compatíveis com uma concepção de conhecimento baseado em representações. Procuro explicar o conceito de conhecimento baseado na informação que Floridi defende, tomando a informação como dados estruturados de acordo com uma sintaxe e uma semântica definidas. De um ponto de vista mais geral, procuro explicitar as consequências relativistas e pragmatistas da abordagem.

Abstract: In this article, I analyse the core concepts and main epistemological and metaphysical implications of the approach to the philosophy of information developed by Oxford philosopher Luciano Floridi. More specifically, he contrasts the concepts of information and data, pointing out the ontological differences between the two. Given such differences, Floridi's proposals are not compatible with a representation-based conception of knowledge. As J.P. Maçorano explains, Floridi defends an information-based concept of knowledge, taking information as data structured according to defined syntax and semantics. From a more general point of view, the relativist and pragmatist consequences of this approach are underlined.

[1] Trabalho realizado graças à bolsa de doutoramento da FCT.

[2] Membro do *Mind Language and Action Group* – MLAG – Bolseiro de doutoramento da FCT.

Introdução

A minha intenção com este artigo é revelar os pressupostos epistemológicos subjacentes à Filosofia da Informação, tal como Luciano Floridi a concebe. A primeira parte do artigo explicita estes pressupostos partindo do conceito de Informação, dado o papel central deste. A análise do conceito de informação é realizada através da contraposição do mesmo com o conceito de dados (*data*). Dadas as diferenças ontológicas existentes entre estes conceitos, defendo que as propostas de Floridi não são compatíveis com as concepções de conhecimento baseadas em representações mentais. Pelo contrário, Floridi assume a defesa de um conhecimento construído pela informação, entendendo a informação enquanto dados estruturados segundo uma sintaxe e uma semântica bem definidas. Defendo que esta posição de Floridi constitui uma crítica ao papel constitutivo dos dados como fontes da informação, sendo necessário caracterizar o papel epistemológico destes como recurso, utilizado na elaboração da informação.

A segunda parte do artigo procura explicitar os pressupostos epistemológicos inerentes às principais metodologias propostas por Floridi para o campo da Filosofia em geral, e para a Filosofia da Informação em particular. Estes métodos consistem no Método de Abstracção, no Minimalismo e no Construcionismo (*constructionism).*Constituindo abordagens específicas da ciência de computadores, é explicitada a aplicabilidade destes métodos à Filosofia, esclarecendo paralelamente quais os pressupostos epistemológicos que incorporam e implicam. Finalmente, a análise do Método de Abstracção leva a considerar a existência de um relativismo e pragmatismo cognitivo, na medida em que a obtenção de conhecimento se encontra dependente do nível de abstracção utilizado para estruturação da informação. Esta posição é coerente com o cepticismo que as teses do construcionismo implicam.

Concluo com uma crítica à aplicabilidade dos métodos propostos por Floridi para a Filosofia, defendendo que a utilização dos mesmos apenas é defensável através do compromisso com os pressupostos epistemológicos explicitados.

Conceito de Informação

A especificação do conceito de informação pressuposto por Floridi

será realizada em duas fases: A) uma análise da revisão da *Standard Definition of semantic Information* (SDI); B) uma análise da defesa do Realismo Informacional (*Informational Realism*) feita por Floridi.

A)

No seu artigo "Is Semantic Information Meaningful Data?" de 2005, Floridi procura contribuir para a definição do conceito de informação semântica através de uma crítica e revisão à SDI, assumindo aqui o conceito de informação um carácter semântico, declarativo e objectivo. Floridi procura com esta restrição metodológica colocar fora do âmbito da sua análise concepções ligadas ao carácter pragmático da informação. Esta opção, na minha opinião, privilegia uma análise ontológica que se mostrará dependente do carácter epistemológico da informação.

A análise de Floridi parte da definição, pela SDI, de condições necessárias e suficientes para a existência de informação semântica:

D.1. A Informação (λ) é constituída por n dados (d), sendo $n \geq 1$;
D.2. Os dados são bem formados (wfd);
D.3. Os wfd são significativos, ou seja, possuem um significado (mwfd = δ).

Floridi retira algumas conclusões acerca desta definição de informação semântica:

1. É necessária a existência de dados para existir informação;
2. Existe Neutralidade Tipológica (NT) – δ, como variável, deixa em aberto que tipo de dados bem formados e significativos constituem informação) – Floridi propõe 4 tipos de informação possíveis[3];
3. Existe Neutralidade Taxonómica (NTx) – numa perspectiva minimalista, a informação pode ter por base apenas um dado, sendo que este dado consiste numa diferença ou ausência de uniformidade, implementável (e, claro, reconhecível) fisicamente. Conclui--se que um dado, por mais simples que seja, é sempre uma relação de diferença entre dois elementos relacionais, ambos constituintes do dado. Dada a classificação subdeterminada destes elementos relacionais, Floridi afirma a

[3] Conferir Floridi, Luciano, 2005a: 354.

NTx, ao mesmo tempo que define a própria relação como binária, simétrica e externa. O dado é definido como uma entidade relacional, constrangedora de possibilidades (*constraining affordances*) de produção de informação através de um processo de "interrogação" e semantização destes dados.

4. Existe Neutralidade Ontológica (NO) – dada a constituição do dado como diferença ou ausência de uniformidade, a possibilidade de implementação desta relação em diversos tipos de suportes (físicos ou não) leva Floridi a afirmar a NO da informação. De qualquer forma, encontra-se implícito que não existe informação sem representação de dados.
5. Existe Neutralidade Genética (NG) – perante o problema da constituição do significado e função dos dados – problema da semantização dos dados, num dado sistema semiótico. No contexto da definição de informação pela SDI, não se coloca o problema de como se processa esta semantização, mas antes se a semantização e o consequente significado dos dados são independentes de um agente cognitivo ou não. Segundo Floridi, os dados significativos podem ter uma semântica independente de qualquer agente cognitivo. No entanto, Floridi afirma que a NG não suporta a tese realista segundo a qual os dados podem possuir a sua própria semântica independentemente de um agente cognitivo inteligente. Ou seja, a NG é uma tese de neutralidade fraca, propondo apenas a possibilidade de existência de informação independentemente de essa informação ser compreendida por um sujeito cognitivo (no entanto, a constituição inicial dessa informação encontra-se dependente de um sujeito). Penso que esta posição de Floridi é inconsistente: por um lado, a NG afirma a existência, em δ, de uma semântica independente de qualquer sujeito cognitivo; por outro lado, defende a existência de uma objectividade da informação, na medida em que esta existe independentemente, ontologicamente e epistemologicamente, de um intérprete, mas só após a criação da mesma por um agente cognitivo. Penso que esta inconsistência se deve ao facto de Floridi pretender defender a existência da informação *per si*, após a sua génese, e, consequentemente, a possibilidade de significado não apenas na mente do intérprete,

salvaguardando uma objectividade da interpretação. No entanto, penso que a semântica implícita numa informação é relativa ao agente cognitivo que realiza a semantização dos dados, de acordo com os constrangimentos que estes dados constituem, o que impossibilita a objectividade interpretativa no seu sentido mais estrito.

6. Existe Neutralidade Alética (NA) – pelo facto da SDI afirmar as condições D.1 – D.3 como suficientes para a constituição de informação semântica, por omissão, defende a existência de uma NA. Segundo Floridi, esta NA é problemática e leva à necessidade de reformular a SDI.

A reformulação crítica da SDI por Floridi parte da crítica à NA. Segundo este filósofo, a SDI, implicitamente, afirma que o valor de verdade da informação é uma propriedade superveniente (no sentido fraco), pelo que δ qualifica-se como informação independentemente de representar ou transmitir algo verdadeiro ou falso, ou mesmo de não possuir qualquer valor de verdade.

Assim sendo, falsa informação ou "desinformação", da mesma forma que tautologias, são classificáveis como informação. Floridi afirma que considerar estes tipos de dados como informação não é sustentável, justificando a sua posição através da apresentação de 9 "más" razões para defender a neutralidade alética e de 2 "boas" razões para afirmar a natureza alética da informação.

Após a apresentação da sua argumentação, Floridi reformula a concepção de informação da SDI, acrescentando-lhe uma nova condição (F4):

D.1. A Informação (λ) é constituída por n dados (d), sendo $n \geq 1$;
D.2. Os dados são bem formados (wfd);
D.3. Os wfd são significativos, ou seja, possuem um significado (mwfd = δ);
F.4. Os δ são verdadeiros.

Os argumentos utilizados não nos interessam no contexto deste artigo, pelo que não são apresentados. O que se mostra relevante são as consequências epistemológicas da concepção de informação apresentada por Floridi.

Primeira consequência (ontológica). Floridi defende que não existe informação sem dados bem formados e significativos. Ora, a

semantização dos dados bem formados terá que ser realizada através de algum processo, que Floridi não analisa. Todavia, Floridi procura afirmar que, neste processo, o papel do agente cognitivo é mínimo, na medida em que existe uma NG que torna independente esta semantização de qualquer agente. Neste ponto, Floridi revela uma inconsistência na medida em que terá sempre que existir um agente capaz de realizar a atribuição de significado aos dados bem estruturados. Só assim é possível compreender a sua definição de dados como constrangedores de possibilidades (*constraining affordances*). Possibilidades de constituição de informação porque interpretáveis por um agente utilizando a semântica que o agente possui; constrangedores porque limitadores das possibilidades de interpretação, por oposição a determinantes objectivos de significado.

A própria definição de dados, como diferenças ou ausência de uniformidade, revela a subdeterminação destes em relação ao seu significado. A própria ontologia relacional que esta definição dos dados define, leva a identificar a ontologia dos dados como externa aos elementos relacionais que compõem esses dados. Ou seja, a capacidade de alguma entidade constituir um dado capaz de ser utilizado na constituição de informação encontra-se dependente da sua relação de diferença face a outra entidade. Estas entidades definem-se ontologicamente *per si*, sem referência mútua. Esta referência apenas é necessária ontologicamente na sua definição ontológica enquanto dados. A NO que Floridi apresenta encontra-se de acordo com a definição relacional dos dados. Os elementos relacionais não interessam *per si* (NTx), inclusivamente nas suas características ontológicas. É a sua relação que se encontra em causa. E é nesta relação que o processo de semantização terá que se basear para a constituição do significado da informação.

Tendo em consideração estas características ontológicas dos dados, não parece possível afirmar a possibilidade de criação de informação a partir de dados, sem a intervenção decisiva de um agente cognitivo, que utiliza o seu aparato linguístico-semântico na interpretação dos dados.

Concluindo, Floridi possui um conceito de informação que assenta numa ontologia da (dupla) relação: relação entre entidades (susceptível de interpretação, logo de constituição de dados); relação entre dados e um agente cognitivo (que atribui significado à relação entre entidades com que se depara).

Segunda consequência (epistemológica). Dada a ontologia

apresentada, a Informação é uma construção interpretativa de um agente cognitivo, o que implica a relatividade da informação. Por um lado, a constituição ontológica dos dados como relação externa entre entidades relacionáveis – mas que em si mesmas (*per si*) carecem de uma necessidade ontológica de tal relacionamento[4] – leva-nos a classificar esta relação como contingente. Contingente porque epistemologicamente construída pelo agente cognitivo. Mesmo que se negue esta contingência adoptando um holismo ontológico que defenda a interdependência ontológica das entidades e, por este meio, a realidade ontológica das relações constituintes destas entidades, o agente cognitivo realiza sempre uma selecção das relações a considerar como dados. Ou seja, é realizada uma selecção de relações que limita e altera epistemicamente as relações e entidades em causa, dado o grande número (possivelmente infinito) de relações que constituem estas entidades e relações. O agente realiza uma selecção do conjunto de diferenças (ou ausência de semelhanças) que constituem os limites constrangedores na criação da informação[5]. Este é o primeiro nível de relatividade da informação.

Por outro lado, o processo de semantização dos dados é em si mesmo interpretação. Senão vejamos. Os dados são definidos como diferenças ou dissemelhanças, passíveis de constituir um recurso para a criação de informação. Existe claramente uma subdeterminação do significado passível de atribuição a estes dados. Caso contrário como seria possível aceitar que os mesmos factos científicos (descritos através dos mesmos dados) possam ser descritos por diferentes teorias (exclusivas)? No próprio dia a dia é possível verificar a existência de diferentes descrições de factos, realizadas por diferentes indivíduos, tendo por base os mesmos dados, e sendo necessário aceitar como válidas várias destas descrições. Voltando ao nosso tema. Dada esta subdeterminação, o processo de semantização encontra-se dependente de: escolhas do agente, implícitas ou explícitas, relativamente ao significado adequado dos dados; capacidade semântica do agente

---

[4] Não pretendo aqui atacar a necessidade de alteridade enquanto necessidade ontológica de diferenciação e afirmação ontológica de entidades distintas. Pretendo salientar apenas que esta alteridade ontológica não constitui nem é identificável com a diferença ou ausência de semelhança com que Floridi define dados. A justificação desta posição deve-se à caracterização que Floridi realiza da relação como exterior às entidades relacionáveis.

[5] Esta selecção pode inclusivamente estar relacionada com as capacidades de instrumentos de observação utilizados (caso dos instrumentos utilizados na ciência).

(domínio semântico da sua linguagem). Estamos situados no nível pessoal da relatividade da semantização.

Existe ainda uma relatividade de nível supra-pessoal, definido pela linguagem (natural ou artificial) utilizada pelo agente. A linguagem subjacente ao processo de semantização dos dados é determinante na definição do significado, na medida em que constitui o repositório das principais categorias e relações de predicação a utilizar no processo. Percebe-se agora a importância, ao nível pessoal, do domínio semântico da linguagem pelo agente.

Dados estes argumentos, defendo que a informação, tal como Floridi a define implicitamente, é relativa[6]. Não no sentido mais radical, mas enquanto dependente da semantização e da capacidade selectiva do agente cognitivo, na medida em que: o papel da linguagem no processo de semantização garante a intersubjectividade, o que permite a objectividade consensual (no limite); os aparatos sensoriais do Homem são partilhados pela espécie e permitem uma convergência mínima da selecção das relações consideradas como dados.[7]

B)

Para aprofundar a análise das implicações epistemológicas do conceito de informação em Floridi é necessário analisar a sua posição a favor de um Realismo Informacional (*Informational Realism*).[8] Floridi defende que a natureza última da realidade é informacional, argumentando que: a) é possível uma reconciliação do Realismo Estrutural Epistémico (REE) com o Realismo Estrutural Ontológico (REO), através da metodologia dos níveis de abstracção, tornando o REO defensável (de um ponto de vista pró-estruturalista); b) o REO é plausível, na medida em que nem todas as entidades relacionadas são logicamente anteriores a todas as estruturas relacionais. Floridi privilegia a relação de diferença para fundamentar esta afirmação; c) é possível desenvolver uma ontologia de entidades estruturais, no âmbito do REO, utilizando objectos/entidades informacionais.

---

[6] Dado o objectivo epistemológico deste artigo, a utilização de "relativo" cinge-se à espécie humana. Não me pronuncio acerca da existência de informação em outros organismos, como sejam animais ou máquinas computacionais.

[7] Estes argumentos necessitariam de uma fundamentação que escapa ao objectivo deste artigo, pelo que se deixa em aberto para futuros trabalhos.

[8] Esta secção do artigo baseia-se em grande parte em Floridi "Informational Realism"

O resultado desta argumentação é, segundo o filósofo, o Realismo Informacional.

Proponho uma análise desta argumentação nos seus pormenores:
a) Reconciliação do REE com o REO. Floridi inicia este argumento definindo sucintamente Realismo Estrutural (RE), REE e REO: RE – o conhecimento do mundo é conhecimento das suas propriedades estruturais; REE – os objectos são o que permanece, em princípio, incognoscível, após colocar de parte as estruturas cognoscíveis da realidade; REO – os objectos são eles próprios estruturas.

Floridi continua, desenvolvendo a relação entre estas três propostas filosóficas. Segundo ele, o RE defende que modelos, bem sucedidos previsional e instrumentalmente, podem ser, nas melhores circunstâncias, crescentemente informativos acerca das relações que obtêm entre os (possivelmente inobserváveis) objectos que constituem o sistema sob investigação (através dos fenómenos observáveis). Assim definido, o RE não especifica a natureza dos objectos relacionados nas estruturas. Estes constituem "caixas negras", epistemicamente inatingíveis.

Saliente-se que a não especificação da natureza dos relacionáveis aponta para duas posições subjacentes possíveis: ou a natureza destes objectos não é estrutural e daí a nossa incapacidade de conhecer os mesmos (segundo o RE); ou a natureza destes objectos é estrutural, o que coloca ao RE o problema de explicar a nossa incapacidade epistemológica de os captar nos seus modelos. Esta conclusão, que vou agora deixar em suspenso, será importante brevemente. Voltando a Floridi.

Dado o problema da não especificação da natureza dos objectos relacionados, Floridi evidencia a consequente reiteração da questão acerca da cognoscibilidade do estatuto ontológico dos objectos. Com o objectivo de especificar as posições do REE e do REO, Floridi apresenta o seguinte formalismo representativo de uma estrutura genérica S: O – conjunto não vazio de objectos (o domínio de S); P – um conjunto não vazio de propriedades de primeira ordem dos objectos em O; R – um conjunto não vazio de relações em O; T – um conjunto potencialmente vazio de regras de transição (operações) em O.

O filósofo parte desta terminologia formal para caracterizar as respostas de REE e REO à questão: o que são (ontologicamente) os objectos no conjunto O?

Floridi conclui que para o REE os objectos apenas podem consistir num resíduo ontológico e, consequentemente, o conjunto P é

incognoscível.

Em relação ao REO, Floridi verifica a existência de uma diferença: os objectos são, em si mesmos, estruturas e na melhor das hipóteses podem ser indirectamente captados nos nossos modelos, pelo menos em princípio. De facto, esta posição parece coerente com a própria designação de REO – realismo estrutural ontológico consiste numa posição ontológica que defende que a constituição última da natureza da realidade é estrutural. Neste sentido, como não concluir que tudo o que existe é, em si mesmo ou naquilo que o constitui de mais básico, estrutura ou relação estrutural?

Todavia, repare-se que Floridi afirma mais do que isto. Afirma que os objectos poderão ser captados pelos modelos (epistémicos). Será que é possível deduzir esta posição epistémica das premissas ontológicas do REO? Não me parece. O REO apenas afirma a natureza ontológica e não possibilidades epistémicas (a não ser que seja capaz de deduzir da ontologia dos seres cognoscentes as suas capacidades cognitivas, o que não acontece explicitamente nos dados apresentados por Floridi acerca do REO). Como explicar então a posição epistemológica de Floridi?

Permitam-me invocar a conclusão relativa ao RE (onde se incluem o REE e o REO) que tinha ficado em suspenso:

> ”Saliente-se que a não especificação da natureza dos relacionáveis aponta para duas posições subjacentes possíveis: ou a natureza destes objectos não é estrutural e daí a nossa incapacidade de conhecer os mesmos (segundo o RE); ou a natureza destes objectos é estrutural, o que coloca ao RE o problema de explicar a nossa incapacidade epistemológica de os captar nos seus modelos.”

Pretendendo afirmar um realismo informacional de natureza estrutural (como veremos), Floridi vê-se confrontado com a necessidade de defender o REO. Ao assumir esta posição vê-se confrontado com uma contradição que tem que resolver, sob pena de ver ruir o seu realismo pelos próprios princípios que o sustentam: os objectos são estruturais mas não captáveis pelos modelos estruturalistas do REE.

O REE defende a possibilidade de captar estruturas e não objectos apenas na medida em que afirma, implicitamente, a natureza dos mesmos como não estruturais. De facto, o critério de possibilidade de

conhecimento do REE baseia-se nas características ontológicas da realidade enquanto estrutural (relações entre objectos) ou não (objectos). Toda a natureza estrutural da realidade terá que ser captável pelos seus modelos, caso contrário o REE vê-se confrontado com outra questão: se existem elementos da realidade de natureza estrutural que podem ser conhecidos e outros que não podem, então o critério de possibilidade de conhecimento da realidade baseado na natureza (ontológica) dos elementos da mesma não é já válido. Como explicar então uma posição (o REE) baseado nesse mesmo critério?

Assim sendo o REE tem que pressupor uma constituição ontológica dos objectos diferente de entidades estruturais. Esta posição é oposta ao que o REO afirma, dizendo que os objectos são estruturas em si mesmos.

É devido a esta oposição que Floridi propõe a possibilidade de um conhecimento indirecto dos objectos (estruturas) através dos modelos desenvolvidos pelo REE, residindo a chave na caracterização deste conhecimento como indirecto. Através desta variação metodológica Floridi pretende conciliar o REE e o REO. No entanto, parte já de uma posição falaciosa: o REO não defende a possibilidade de conhecimento (indirecto ou não) dos objectos (assumindo-os como estruturas); e, assumindo o REO, esses objectos teriam mesmo que ser captáveis nos modelos da realidade do REE, dada a sua natureza estrutural (o que não acontece segundo os seus defensores).

Apresentada esta crítica, assumamos as propostas de Floridi como válidas e continuemos a análise.

A conciliação do REE com o REO é tentada através da caracterização do possível conhecimento das relações e dos objectos como indirecto, face a um conhecimento directo dos mesmos (como explicação das propriedades intrínsecas do sistema). Como é que Floridi realiza esta diferenciação?

Primeiro Floridi caracteriza o RE como consistindo num *trade-off* que assume um compromisso ontológico enfraquecido em troca de uma elasticidade epistemológica que permite incorporar uma elasticidade no conhecimento, incorporando o "avanço científico", por exemplo. Este *trade-off* resulta da separação, dentro do âmbito epistémico, entre as descrições das características estruturais cognoscíveis do sistema e as explicações das suas propriedades intrínsecas (explicações demasiadamente comprometidas com um nível ontológico, logo, aqui não consideradas possíveis pelo filósofo). O conhecimento passa aqui a ser entendido como uma relação indirecta, descritiva, entre um agente

epistémico e o sistema em análise.

Esta caracterização do conhecimento permite a Floridi a introdução de um método proposto por ele para o âmbito da filosofia: o método de abstracção[9]. Este método é utilizado na elaboração dos argumentos a favor da não incompatibilidade entre REE e REO, assim como para a caracterização do conhecimento obtido pelos seres humanos.

Retomando a caracterização do conhecimento como relação indirecta, logo ontologicamente menos comprometido, Floridi acaba por caracterizar a relação epistémica através do método de abstracção (segundo a nossa interpretação). De facto, apesar deste método consistir nisso mesmo, uma metodologia, Floridi parece afirmar que o método consiste numa formalização de um processo cognitivo presente na relação dos agentes cognitivos com o mundo. Ou seja, não formalmente ou estruturadamente, o processo cognitivo humano apresenta características semelhantes ao método de abstracção. Reforçando esta interpretação, Floridi afirma que uma das vantagens e objectivos do método de abstracção é tornar visível e consciente o nível de abstracção em que determinado conhecimento é criado.

Uma caracterização do método em causa será esclarecedora da relação epistémica em causa.

O método de abstracção é proveniente da modelação realizada em ciência. É influenciado por uma área específica da Ciência de Computadores denominada Métodos Formais, onde é utilizada matemática para especificar e analisar o comportamento de sistemas de informação. Este método pode ser caracterizado como consistindo num conjunto de entidades matemáticas que pretendem representar um dado sistema, sempre a partir de um dado nível de observação. As entidades em causa são:

- observáveis (uma colecção de observáveis) – um observável é uma variável declarada e interpretada; ou seja, uma variável cujos valores possíveis se encontram bem definidos (declarados) e uma variável que se encontra associada a determinadas propriedades, que pretende representar, do sistema em análise.
- nível de abstracção (LOA) – um determinado nível de abstracção é constituído por um conjunto de observáveis e por

[9] Analisado na segunda parte deste artigo.

> um comportamento que define o relacionamento entre os observáveis, ou seja, o comportamento que o sistema pode assumir.

Através destas entidades, Floridi propõe uma formalização do conhecimento humano, baseando-se na possibilidade da existência de múltiplos e diversos níveis de abstracção possíveis para representar um dado sistema. Dado o facto deste método permitir a criação de modelos dos sistemas reais, a utilização de um dado conjunto de observáveis (LoA) irá ter como consequência um determinado modelo, diferente daquele que se obteria pela utilização de outro LoA.

Neste sentido, em que medida se relaciona este método com a proposta de conhecimento indirecto dos sistemas reais?

Utilizando este método, o filósofo analisa o conhecimento considerando-o como resultado de um modelo da realidade criado epistemicamente pelo agente cognitivo. Modelo este que corporiza um LoA, ou seja, um conjunto de observáveis que significa um compromisso ontológico face à realidade (subdeterminada epistemicamente). O conhecimento do mundo é realizado através da análise de um modelo dessa realidade, construído tendo como referência um conjunto de observáveis e o comportamento do LoA em causa, definidos e utilizados pelo agente cognitivo. A possibilidade de diversos modelos, ou seja, de diversos LoA, explica-se pela subdeterminação epistémica da realidade face ao agente cognitivo, o que, no entanto, não altera o estatuto ontológico da realidade. É apenas na relação entre o agente cognitivo e a realidade que se encontra a subdeterminação. Floridi não explora esta relação, no entanto somos levados a assumi-la como essencial na compreensão dos seus pressupostos epistemológicos.

Nesta relação é possível analisar a relação entre o conhecimento e a realidade (ontológica) em Floridi. No fundo trata-se de responder à questão do relativismo (aparente) desta proposta. Numa primeira análise, o conhecimento é de facto relativo, mas relativo a um dado LoA. O que é que isto significa? Significa que o modelo desenhado pelo cognoscente é função do conjunto de variáveis e comportamentos considerados pelo mesmo na elaboração do modelo. O que é que fica aqui em aberto? A origem deste conjunto de observáveis e comportamentos, assim como a completude cognitiva dos mesmos na explicação da realidade. Segundo o que Floridi parece querer indicar, o agente cognitivo realiza uma determinação dos observáveis e dos

comportamentos. No entanto, esta determinação parece ser uma selecção e não um acto constituinte dos mesmos. O agente selecciona, mas selecciona algo ontologicamente válido, na medida em que ontologicamente fundado em dados (em diferenças constituintes da realidade, em relações). Os observáveis surgem como interpretações possíveis dadas as limitações ontológicas. Limitações ontológicas porque limitações da interpretação da realidade tendo sempre em consideração a subdeterminação epistémica da realidade (dos dados).

Através desta análise as propostas de Floridi, apesar de aparentemente relativistas, têm que ser apresentadas como diferentes desta posição. De facto, há uma fundação ontológica do conhecimento nos dados. Existe um vínculo ontológico do conhecimento, ainda que enfraquecido, que permite delimitar as possibilidades explicativas e incorporar as limitações de agentes cognitivos epistemicamente limitados (nunca capazes de incorporar a totalidade dos observáveis e comportamentos num modelo e/ou de unificar os diferentes níveis de abstracção possíveis numa "abstracção unificada"[10]).

Retomemos a argumentação de Floridi a favor da reconciliação entre o REE com o REO. Dado o carácter indirecto do conhecimento, Floridi pode afirmar que um LoA permite a uma teoria analisar um sistema (uma parcela da realidade) e criar um modelo dessa realidade onde se encontra identificada a estrutura desse sistema a esse nível de abstracção. O REE é englobado por esta descrição de Floridi, ao mesmo tempo que é caracterizado como constituindo uma aproximação minimalista: na medida em que existe um compromisso ontológico apenas com uma interpretação realista das propriedades estruturais do sistema. Este compromisso ontológico é caracterizado como primário, ou seja, é relativo a um conhecimento de primeira-ordem da estrutura do sistema. É com esta caracterização que Floridi prepara a resolução do antagonismo entre REE e REO. Floridi assume este compromisso ontológico como respeitando a navalha de Ockam, ou seja, como não constituindo compromissos desnecessários ou demasiado arriscados ou suspeitos. Como pode então o REO ser incorporado como uma posição defensável e conciliável?

De uma forma simples: Floridi afirma que o REE e o REO trabalham em níveis epistémicos diferentes, ou seja, em LoA diferentes. O REO analisa a realidade de forma derivada, ou seja, assume um

[10] O termo é nosso, pretendendo significar um paralelismo com as aspirações de unificação da ciência em torno de Teorias Unificadas.

compromisso ontológico, e possibilita um conhecimento, de segunda-ordem. Dado este compromisso de segunda-ordem implicar a consideração dos objectos como entidades relacionais (visão minimalista dos mesmos, segundo Floridi), a navalha de Ockam não contesta as pretensões epistémicas do REO ao conhecimento dos mesmos, na medida em que estas pretensões se realizam sem implicar compromissos ontológicos mais profundos que os do REE. De facto, o REO é definido, através da metodologia dos níveis de abstracção, como um nível de abstracção superior (apenas em termos formais) face à realidade. O REO acede ao conhecimento dos objectos de uma forma indirecta das inferências possíveis a partir dos modelos relacionais obtidos pelo REE. Dada a estrutura do modelo da realidade do REE, o REO pode inferir as propriedades relacionais/estruturais que os objectos que constituem esse modelo têm que possuir para que tal modelo seja possível. Subimos um nível na análise da realidade. Passamos para uma análise dos objectos, via a consideração das condições de possibilidade de conhecimento das propriedades estruturais do sistema.

Saliente-se que esta posição de Floridi é coerente com as suas propostas relativas à consideração dos dados como entidades ontológicas relacionais. Os objectos serão, em si mesmos, relacionais, logo estruturalmente captáveis como relacionáveis. No entanto coloca-se a questão (aqui não analisada): até que ponto se considera a realidade como constituída por entidades relacionáveis? *Ad infinitum*?

Após a conciliação entre o REO e o REE através da sua identificação com diferentes níveis de análise e com uma perspectiva transcendental de conhecimento, Floridi procura consolidar a sua ontologia defendendo que os relacionáveis (*relata*) não são logicamente anteriores a todas as relações.

b) Floridi procura defender a plausibilidade do REO mostrando que nem todas as entidades relacionadas são logicamente anteriores a todas as estruturas relacionais. Caso isto se verifique Floridi pode afirmar que, em si mesmas, estas entidades relacionadas, ou relacionáveis, se definem ontologicamente como entidades estruturais. Esta conclusão solidificaria toda a sua concepção de informação e a sua ontologia informacional.

Com este objectivo, e procurando não cair na crítica de regressão ao infinito, Floridi introduz a distinção entre relações externas e internas. Os exemplos apresentados para clarificar cada uma são, respectivamente, a relação de distância e a relação de casamento. Sendo o casamento uma relação interna, é caracterizada como constituindo os

elementos relacionados naquilo que eles são. No entanto, ao contrário da intenção de Floridi, as relações internas como a de casamento parecem ser supervenientes, na medida em que parecem aplicar-se aos elementos relacionados posteriormente à sua existência e parecem caracterizá-los contingentemente. O objectivo de Floridi passa precisamente por conseguir defender a prioridade lógica das relações internas sobre os relacionáveis. O filósofo introduz aqui uma "definição operacional" que lhe permite atingir o seu objectivo. Floridi define que a afirmação da prioridade lógica das relações internas passa por mostrar que as propriedades essenciais dos objectos relacionados em questão dependem de algumas propriedades internas fundamentais.

Com esta "definição operacional", desloca-se a discussão das relações internas para o campo das propriedades internas fundamentais. A partir daqui é utilizada uma relação metafísica, segundo Floridi ainda mais fundamental do que a definição da essência dos relacionáveis: a relação de diferença, na medida em que esta constitui a própria possibilidade de existência dos relacionáveis. Apesar do filósofo não elaborar esta fundamentalidade da relação de diferença, parece-nos que Floridi utiliza a diferença como a possibilidade de diferenciação ou alteridade ontológica do elemento (qualquer que ele seja) face a um fundo ontológico indiferenciado e a outros elementos que apenas se constituem como tal na medida em que são distinguíveis (ontologicamente). Repare-se que esta visão aparece já inserida num paradigma de diferenciação/alteridade, logo de relação.

Floridi consente que esta relação de diferença nos diz muito pouco acerca da natureza dos relacionáveis, todavia consegue apresentar um argumento razoável da prioridade de uma relação face aos elementos relacionados. Resta-nos uma dúvida: será que é extrapolável afirmar a prioridade lógica de outras relações face aos relacionáveis? E será que esta relação de diferença é de facto interna? Não será antes um resultado da existência de diferentes propriedades intrínsecas nos relacionáveis? Não poderá a existência ser uma instanciação de propriedades: caso estas propriedades sejam diferenciadoras fala-se de existência; caso o não sejam fala-se de não existência? Ou seja, o fluxo lógico não será das propriedades para a relação de diferenciação, apesar de esta relação se constituir como condição transcendental de existência? Repare-se que estas objecções se situam já fora do âmbito restrito da lógica, o que incita outra questão: será que a prioridade lógica desta relação permite fundamentar uma ontologia da relação e da informação, como o filósofo pretende? Da prioridade lógica deriva uma

prioridade ontológica? São muitas questões a esclarecer para afirmar a validade dos argumentos de Floridi…

Mas continuemos com a análise da sua defesa do realismo informacional.

c) É possível desenvolver uma ontologia de entidades estruturais, no âmbito do REO, utilizando objectos/entidades informacionais. Floridi assume neste passo que conseguiu estabelecer a validade de uma visão estrutural sobre a realidade, tanto ontológica como epistemológica. Procede então à tentativa de convergência entre a ontologia estrutural e uma ontologia baseada em objectos informacionais.

A proposta é que os objectos relacionais sejam considerados objectos informacionais, ou seja, aglomerados (*clusters*) de dados no sentido ontológico de diferenças ou ausências de uniformidade. A concepção de dado foi já esclarecida quando abordado o conceito de informação. Segundo esta, o dado só pode ser ontologicamente compreendido como relação entre dois elementos contrastantes, logo como ausências de uniformidade. Esta relação constitui em si mesmo o dado. Floridi estabelece o dado como diferenças ontológicas ainda não qualificadas, no sentido de não completamente determinadas, logo como entidade relacional capaz de se enquadrar no estruturalismo proposto. Um aglomerado de dados constitui o conjunto de entidades que constitui em si mesmo um objecto relacional. Desta forma, o objectivo epistemológico do REO proposto por Floridi consiste neste mesmo aglomerado de dados enquanto entidades relacionais constitutivas dos objectos relacionados.

O resultado desta ontologia de objectos informacionais é o que Floridi designa por Realismo Informacional. Esta posição afirma a existência de uma realidade independente da mente que os agentes cognitivos podem apreender, epistemicamente e nunca completamente, como uma realidade estrutural e informacional. Esta posição integra uma humildade epistémica ao afirmar que a natureza última da realidade poderá ser ou não substancial, todavia não possuímos razões para a considerar como tal. Como precaução, Floridi assume o mínimo compromisso ontológico afirmando uma espécie de compromisso transcendental: apenas podemos afirmar que qualquer que seja a natureza última da realidade, ela terá que possuir determinadas características estruturais e informacionais (tal como Floridi as descreve). Com esta conclusão, epistemicamente os agentes encontram as suas limitações.

Podemos então dizer que:

1. Floridi incorpora na sua ontologia características derivadas da análise epistemológica (por outras palavras, submete a ontologia à epistemologia). Este facto revela a importância da abordagem transcendental para a ontologia do filósofo, assumindo a situação do filósofo como ponto de observação inultrapassável e a considerar de forma determinante na realização de uma ontologia.
2. A concepção de conhecimento inerente em Floridi não é compatível com concepções baseadas em representações mentais. Existe um problema epistémico de acesso à realidade. Este é sempre um processo mediado por LoA que nos dá apenas acesso indirecto às propriedades estruturais da realidade. Logo, não se trata de um acesso imediato às características do real, mas sim de um processo de inferência de propriedades estruturais não representativas do real, no sentido de não constituir um *faxsimilae* do real mas uma criação de conhecimento. Este processo de criação assume os dados (entidades relacionais) como um recurso e não como uma fonte privilegiada (como representação da realidade) e determinante do conhecimento. Os próprios dados são ontologicamente sub determinados (enquanto entidades relacionais), o que significa que os mesmos dados podem ser epistemicamente interpretados de forma diversa por diferentes sujeitos (dentro dos limites ontológicos existentes).
3. Apesar de Floridi propor a designação de Realismo Informacional, o compromisso ontológico na sua posição é mínimo e deixa em aberto a possibilidade da realidade se constituir como algo mais do que objectos informacionais. A sua resposta não é última, é a possível. Mais importante ainda, epistemologicamente esta posição incorpora um relativismo epistemológico parcial. Parcial porque existem fronteiras informacionais à nossa interpretação dos dados, que impedem de assumir um relativismo completo ou ainda um cepticismo radical. Existe um relativismo linguístico-sintáctico, fruto do processo de criação da informação utilizando uma determinada linguagem natural (ou artificial), formalizada nos LoA, e um

determinado conjunto de dados (seleccionados através do compromisso com um determinado LOA.

Metodologias Propostas por Floridi

Esta parte examina três metodologias propostas por Floridi: a) Minimalismo, b) Construcionismo e c) Método de Abstracção.

a) Minimalismo

Este método pretende lidar com a complexidade inerente aos problemas filosóficos *tout court*. O ponto de partida consiste na constatação de que a robustez de uma resposta a um problema filosófico depende da robustez de pressupostos e/ou assumpções. Estes constituem respostas que permitem lidar com outros problemas inerentes ao problema central em análise.

O minimalismo defende que é necessário determinar um problema, um ponto de partida, que se comprometa ou dependa o menos possível de outros problemas que não o central, o que permite aumentar a robustez da resposta que se procura. Até este ponto a proposta não surpreende. Mas como operacionalizar de uma forma estruturada este método?

Neste ponto Floridi inspira-se claramente na área das Ciências de Computadores. O filósofo introduz a ideia de que a facilidade de tratamento de um problema filosófico pode ser melhorada através da mobilização de sistemas discretos para análise do mesmo. Por si só esta concepção já merece uma reflexão. Em que consistem estes sistemas discretos? Apesar de Floridi não ser claro, dado o seu trabalho teórico adjacente, penso que se refere a sistemas formais como os níveis de abstracção (LoA) de que já falámos: construções matemático-formais, baseadas em variáveis definidas e relacionadas com características observáveis, mediadas por comportamentos que permitem ao sistema a sua evolução entre estados. A proposta deste tipo de sistemas para estudo da realidade pressupõe já uma aceitação da ontologia relacional e informacional do filósofo. Caso contrário pode colocar-se a questão da sua capacidade de captação epistémica da realidade.

Continuando.

Assumindo a necessidade de utilização deste tipo de sistemas discretos para estudo de problemas filosóficos, Floridi apresenta 3 critérios para que a sua selecção seja conforme ao minimalismo:

1. Controlabilidade – um sistema é controlável quando a sua estrutura pode ser modificada de acordo com determinado objectivo. Desta forma, o sistema pode ser utilizado como instanciação e teste de possíveis soluções.
2. Implementabilidade – os sistemas têm que ser implementáveis fisicamente ou por simulação. Este critério tem inerente o construcionismo que se analisará em b). Sumariamente, a implementabilidade implica para Floridi o conhecimento completo dos componentes do sistema e das suas leis transicionais. Só assim o sistema pode ser utilizado para instanciação e teste de hipóteses.
3. Previsibilidade – este critério decorre dos anteriores, sendo que a previsibilidade do comportamento do sistema só é possível pela controlabilidade e implementabilidade do mesmo.

O minimalismo apresenta as seguintes características:

- É relacional. O problema e o sistema investigados nunca conseguem ser completamente minimalistas. Encontram-se sempre relacionados com um espaço de problemas e sistemas ligados ao problema em análise.
- Este método permite-nos escolher o ponto de partida da análise do problema de uma forma crítica. A tratabilidade de um problema é função dos 3 critérios apresentados.
- O minimalismo foca-se nas relações inferenciais entre um problema e o espaço em que este se movimenta. Isto não significa que apenas os problemas simples sejam abordados. O que está em causa é a abordagem dos problemas e a consciência das inferências necessárias (ou evitáveis) face ao espaço desse problema.

Como resultado destas características, o minimalismo é aplicado tendo sempre em atenção um dado problema e o seu espaço envolvente, ou seja, constitui uma abordagem sempre relativa a um contexto de aplicação.

b) Construcionismo

Este método tem por base a concepção segundo a qual o construtor de um sistema sabe tudo acerca desse sistema na medida em que o construiu, mais ainda, apenas quem constrói um sistema sabe como este funciona. Paralelamente, Floridi faz assentar o construcionismo na impossibilidade de conhecimento da natureza da realidade em si mesma. Note-se que estes pressupostos se encontram em linha com a ontologia e epistemologia analisadas no capítulo anterior: o realismo informacional deixa em aberto a própria constituição ontológica da realidade, podendo esta incorporar algo para além dos objectos relacionais e as propriedades estruturais que este realismo identifica; este realismo remete ele mesmo enquanto aberto para o problema epistemológico do conhecimento da realidade em si mesma, conhecimento este indirecto. Estes elementos teóricos são reconhecíveis no construcionismo.

Neste contexto, Floridi salienta o conhecimento atingido pelo construtor do sistema. Não colocando em questão a impossibilidade de conhecimento da realidade em si, Floridi alinha por uma posição que defende a possibilidade de melhorar o nosso conhecimento da realidade através da melhoria do nosso conhecimento acerca das técnicas que utilizamos para investigar a realidade. Trata-se de uma concepção tecnológica que o próprio relaciona com a filosofia da tecnologia de Francis Bacon.

O construcionismo encara a possibilidade de aproximar o nosso conhecimento da realidade, de uma forma indirecta, através da investigação e construção dos próprios sistemas utilizados na elaboração do conhecimento. Estes sistemas são privilegiados na medida em que constituem construções humanas, logo cognoscíveis por nós, e os únicos meios que possuímos para aceder (indirectamente) à realidade.

Formalmente, o método consiste em 5 princípios:

1. Princípio do Conhecimento – apenas o que pode ser construído pode ser conhecido. Aquilo que não pode ser construído pode na melhor das hipóteses originar uma hipótese de trabalho;
2. Princípio da Possibilidade de Construção (*constructability*, no original) – as hipóteses de trabalho (só) são investigadas utilizando simulações (teoréticas ou práticas) baseadas nas mesmas.
3. Princípio da Possibilidade de Controlo – as simulações têm

que ser controláveis: têm que ser modificáveis em termos de composição e de previsão de resultados.

4. Princípio da Confirmação – qualquer confirmação ou infirmação das hipóteses de trabalho diz apenas respeito à simulação e não ao objecto (real) simulado. Incorpora um sub-princípio:
5. Princípio da Dependência do Contexto – o isomorfismo existente entre a simulação e o simulado é apenas local e não global.
6. Princípio da Economia – deve ser utilizado o menor número possível de recursos conceptuais. Os recursos devem sempre ser menores do que os resultados obtidos.

A influência das Ciências de Computadores é visível na descrição deste método. A resposta a problemas é realizada através de uma investigação. Esta investigação toma uma forma operacional em que é utilizado um sistema construído para o efeito. Este sistema (no caso de Floridi podemos remeter para os LOA como exemplo) serve para implementar e testar as hipóteses de trabalho que procuram responder ao problema em análise. O sistema pode ser modificado na sua estrutura ou regras internas, pode ser expandido ou restringido, consoante a necessidade de testar variações teóricas às hipóteses de trabalho.

Todavia, desta investigação, o filósofo pode apenas retirar conclusões acerca do modelo (sistema) que utiliza para testar as suas hipóteses. O modelo representa apenas o real num isomorfismo local (Principio 4.1). A relação entre o simulado e a simulação encontra-se abrangida pela incapacidade epistemológica de conhecer o real. Assim se compreende a designação de "hipóteses de trabalho" que Floridi utiliza. A localidade do isomorfismo revela a limitação desse isomorfismo e assim da própria relação epistemológica em que esse isomorfismo foi construído.

Confirma-se a análise já efectuada da relação do mental com o real. O conhecimento não é constituído de uma forma mimética, representacional. O mecanismo constituinte do conhecimento proposto pelo construcionismo em certos aspectos é o inverso. Em primeiro lugar nega a existência de um qualquer mecanismo misterioso que crie "cópias" mentais da realidade. Em segundo lugar, afirma que a cognição é um processo de modelação. Modelação na medida em que o agente cognitivo dá forma e modifica a realidade procurando torná-la inteligível para ele. Ou seja, de um processo em que a realidade exercia

a sua acção no sujeito (originando representações de si mesma), passa-se para a acção do sujeito que permite a construção e modelação de entidades inteligíveis. Obviamente, só os modelos são conhecidos. Onde fica então a realidade modelada? Esta é passível de uma apreensão indirecta e sempre por aproximação. O conhecimento cai, segundo a minha perspectiva, para uma concepção relativa (um relativismo limitado como já explicitado) e potencialmente pragmática. Dada a indeterminação ontológica, que se repercute e vê potenciada num processo epistemológico indirecto, o conhecimento deixa de se poder fundar ontologicamente e tem que procurar outro "fundamento" ou "âncora", capaz de o legitimar. Floridi não se pronuncia acerca da fundamentação do conhecimento senão implicitamente ao afirmar barreiras ontológicas que não permitem ultrapassar um certo nível de interpretação dos dados sem cair na contradição.

Com o Princípio da Economia, o construcionismo procura assumir o mínimo de compromissos ontológicos, respeitando o minimalismo e a navalha de Ockam. Assim sendo, verifica-se que existe uma relação e coerência na adopção dos diferentes métodos. Analisemos o último método.

c) Método de Abstracção

Tendo este método sido objecto de abordagem no capítulo anterior, a sua explicitação incidirá nos aspectos essenciais à análise.
Este método tem por objecto uma análise fenomenológica ou conceptual, orientada para a explicitação do nível de abstracção (LoA) em que o discurso é realizado, ou o sistema é considerado. Ao realizar esta análise explicita-se quais os observáveis considerados na elaboração do discurso[11], assim como quais os predicados, comportamentos e, possivelmente, os LoA (gradiente de abstracções) em redor do LoA utilizado.

Encontrando-se os restantes elementos apresentados, definimos apenas o gradiente de abstracções: um gradiente de abstracções consiste num conjunto de LoAaplicado sobre um mesmo sistema, sendo que a relação entre estes LoA é definida através de elementos relação, incluídos no gradiente de abstracções. Esta relação estabelece as correspondências entre observáveis de diferentes LoA, de tal forma que

[11] Assume-se aqui discurso no sentido de verbalização de conhecimento, logo como sinónimo deste.

é possível a comunicação entre LoA.

A aplicação dos níveis e gradientes de abstracção é possível em qualquer sistema, dada a hipótese de formalização do mesmo. Como já explicitado, o construcionismo remete para o uso de sistemas no processo cognitivo, mais concretamente para sistemas que modelam a realidade. Como explica o método de abstracção a constituição destes modelos?

Estes modelos constituem uma particular relação entre diferentes níveis de abstracção: a relação de simulação. Esta relação consiste na relação entre observáveis de um sistema simulador e os observáveis de um simulado. Na medida em que se pretende congruência entre os sistemas, a relação tem que existir entre pares de observáveis. Desta forma, procura-se que a evolução dos sistemas entre estados ocorra em paralelo e de facto se possa falar de simulação.

Segundo Floridi, este tipo de utilização dos LoA implica que um agente epistémico crie ou tente criar uma relação de equivalência entre estes dois sistemas, observando-os em diferentes LoA e procurando compreender em que níveis de abstracção estes sistemas são equivalentes.

Constituintes do método:

1. **Observáveis** (*observables*)
2. **Predicados** (*predicates*)
3. **Comportamento** (*behaviour)*
4. **Gradiente de abstracções** (*gradient of abstractions*)

O Gradiente de abstracções formaliza as condições de consistência mínima que as abstracções escolhidas devem respeitar.

Resta esclarecer a aplicabilidade destes métodos à Filosofia, esclare- cendo paralelamente quais os pressupostos epistemológicos que incorporam e implicam. Finalmente, a análise do Método de Abstracção leva a considerar a existência de um relativismo e pragmatismo cognitivo, na medida em que a obtenção de conhecimento se encontra dependente do nível de abstracção utilizado para estruturação da informação. Esta posição é coerente com o cepticismo que as teses do construcionismo implicam, assim como com a sua posição materialista.

Referências

Floridi, Luciano, 2005a, "Is Semantic Information Meaningful Data". *Philosophy and Phenomenological Research*, vol. LXX, Nº.2, 351-370.

Floridi, Luciano, 2005b, "Information", invited contribution to the *Encyclopedia of Science, Technology, and Ethics*, (ESTE) edited by Carl Mitcham (Macmillan, 2005).

Floridi, Luciano, & Sanders, Jeff W., 2004a, "The Method of Abstraction", invited chapter for the *Yearbook of the Artificial* (Issue II, 2004, Peter Lang) dedicated to "Models in contemporary sciences", pp. 177-220.

Floridi, Luciano, 2004b, "Informational Realism", in Computers and Philosophy 2003 – Selected Papers from the Computer and Philosophy conference (CAP 2003),

Floridi, Luciano, 2004c, "From the Philosophy of AI to the Philosophy of Information", invited contribution to *The Philosophers' Magazine* 2004, 28.4, pp. 56-60.

Floridi, Luciano, 2004d, "Open Problems in the Philosophy of Information, Metaphilosophy", 35.4, pp. 554-582. Revised version of The Herbert A. Simon Lecture on Computing and Philosophy given at Carnegie Mellon University in 2001

Greco, Gian Maria; Paronitti, Gianluca; Turilli, Matteo; Floridi, Luciano; 2005b, "How to Do Philosophy Informationally", Lecture Notes on Artificial Intelligence, 3782, 623–634.

Greco, Gian Maria; Paronitti, Gianluca; Turilli, Matteo; Floridi, Luciano; 2005c, "The Philosophy of Information – A Methodological Point of View".

Conceito de crença, triangulações e atenção conjunta

Sofia Miguens[1]

Resumo: Na última fase da sua obra D. Davidson desenvolve um conjunto de teses em torno da ideia de triangulação. Estas teses trazem modificações à anterior concepção da interpretação radical. O propósito continua a ser compreender como podemos 'ler' mente e significado a partir do comportamento de agentes no mundo, no entanto um novo elemento aparece: a intersubjectividade. A intersubjectividade é trazida à consideração através do conceito de triangulação. Situações de triangulação são situações em que dois agentes reagem coordenadamente entre si e relativamente a um objecto terceiro no mundo. Segundo Davidson, elas estão presentes desde o comportamento animal até à comunicação linguística entre humanos. Davidson descreve dois tipos de triangulação: i) pré-conceptual e pré-linguística, ii) conceptual e linguística. Entre a primeira e a segunda triangulação dá-se a emergência do pensamento (o pensamento característico do nosso tipo de mentes: pensamento acerca de um mundo objectivo, envolvendo os conceitos de crença e de verdade). O problema é que Davidson descreve os dois tipos de triangulação mas não avança hipóteses acerca do que permitiria a passagem entre uma e outra (quando essa passagem obviamente acontece – ela ocorreu por exemplo em cada um de nós). Neste artigo parto de interpretações dos estudos da atenção conjunta na ciência cognitiva para a formulação de hipóteses acerca de tal passagem. O resultado é uma crítica ao uso que Davidson faz do conceito de crença na teoria da mente.

Abstract: In his last writings, Davidson develops a set of theses around the concept of triangulation which bring changes to the way radical interpretation was formerly conceived. While his aim is still to understand how it is possible to read mind and meaning out of agents' behaviour, the main difference between earlier formulations of radical interpretation and these late views is the role of intersubjectivity. Intersubjectivity is brought in through the concept of triangulation. Davidson calls 'triangulations' those settings in which two agents react in a coordinate way to each other and to a third element in their common environment. According to Davidson, triangulations structure a range of situations, from animal behaviour to linguistic communication among humans. Davidson describes two kinds of triangulation: (i) pre-conceptual and pre-linguistic, (ii) conceptual and linguistic. Between the first and the second triangulation, he situates the emergence of thought (thought as it is characteristic of our kind of minds, i.e. thought about the objective world, involving the concepts of *belief* and *truth*). Still, he puts forward no hypothesis as to what explains it that a creature may change from being a participant in the first kind of triangulation to being a participant in triangulations of the second kind (and this, we know, happens – it happened, namely, in each one of us). In this article, I start from interpretations of studies of joint attention in cognitive science to explore a hypothesis concerning the change from the first to the second kind of triangulation. The upshot is a criticism of the use Davidson makes of the concept of belief in his theory of mind.

[1] Membro e investigadora do *Mind Language and Action Group* – MLAG – do Instituto de Filosofia da Universidade do Porto e Professora do Departamento de Filosofia da Faculdade de Letras da Universidade do Porto

Introdução

Donald Davidson é um dos autores de referência do nosso Projecto de Investigação (*Rationality, Belief, Desire II – from cognitive science to philosophy*), e por essa razão temos dedicado atenção a vários aspectos específicos da sua obra. O meu propósito de fundo neste artigo é criticar a forma como Davidson utiliza o conceito de crença na teoria da mente. Escolhi um ângulo específico: a intersubjectividade, i.e. a relação entre duas mentes. Não entendo aqui 'intersubjectividade' como *mindreading,* embora as relações mente-mente tenham sido anteriormente tratadas desse modo no âmbito do Projecto, com a análise do contraste entre as abordagens *theory theory* e *simulation theory* da capacidade humana de 'ler mentes' no mundo [2]. A questão é ainda a das relações entre uma mente e outra mente, no entanto para Davidson ela é mais profunda do que as questões acerca de arquitectura cognitiva de agentes que subjazem à alternativa entre *theory theory* e *simulation theory*. A diferença da perspectiva de Davidson reside no facto de ele relacionar a intersubjectividade com a própria possibilidade de pensamento objectivo acerca do mundo. Essa é uma das questões de fundo do presente artigo.

Enquanto estudo da obra de Davidson, este artigo segue-se a *Why Can't There be a Science of Rationality?*[3]. Nesse artigo a minha referência era a Teoria Unificada do Pensamento, Significado e Acção. A questão da triangulação de que aqui trato foi introduzida por Davidson apenas posteriormente. Com ela Davidson pretende, em traços largos, corrigir o 'exteriorismo' da anterior formulação da sua teoria da interpretação – convém aqui sublinhar que quer a interpretação radical de Quine quer a tradução radical de Davidson são teorias da interpretação exterioristas tanto quanto assumem que a teoria da mente começa pela terceira pessoa. O que isto significa é que se considera que a única evidência disponível para o teórico da mente é o comportamento de agentes no mundo; mente e significado vêm a ser atribuídos depois. As teses desenvolvidas em torno da ideia de triangulação constituem o princípio de uma alternativa a essa perspectiva exteriorista.

---

[2] Cf. Nichols e Stich 2003, *Mindreading*, Oxford University Press e Clara Morando 2005 *Intelectu nº 11.*

[3] Miguens 2006, in Miguens, Pinto e Mauro 2006, *Análises,* Porto, Faculdade de Letras da Universidade do Porto, pp. 91-98.

1. Estudos empíricos da atenção conjunta e triangulação davidsoniana

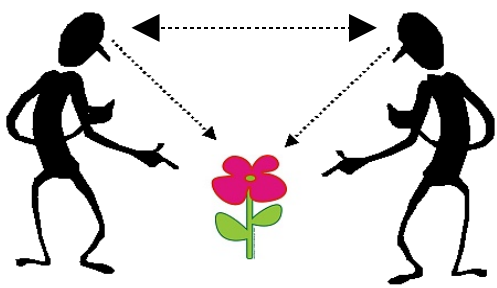
Na verdade, o que me interessa aqui fundamentalmente não é tanto Davidson mas a forma como se utiliza o conceito de crença na teoria da mente (quando falo de 'teoria da mente' penso quer na filosofia quer na ciência cognitiva). Noutras palavras, interessa-me compreender o espaço entre o comportamento de uma criatura, observável por outra criatura, e a atribuição de um interior mental, atribuição feita por criaturas com vidas cognitivas diversas, humanas e não humanas. Ora, nós temos aqui um preconceito de partida, que tem a ver com a forma como nós próprios 'lemos mente' no comportamento de outras criaturas, e devemos ter consciência dele. Como Michael Tomasello, o psicólogo do Max Planck Institut for Evolutionary Psychology que ganhou o Prémio Nicod de filosofia da mente este ano (2006) no âmbito dos seus estudos com chimpanzés, gosta de sublinhar, para nós, humanos, é até difícil imaginar que outros organismos possam observar o comportamento de membros da mesma espécie, e de humanos, e não o compreender em termos intencionais. É estranho imaginar criaturas que percebem corpos e movimentos no mundo, e que não vêem aí imediatamente acções, intenções, crenças, desejos. Mas aparentemente isso é possível, nomeadamente em indivíduos de outras espécies. Então, o que explicará que se possa ler ou não interior mental em comportamento observável? Que pistas são eficientes? O que é que tem que estar na mente da criatura que percebe para que exista um apercebimento mentalista do comportamento observado na outra criatura? Parece claro que algo de natural em nós, humanos adultos, não está ainda aí noutros agentes – mas será isso razão suficiente para defender, como Davidson defende, que apenas os humanos têm crenças, e as utilizam para ler o comportamento de outras criaturas, porque apenas os humanos são criaturas linguísticas na posse do conceito de crença? É disto que vou tratar, e vou argumentar que não.

A minha estratégia será apoiar-me nas interpretações que vários filósofos têm feito de estudos empíricos do fenómeno da atenção conjunta para criticar a concepção – puramente filosófica – da triangulação em Davidson. A atenção conjunta é um fenómeno estudado na ciência cognitiva por psicólogos e primatologistas. As interpretações filosóficas dos estudos da atenção conjunta nas quais me baseio são as de N. Eilan (2005), C. Peacocke (2005) e J. Campbell (2005). As interpretações destes autores não coincidem; vou utilizar as

que me interessam para criticar Davidson. Avanço desde já que, na terminologia de Eilan, o problema de Davidson reside na forma como trata a primeira triangulação, que o coloca do lado daquelas a que Eilan chama 'explicações pobres da atenção conjunta', quando os estudos empíricos oferecem boas razões para uma interpretação 'rica' do fenómeno. Parece-me que Eilan tem razão e que ela localiza uma das insuficiências importantes da teoria da mente defendida por Davidson.

2. Davidson, triangulações e linguagem.

Na última fase da sua obra, Davidson (especificamente nos ensaios sobre intersubjectividade reunidos no volume *Subjective, Intersubjective, Objective*[4]) desenvolve um conjunto de teses acerca da natureza do pensamento em torno da ideia de triangulação. Situações de triangulação são situações em que dois agentes reagem coordenadamente entre si e relativamente a um objecto terceiro no mundo (Davidson fala de «respostas mútuas e simultâneas de duas ou mais criaturas a um estímulo distal comum e às respostas da outra»[5]). Segundo Davidson, estas triangulações ocorrem desde o comportamento animal (os dois exemplos que usa em *The Emergence of Thought* são de cardumes de peixes e vocalizações de macacos) até à comunicação linguística entre humanos.



Davidson pensa que existem dois tipos de triangulação, i) pré-cognitiva e pré-linguística, ii) conceptual e linguística. A primeira envolve animais não humanos e crianças, a segunda apenas humanos

[4] Esses ensaios são *Rational Animals*, *The Second Person* e *The Emergence of Thought* (Davidson 2001b, Davidson 2001c, Davidson 2001d).
[5] Davidson 2001 a: xv.

com domínio de uma linguagem. Entre a primeira e a segunda triangulação dá-se aquilo a que Davidson chama a 'emergência do pensamento'. Davidson pensa que a linguagem tem aqui um papel essencial.

Convém enunciar este ponto claramente: Davidson defende que *apenas certas mentes são capazes de pensamento*. Ele não pode, obviamente, estar a falar de percepção, ou de cognição em termos mais gerais: quando diz 'pensamento' refere-se a algo que é possível no nosso tipo de mentes e não em todos os tipos de mentes. E o que é esse algo? De acordo com Davidson é pensamento objectivo *tornado verdadeiro por um mundo independente daquilo que o pensador pensa*. Segundo Davidson, o pensamento objectivo requer nas mentes que dele são capazes a presença dos conceitos de crença e de verdade, e é por essa razão que ele defende que o pensamento objectivo existe apenas em mentes que estão entre si na 'situação griceana' tornada possível pela comunicação linguística. Mentes que estão entre si na 'situação griceana' tornada possível pela comunicação linguística são mentes capazes de atribuir crenças, e crenças acerca de crenças, a outras mentes, e que são por isso capazes da intenção característica da comunicação linguística, que é a intenção de ter a sua intenção reconhecida[6].

A emergência do pensamento tem como condições necessárias duas triangulações, que Davidson caracteriza da seguinte forma: 1) A primeira triangulação acontece quando uma criatura correlaciona as suas reacções a um fenómeno exterior com as reacções de outra criatura; nesta posição espera o fenómeno exterior quando percebe a reacção da outra; é o facto de a expectativa poder falhar que introduz a possibilidade de erro na representação de alguma coisa.

A primeira triangulação é necessária mas não suficiente para a existência de pensamento objectivo: a situação descrita é possível entre agentes que não dispõem do conceito de crença, e que são incapazes de imputar crenças a outros agentes. Ora para Davidson o conceito de crença é, também ele, condição necessária do pensamento objectivo. A segunda triangulação envolve mentes dotadas do conceito de crença, e

[6] A exigência de intenções griceanas embebidas numa situação comunicacional significa que um acto comunicacional é bem sucedido apenas se a intenção com que é praticado é reconhecida. Não é isso que se passa com qualquer um dos nossos actos intencionais; frequentemente o que estamos a fazer pode ser bem sucedido independentemente de qualquer reconhecimento por outro agente. Um ponto básico da filosofia da comunicação é, assim, que a comunicação é um jogo de coordenação tácita que necessariamente envolve mais do que um agente.

Davidson caracteriza-a em termos genericamente griceanos. É ela que estabelece aquilo a que Davidson chama a sua pretensão wittgensteiniana: o 'carácter social da linguagem e do pensamento'. Esta pretensão diz respeito, portanto, apenas a mentes que se encontram entre si numa situação griceana de intenções embebidas, e portanto a mentes linguísticas de humanos.

Note-se que uma vez presente numa mente, o conceito de crença tem para Davidson duas funções: (i) aquela que mais frequentemente se refere, que é permitir interpretar o comportamento de outra criatura, (ii) uma menos frequentemente sublinhada, que é ser o veículo para o sujeito capturar o conceito de verdade objectiva[7].

Vamos então ao que me interessa verdadeiramente aqui. Tudo o que Davidson afirma acerca das triangulações supõe uma forma específica de conceber a segunda pessoa, i.e. o segundo vértice do triângulo, a segunda criatura em relação com a primeira. Para Davidson, a mutualidade desta relação só pode ser caracterizada em termos do conceito de crença (numa criatura existem crenças acerca das crenças da outra, tomando como evidência o comportamento). Mas será mesmo que toda a mutualidade exige representações de segunda ordem deste tipo, supostamente existentes apenas em mentes linguísticas? Não existirá algum antecessor que desempenhe um papel análogo ao papel da crença no apercebimento mútuo de criaturas e no envolvimento deste na ideia de mundo objectivo? Uma coisa é certamente estranha: Davidson descreve os dois tipos de triangulação mas não avança qualquer hipótese acerca do que permite, pelo menos em algumas mentes, a passagem entre uma e outra. Chega a dizer que não vê como teríamos vocabulário para falar sobre essas coisas e que não gostaria nada de trabalhar no campo da psicologia do desenvolvimento. Mas a verdade é que tal passagem acontece – aconteceu em cada um de nós – e podemos procurar compreendê-la por outras vias, que não a da análise conceptual. Daí o interesse dos estudos da atenção conjunta. Mas antes de passar a estes quero dizer um pouco mais acerca da forma como Davidson se serve da triangulação para conceber as origens da linguagem (afinal, esse é um interesse coincidente com o dos psicólogos que estudam a atenção conjunta). Para Davidson é a

[7] «A não ser que a linha de base do triângulo, a linha que liga os dois agentes, seja fortalecida até ao ponto em que é possível implementar a comunicação de conteúdos proposicionais, não há forma de fazerem uso da situação triangular para formarem juízos acerca do mundo. Só quando a linguagem está presente é que as criaturas podem apreciar o conceito de verdade objectiva», Davidson 2001d: 130.

linguagem que faz a grande diferença entre a primeira e a segunda triangulação; a linguagem faz uma diferença profunda, no que respeita a tornar as mentes humanas propriamente humanas, e os humanos 'animais racionais'. A linguagem de que Davidson fala aqui não é um objecto abstracto, definido por uma lista finita de expressões, regras para construir concatenações e uma interpretação semântica; o que está em causa são as línguas naturais, que as crianças começam a falar, e elas não vêm todas de uma vez. Davidson defende, como se sabe, que «There's no such thing as a language»[8], i.e. que não existem coisas tais que sejam línguas; estritamente falando, não há sequer duas pessoas que falem a mesma língua: significados não existem fora de práticas humanas, apenas existem comportamentos linguísticos de indivíduos[9].

É precisamente para compreender como é possível começar a imputar algum tipo de significação a acontecimentos tais como elocuções ou marcas que por si próprios, intrinsecamente, nada significam que a segunda pessoa é essencial. É a segunda pessoa e não uma suposta língua exterior existindo abstractamente, como o Inglês ou o Português, que é essencial para a forma como uma mente se torna – e é – uma mente linguística (tão essencial que Davidson chega a dizer que «se tu e eu fossemos as únicas pessoas no mundo e eu falasse inglês e tu sherpa, ainda assim poderíamos entender-nos um ao outro e comunicar»[10]). Para explicar porquê, em *The Second Person* Davidson introduz a triangulação para falar da forma como criaturas aprendem a responder de forma específica a estímulos específicos com palavras, e portanto a referir objectos no mundo. É ilustrativo que use os dois exemplos que usa: um exemplo de comportamento animal condicionado e um comportamento de aprendizagem de linguagem num humano. Um cão saliva quando ouve a campainha, uma criança chama 'mesa' a uma mesa. Nós achamos muito natural chamar 'estímulo' à campainha ou à mesa (estímulos distais) e não à estimulação das terminações nervosas do cão e da criança, na periferia de cada uma das criaturas. E é de facto natural. Davidson defende que é a triangulação que explica isso: achamos natural que seja assim para o cão e para a criança porque é natural *para nós*. O caso da criança que aprende a chamar 'mesa' à mesa envolve uma forma de triangulação: uma linha

[8] Cf. *A Nice Derangement of Epitaphs*, Davidson 2005: 107.

[9] «o principal propósito do conceito de língua e conceitos adjacentes como predicado, frase e referência, é ajudar-nos a entender o comportamento linguístico dos humanos e aquilo que falantes intérpretes sabem que lhes permite comunicar», Davidson 2001c:109.

[10] Davidson 2001c:114.

vai da criança à mesa, outra de nós à mesa, outra entre nós e a criança. Aí onde as linhas convergem, é localizado o estímulo distal – se considerássemos uma criatura por si, nos seus limites corporais, não teríamos nenhuma razão para dizer que ela está a reagir a mais do que à estimulação que chega à superfície do seu corpo e às suas terminações sensoriais.

3. Atenção conjunta e conhecimento mútuo ou tácito

Passo então aos estudos sobre atenção conjunta, nomeada- mente os que visam a primeira triangulação, entre mentes não linguísticas[11], e que nos permitem imaginar um passo intermédio entre a primeira e a segunda triangulações (ou, de uma forma mais concreta, entre comportamentos animais de coordenação e mentes em comunicação linguística acerca do mundo). Sob o título de atenção conjunta esta primeira triangulação tem sido bastante estudada em crianças, e também no domínio da psicologia animal, por exemplo em estudos com chimpanzés acerca do que é perceber uma outra criatura como vendo.[12]

O indício comportamental mais óbvio de atenção conjunta é o acto de uma criatura seguir o olhar de outra criatura dirigido a algo no ambiente: nós somos criaturas que seguem o olhar e a atenção de outras criaturas, criaturas capazes de levar outras criaturas a seguir o nosso olhar e partilhar a atenção a um objecto terceiro no mundo. Os psicólogos do desenvolvimento começaram a estudar a atenção conjunta em grande medida para compreender as origens da comunicação verbal nos humanos. Quem está interessado em atenção conjunta está usualmente interessado no desenvolvimento da linguagem, e estuda assuntos tais como: i) o contraste entre apontar imperativo e apontar declarativo em chimpanzés e humanos (é consensual que por exemplo chimpanzés, ao contrário de crianças humanas, não produzem o apontar declarativo que as crianças caracteristicamente produzem, mas apenas apontar imperativo[13]; ii) as etapas da direcção do olhar das crianças humanas nos primeiros meses do desenvolvimento (após olhares dirigidos apenas para quem cuida, há olhar dirigido a objectos no meio, depois os olhares oscilam entre

[11] A atenção conjunta não acontece apenas entre mentes não linguísticas, mas estende-se a estas.
[12] Cf. Call & Tomasello 2005, Povinelli & Eddy 1996.
[13] A diferença é formulável como: 'Quero!' versus 'Olha!' – o primeiro ocorre em chimpanzés, o segundo aparentemente apenas em humanos.

objectos e adultos, e finalmente dá-se uma convergência entre apontar e seguir o olhar – palavras em línguas naturais humanas começam a ser usadas apenas aqui).

Não vou entrar em discussões específicas dos estudos empíricos da atenção conjunta, apenas notar que existe um acordo relativo acerca de dados – o diferendo é acerca de interpretações. Tomemos Tomasello. Ele sugere o seguinte: dá-se, entre os 9 e 18 meses, uma revolução cognitiva comparável com aquela que se dá entre os 3 e 4 anos de idade numa criança (esta é a altura em que emerge o conceito de crença, e portanto o de crença falsa, a ideia de que os outros podem ter crenças diferentes das nossas próprias, e isso passa a poder ser utilizado na previsão de comportamento de outrem). Aquilo que se desenvolve nesta primeira revolução é um entendimento baseado na agência, e um progresso da ideia de que os outros seres têm intenções para a ideia de que os outros seres podem ter intenções diferentes das nossas, eventualmente intenções que não correspondem (*match*) a estados de coisas actuais (intenções não preenchidas, não satisfeitas). Para tal revolução é necessário o desenvolvimento do raciocínio meios/fins, a distinção entre fins e meios para os atingir, a aplicação disto à acção própria e por simulação à de outrem. É a aplicação deste mecanismo cognitivo ao comportamento de olhar intencionalmente, o comportamento próprio e o comportamento de outrem, que permite que a atenção comece a ser pensada de forma nova[14]. Aos 12 meses, as crianças começam a procurar manipular e controlar o comportamento de atenção de outrem, apontando e articulando as primeiras palavras. Estas manipulações da atenção constituem uma forma primitiva de intenção comunicativa. É uma forma primitiva porque utiliza apenas a agência e a atenção, e não o conceito de crença. No entanto, tem a estrutura embebida das intenções griceanas, que não se encontra noutros primatas[15].

Para compreendermos que há aqui qualquer coisa de novo, que tem a ver com algo como a instauração de uma 'ligação entre mentes', pensemos nos seguintes casos, estudados na psicologia animal. Se tomarmos duas criaturas dotadas de visão e que estão dentro do campo

[14] É claro que isto envolve saber o que se entende por atenção. Tomasello assume que a atenção é concebida pelas crianças como percepção intencionalmente dirigida, estando o seu uso ligado ao entendimento dos outros como agentes capazes de acções intencionais. Esta posição não resolve obviamente todos os problemas em torno da atenção – será por exemplo que ela é concebida como uma actividade mental?

[15] Cf. Call & Tomaselli 1995: 61.

de visão uma da outra, será que podemos afirmar com segurança que elas não apenas se vêem uma à outra, mas também sabem que se vêem uma à outra? A resposta é negativa, ou, pelo menos, não podemos fazer essa afirmação sem hesitar e sem a 'atenuar': basta recordar as experiências de Povinelli e Eddy[16], em que chimpanzés dirigem pedidos a humanos que estão de frente para eles mas de olhos vendados, semelhantes aos que dirigem a humanos com quem estão face a face e que olham olhos nos olhos, mas não os dirigem àqueles que estão de costas mas com o rosto voltado de forma a poderem vê-los.

São fenómenos deste género que permitem conceber um hiato entre essas mentes e as nossas; mas o ponto importante aqui é que crianças pequenas têm performances completamente diferentes (bem sucedidas) em tarefas de atenção conjunta, relativamente a outras mentes não linguísticas. A hipótese que parece impor-se é que nos humanos se dá qualquer coisa como um desenvolvimento da mutualidade mesmo antes da posse do conceito de crença. À luz das diferenças entre primatas e crianças humanas torna-se então razoável pensar que a atenção conjunta está ligada a uma forma de estratégia intencional que não utiliza ainda o conceito de crença, e que é fundamental no desenvolvimento do tipo humano de mentes.

Vou agora introduzir um caso de atenção conjunta mais complicado e uma definição preliminar (embora polémica), para passar à discussão filosófica. O exemplo é de J. Campbell (filósofo, UC-Berkeley), a definição de Naomi Eilan (filósofa, Warwick).

Supõe que estás sentado num banco de um parque a olhar para um cisne. Alguém se senta a teu lado, uma conversa inicia-se, durante a qual ambos observam o cisne. De alguma forma passaste de uma situação em que tu tinhas percepção do cisne para uma situação em que há apercebimento não apenas do objecto apercebido mas apercebimento de que o objecto está a ser simultaneamente apercebido por outrem. O problema que se coloca é o seguinte: será que ocorreu alguma mudança relevante na tua experiência perceptiva do cisne? Se for esse o caso, como se deve descrever tal mudança?

Antes de discutir alternativas de resposta, vou introduzir a definição preliminar que Eilan dá de atenção conjunta (até aqui falei de forma psicológica). Afirmar de um evento que se trata de um evento no qual existe atenção conjunta de dois (ou mais) sujeitos a um mesmo objecto implica estar comprometido com as seguintes quatro pretensões

[16] Povinelli & Eddy 1996.

acerca do evento:

1. Existe um objecto ao qual é dirigida a atenção de cada um dos sujeitos, o que implica (i) uma conexão causal entre o objecto e cada sujeito, e (ii) apercebimento (*awareness*) do objecto por cada sujeito.

2. Existe uma conexão causal de algum tipo entre os actos de atenção dos dois sujeitos.

3. Na sua experiência os sujeitos utilizam o conceito de atenção.

4. Cada sujeito apercebe, de alguma forma, o objecto como um objecto que está presente para ambos os sujeitos. Há, quanto a isto, um encontro de mentes entre os sujeitos, de tal forma que o facto de ambos estarem a ter atenção ao mesmo objecto é mutuamente manifesto (o termo é utilizado, e isto não é indiferente, por Sperber e Wilson na sua teoria da comunicação[17]).

Não é suficiente para que exista atenção conjunta que duas criaturas prestem atenção ao mesmo objecto e que a atenção de uma seja a causa da atenção da outra: o facto de ambos estarem a prestar atenção ao mesmo objecto tem que ser mutuamente manifesto. O seguinte exemplo de Peacocke é bem ilustrativo: «Consideremos duas pessoas que estão em pé, em frente uma à outra, separadas por um espesso painel de vidro. Suponhamos que cada uma delas acredita falsamente que este vidro é um espelho numa só direcção, que lhe permite ver o outro, mas impedindo o outro de a ver a ela. Então cada uma realmente vê a outra, ao mesmo tempo que acredita que a outra pessoa não pode vê-la. (…) Da mesma forma, vamos supor que nesta situação, ambas estão a prestar atenção a alguma coisa – um animal, por

[17] Sperber e Wilson consideram que a partilha de informação é essencial para que a comunicação possa existir, e é a tentativa de explicitar em que pode consistir essa partilha que os leva a falar de 'manifestação mútua'. Não apenas factos perceptivos como também suposições ou crenças de vária ordem podem ser, segundo os autores, mutuamente manifestas. A noção 'ser manifesto' é mais fraca do que a noção 'ser conhecido', e por isso mesmo os autores consideram poder desenvolver uma noção de manifestação mútua que seja psicologicamente menos implausível do que a noção de 'conhecimento mútuo'. Cf. Sperber & Wilson 2001: p. 78 e seguintes.

exemplo – no seu horizonte visual comum, de um dos lados do vidro que está entre elas. Cada uma pode ter uma percepção genuína da outra prestando atenção exactamente à mesma coisa a que ela está a prestar atenção, nomeadamente o animal. Mas porque cada uma acredita que a outra não a pode ver, isto está longe de ter a abertura característica dos casos paradigmáticos de atenção conjunta.»[18].

Nesta situação não existe atenção conjunta porque falta a manifestação mútua. O núcleo do problema epistemológico da atenção conjunta é a natureza desta manifestação mútua. Para o tratar é decisivo saber que conceitos estão disponíveis nas mentes que interagem. Tudo nos leva a crer que a atenção conjunta pode dar-se envolvendo mentes não linguísticas, que ainda não possuem o conceito de crença, como as mentes de crianças de dois anos. Ora se a atenção conjunta acontece envolvendo, como parece ser o caso, crianças com dois anos (mentes pré-linguísticas sem o conceito de crença), então o fenómeno não pode ter as características daquilo a que os filósofos chamaram conhecimento comum ou tácito[19]. Este é frequentemente considerado a chave para compreender a comunicação linguística entendida como actividade racional de cooperação e mais em geral qualquer actividade racional de cooperação. As questões epistemológicas que coloca são muito semelhantes às da atenção conjunta. No entanto as suposições usuais no tratamento do conhecimento tácito não são transponíveis para aqui. Tomemos um exemplo clássico na literatura de conhecimento mútuo: tu e eu estamos sentados a uma mesa com uma vela entre nós. Numa tal situação, em condições normais, teremos conhecimento mútuo do facto de que ambos vemos a vela. Em que é que isto se traduz? Como pode ser analisado? Uma análise filosófica do conhecimento mútuo tipicamente atribui-me pelo menos a crença de que tu vês a vela, a crença de que tu acreditas que eu vejo a vela, a crença de que tu

[18] Peacocke 2005: 299.

[19] Cf. por exemplo D. Lewis *Convention (1969*) e S. Schiffer *Meaning* (1988). O conhecimento comum permite--nos interagir com outros humanos sem sobressalto, assumindo coisas que não vêm directa e explicitamente ao caso nas situações concretas, tais como que as pessoas em Portugal falam português e usam euros como moeda (imagine-se que abordamos alguém na rua em Lisboa («Olhe, desculpe…»), e a pessoa se volta para nós a falar convictamente uma língua que nos é totalmente ininteligível; ou que vamos ao supermercado e a pessoa da caixa nos quer dar o troco em rublos, como se fosse a coisa mais normal do mundo). O sentido em que pessoas têm estas crenças é contrafactual: elas inferi-las-iam a partir das crenças que têm, por princípios que aceitam. Note-se que isto é qualquer coisa que se pode dizer acerca do conhecimento mas não se pode dizer acerca da percepção.

acreditas que eu acredito que tu vês a vela,... [20]. Não precisamos de perguntar onde pára esta iteração para notar que se a atenção conjunta é possível entre mentes não linguísticas que não possuem o conceito de crença ela não poderá ser descrita de forma análoga. E um dado incontornável da psicologia para a filosofia é que crianças com menos de quatro anos não dispõem do conceito de crença; sabemos isso porque elas falham no teste das crenças falsas. No entanto parecem capazes de atenção conjunta. Temos certamente a opção de defender que é por isso mesmo que não são capazes de atenção conjunta plenamente desenvolvida (é a linha que Tomasello escolhe). Podemos também procurar investigar se será possível qualquer coisa como um apercebimento que não tenha a estrutura reflexiva do conhecimento mútuo.

Posso agora formular o meu problema da seguinte maneira. Se a atenção conjunta ocorre envolvendo não apenas mentes linguísticas em situação griceana de comunicação mas também mentes pré-linguisticas, isso coloca a hipótese de existir algum tipo de apercebimento mútuo em mentes pré-linguísticas que não tem características griceanas explícitas – a ser esse o caso, quais serão as implicações para os problemas que Davidson trata através da triangulação[21]?

O que me parece que os estudos da atenção sugerem é que existem diferentes componentes da compreensão de um comportamento como intencional – não apenas crença mas agência e atenção – e que eles não vêm todos ao mesmo tempo nem têm que estar todos lá, em todos os tipos de mentes, para se perceber uma outra parte do mundo como mental. Voltemos ao exemplo de Campbell. A resposta à questão (ocorreu alguma mudança relevante na tua experiência perceptiva do cisne?) diferirá conforme a teoria da atenção conjunta defendida[22]. No

[20] Convém notar que para Peacocke (Peacocke 2005) este não é um fenómeno de conhecimento mútuo, na medida em que para o tratar não é necessário todo o aparato adstrito a este.

[21] Os problemas são, recordo, a emergência do pensamento objectivo e a importância da segunda pessoa nessa emergência.

[22] Nos termos de Campbell, uma teoria experiencialista ou uma teoria não experiencialista, podendo uma teoria experiencialista ser relacional ou reducionista. Cf Campbell 2005. De acordo com uma teoria não experiencialista, a experiência perceptiva de um indivíduo é a mesma quando este presta, isoladamente, atenção a alguma coisa e quando o faz numa situação de atenção conjunta. De acordo com uma teoria experiencialista, dá-se uma alteração da própria experiência perceptiva quando se passa da atenção isolada para a atenção conjunta. No caso das teorias experiencialistas, Campbell fala ainda de uma perspectiva reducionista (se é possível identificar quais os estados individuais que importam aqui) e de uma perspectiva relacional (se a atribuição

entanto, o que está basicamente em causa é a descrição que queremos adoptar da forma como uma mente apercebe a atenção de outrem.

A opção de Tomasello é atribuir ao conceito de atenção uma função análoga ao papel explicativo da crença relativamente ao comportamento. No entanto, porque considera que as coordenações que produzem apercebimento mútuo genuíno são uma forma de cooperação racional, declara que elas não estão ainda presentes na mente da qual está ausente o conceito de crença. Mas já que estamos a falar da forma como a atenção 'precede' a crença na função explicativa desta, devemos recordar que o conceito de crença tem para Davidson não apenas uma mas duas funções: (i) interpretar o comportamento de outrem, (ii) ser o veículo para o sujeito capturar o conceito de verdade objectiva. Vemos então que Tomasello apenas considera a primeira função; o que ele não pergunta é se existe algum 'antepassado' da segunda função, i.e. do papel da crença no que diz respeito ao conceito de verdade objectiva na vida mental da criança. Segundo N. Eilan, H. Werner e B. Kaplan foram dos primeiros psicólogos a explorar essa ideia: a ideia é ver as triangulações da atenção conjunta como uma primeira manifestação de uma estratégia contemplativa, por oposição a uma estratégia puramente prática na forma de lidar com o mundo, de forma a que o acto de referência emerge como um acto social e não individual. Não a exploraram no entanto suficientemente[23].

Em que é que tudo isto poderia redundar numa crítica a Davidson? Antes de mais convém dizer claramente que o interesse dos estudos de fenómenos de atenção conjunta envolvendo mentes não linguísticas reside no facto de estes parecerem revelar aspectos da relação mente-mente que Davidson situa apenas na segunda triangulação. Isto abre a possibilidade de elaborar uma teoria da relação mente-mente que não conceba a 2ª pessoa como derivada da forma como conceitos mentais são utilizados em explicações de 3ª pessoa. A ideia de Eilan é que os fenómenos da atenção conjunta nos ajudam a fazer isso mesmo.

Queria agora recordar o exemplo de atenção conjunta e a definição preliminar de Eilan e introduzir a diferença entre as interpretações que John Campbell e Christopher Peacocke fazem do fenómeno da atenção conjunta e da natureza da manifestação mútua. O que está em causa é uma disputa acerca dos pontos 3 e 4 da definição, e do mecanismo que

---

dos estados psicológicos relevantes a x já implica que exista alguém com quem x presta atenção em conjunto a algo). Na perspectiva experiencialista relacional, a atenção conjunta é um fenómeno primitivo da consciência.

[23] Cf. Eilan 1995, para a referência a W. Werner e B. Kaplan 1963, *Symbol Formation*.

os relaciona. Recordo os pontos 3 e 4: 3. Na sua experiência os sujeitos utilizam o *conceito de atenção*; 4. Cada sujeito apercebe, de alguma forma, o objecto como um objecto que está presente para ambos os sujeitos. Há, quanto a isto, um encontro de mentes entre os sujeitos, de tal forma que o facto de ambos estarem a ter atenção ao mesmo objecto é *mutuamente manifesto*).

A alternativa diz basicamente respeito à riqueza relativa do input e dos mecanismos cognitivos. Para Peacocke, o input perceptivo é percepção da percepção que o outro tem do objecto; é necessária uma reflexão de segundo grau para pôr a funcionar a relação entre as duas percepções (Peacocke defende ainda que para existir atenção conjunta plenamente desenvolvida – *full blown*, tal como não existe numa criança de dois anos e existe num humano adulto, são necessários pensamentos auto-referenciais). Percepção da percepção que o outro tem do objecto e reflexão de segundo grau necessária para pôr a funcionar a relação entre as duas percepções é também o que M. Tomasello considera ter que existir na atenção conjunta em adultos; é por isso que defende que crianças com dois anos não são capazes de atenção conjunta plena. Na teoria (experiencialista) relacional da atenção conjunta defendida por Campbell, o outro em co-atenção ao objecto é um traço de conteúdo da experiência de atenção conjunta e é um fenómeno primitivo da consciência e não uma meta-representação (o apelo à reciprocidade da regulação afectiva pode ser uma forma de especificar isto). Temos que ser claros aqui pelo menos acerca do seguinte: só precisamos de reflexão de segundo grau e do conceito de crença se assumirmos à partida a opacidade das mentes umas às outras e portanto a incontornabilidade da postura explicativa para tratar da questão do conhecimento da outra mente. Mas teremos boas razões para a assumir? O interesse das experiências com atenção conjunta reside no facto de elas parecerem indicar que não é só isso que está em jogo na relação mente-mente[24]. E isso deve fazer-nos repensar a forma como concebemos a diferença entre mentes linguísticas e não linguísticas.

[24] Nos termos de N. Eilan, trata-se de uma escolha de caminho para responder à questão «Como é que a ideia de um mundo vem a estar numa mente?» Davidson e Kant representam alternativas básicas na resposta a essa questão: a resposta de Davidson é baseada na comunicação, a resposta de Kant é baseada na percepção. Se escolhermos, para responder à questão, o ramo davidsoniano, deparar-nos-emos com o tipo de críticas que estamos a considerar.

Conclusão

Penso que Davidson tem razão quando diz que é impossível resolver a questão dos conteúdos e da objectividade do pensamento relativamente a uma só criatura, e aos limites corporais da criatura, e que a triangulação é uma boa entrada no tratamento destes problemas. Mas as posições específicas de Davidson quanto à primeira e à segunda triangulação não são defensáveis.

Na forma como caracteriza as triangulações, Davidson coloca-se do lado de Peacocke e Tomasello, i.e. daquelas a que N. Eilan chama 'teorias pobres' (teorias de acordo com as quais em crianças com idade entre 1 e 2 anos apenas as condições 1) e 2) da definição preliminar de atenção conjunta se aplicam; noutras palavras, teorias de acordo com as quais o que há antes do pensamento objectivo não é essencialmente social, a mutualidade vem apenas depois, com a introdução do conceito de crença). Para as 'teorias pobres' a interpretação em 3ª pessoa continua a ser, mesmo no âmbito das propostas sobre triangulação, a base para qualquer sentido que se possa fazer das relações entre 1ª e 2ª pessoa: a evidência é o comportamento, e o intuito é explicativo (não é descabido recordar que Davidson é um quineano).

Não é por isso por acaso que Davidson não caracteriza a conexão entre os dois níveis de triangulação – ele não poderia fazê-lo dadas duas teses que defende: 1) a tese segundo a qual os nossos conceitos de atitudes proposicionais (crenças, desejos) obtêm significado em virtude de certos padrões de explicação em que esses conceitos entram todos ao mesmo tempo, de forma holista, 2) a tese segundo a qual todo o conteúdo representacional é conceptual e governado por constrangimentos de racionalidade. Mas se pensarmos que tem que haver uma conexão entre os dois níveis de triangulação, uma vez que algumas mentes se desenvolvem de forma a passar de um para o outro, teremos razões para pôr em causa essas duas teses de Davidson. Se admitirmos que do ponto de vista do desenvolvimento do tipo de mentes que são as mentes humanas há uma passagem e uma herança entre a primeira e a segunda triangulação, e que antecessores de apercebimento mútuo e de mundo objectivo, e não apenas coordenação causal, devem já estar lá em algumas formas de triangulação pré-cognitiva (as de atenção conjunta, inexistentes noutros animais), teremos que 'transferir' para a primeira triangulação algumas características que Davidson atribui apenas à segunda triangulação. A triangulação pré-cognitiva entre alguns tipos de criaturas tem que ser

mais rica do que Davidson está pronto a admitir. A nova forma de encarar a primeira triangulação mudará também a forma como vemos a segunda triangulação, bem como o estatuto da pretensão wittgensteiniana acerca da 'natureza social do pensamento' que a acompanha (a 'natureza social' do mental tem que ir mais fundo do que a linguagem natural).

É útil aqui perguntar o que Davidson pretende da segunda triangulação no seu tratamento das questões da intersubjectividade (que dizem respeito, como vimos, a 'animais racionais', 'segunda pessoa', 'emergência do pensamento'). Ele caracteriza através da segunda triangulação aquilo que existe em mentes linguísticas e não existe em mentes não linguísticas. Ora a caracterização ela própria presume que o conceito de crença é essencial ao papel da segunda pessoa no pensamento objectivo. Acontece que os estudos da atenção conjunta – por mais que Davidson diga que não temos, a partir das nossas mentes linguísticas já dotadas dos conceitos de verdade e crença, vocabulário para compreender estados de desenvolvimento anteriores, mesmo que tenham ocorrido em nós próprios – nos fazem pensar que de facto podem existir estádios intermédios de 'estratégia intencional'. De acordo com as interpretações ricas dos estudos da atenção conjunta, tais fenómenos evidenciam uma etapa intermédia da intersubjectividade em que o que mentes utilizam para obter mutualidade e como utensílio explicativo não é o conceito de crença. Que isto seja possível deve perturbar o quadro holista explicativo do comportamento de agentes a partir de conceitos mentais (crença, desejo, intenção). Mas sobretudo mostra que, por mais que Davidson enfatize, na última fase da sua obra, o relevo à intersubjectividade – a triangulação pretende ser uma alternativa ao 'exteriorismo' das teorias da interpretação (quer a interpretação radical de Quine quer formulações anteriores da tradução radical do próprio Davidson), quadros em que a teoria da mente começa pela terceira pessoa – Davidson continua a derivar a 1ª e 2ª pessoas da 3ª pessoa, e é em parte isso que nos obriga a admitir que todo o conteúdo representacional é conceptual, e que provoca uma restrição severa daquilo que entendemos por 'racionalidade'. Faz além do mais com que, o que parece estranho, os laços causais, presumivelmente perceptivos que supostamente estão na primeira triangulação, pareçam desaparecer na segunda triangulação quando entra em cena a pretensão wittgensteiniana acerca da natureza social da linguagem e do pensamento. Mas esses laços causais não podem pura e simplesmente desaparecer (uma crítica comum a Davidson é que na sua teoria da

mente a percepção 'desaparece' – o que não é estranho: ele é um quineano).

O problema geral da teoria da mente é conceber a passagem entre uma *mindless nature* e as explicações mentais. Este é um problema acerca de continuidade/descontinuidade. O principal defeito da ideia davidsoniana de triangulação é fazer-nos conceber esta passagem como um salto abrupto ligado à presença do conceito de crença, com os seus dois papéis, nas mentes linguísticas. Não creio que exista esse salto abrupto, e um argumento a favor desta tese é a existência de antecessores do conceito de crença na estratégia intencional. Mostrá-lo foi o meu principal propósito aqui.

Referências:

Call, J. & Tomasello, M., 2005, «What Chimpanzees Know about Seeing», in Eilan, Hoerl, McCormack & Roessler 2005

Campbell, John, 2005, «Joint attention and common knowledge», in Eilan et al 2005.

Davidson, D, 2001a, «Introduction», in Davidson 2001, *Subjective, Intersubjective, Objective, Oxford*, OUP.

Davidson, D, 2001b, «Rational Animals», in Davidson 2001, *Subjective, Intersubjective, Objective, Oxford*, OUP.

Davidson, D, 2001c, «The Second Person», in Davidson 2001, *Subjective, Intersubjective, Objective, Oxford*, OUP.

Davidson, D, 2001d, «The emergence of thought», in Davidson 2001, *Subjective, Intersubjective, Objective, Oxford*, OUP.

Davidson, D. 2005, «A Nice Derangement of Epitaphs», in Davidson 2005, *Truth, Language and History*, Oxford, OUP.

Eilan, Hoerl, McCormarck & Roessler 2005, Joint attention: communication and other minds, Oxford, OUP.

Eilan, «Joint Attention, Communication and Mind», in Eilan, Hoerl, McCormarck & Roessler 2005

Gomez, Juan Carlos, 2005, «Joint Attention and the notion of subject: insights from apes, normal children and children with autism», in Eilan, Hoerl, McCormarck & Roessler 2005

Lewis, David, 1969, *Convention*, Cambridge Mass, Harvard UP.

Peacocke, Christopher, 2005, «Joint attention: its nature, reflexivity and relation

to common knowledge», in Eilan, Hoerl, McCormarck & Roessler 2005.

Povinelli D.J. & Eddy T. J. 1996, What young chimpanzees know about seeing, *Monographs of the Society for Research in Child Development*, 61-3, 1-190.

Schiffer, Stephen, 1988, *Meaning*, Oxford, OUP.

Tomasello, M, 1995, Joint attention as social cognition, in Moore, C. & Dunham, P, (eds) *Joint attention: its origins and role in development*, Hillsdale, NJ, Erlbaum, 103-130.

Werner, H. & Kaplan, B., 1963, *Symbol formation*, Hillsdale, NJ, Erlbaum.

Emoções, Protoemoções e Racionalidade

Tomás Carneiro[1]

Resumo: Neste artigo procuro investigar o estatuto a atribuir ao background não cognitivo numa teoria filosófica da racionalidade. Para isso discuto nomeadamente as consequências dos resultados da psicologia evolucionista para o campo das teorias da racionalidade, sobretudo os trabalhos sobre emoções e racionalidade das emoções. Procuro ainda saber como será possível encontrar um critério normativo de racionalidade.

Abstract: In this article I analyse the status of the non-cognitive background of agents in a philosophical theory of rationality and discuss the implications of evolutionary psychology results, especially the work on emotions and on the rationality of emotion. I also investigate the possibility of finding a normative criterion of rationality for dealing with such issues.

[1] Membro e investigador do *Mind Language and Action Group* – MLAG – do Instituto de Filosofia da Universidade do Porto.

Introdução

Os dois problemas centrais deste artigo são os seguintes: 1) que estatuto devemos atribuir ao nosso *background* não cognitivo numa teoria filosófica da racionalidade?, e 2) como encontrar um critério normativo de racionalidade?

Na primeira secção, 1) "*Critérios de racionalidade: dois problemas levantados ao critério consequencialista*", procuro contextualizar o artigo no estado da arte fazendo alusão ao trabalho de Stephen Stich no sentido de trazer os resultados da psicologia evolucionista para o campo das teorias da racionalidade. Interessa-me sobretudo o seu trabalho sobre racionalidade das emoções, nomeadamente o critério consequencialista que Stich usa para avaliar a racionalidade e a irracionalidade das emoções.

Em seguida apresento duas objecções ao critério consequencialista. Numa primeira objecção critico Stich por este não deixar bem claro onde devemos parar a atribuição de racionalidade. Se seguirmos apenas o critério consequencialista, parece-me que nos vemos livres para atribuir racionalidade a fenómenos não cognitivos tais como algumas emoções básicas. Numa segunda objecção ao critério consequencialista defendo a necessidade de um meta-critério normativo com o qual possamos avaliar a racionalidade do agente, suas acções e processos cognitivos.

Na segunda secção, 2) "*Racionalidade derivada e avaliação de racionalidade*" pretendo responder a estes dois problemas. Nesse sentido procuro argumentar em favor de um maior enquadramento empírico/biológico das teorias da racionalidade e para isso (inspirado no conceito de "intencionalidade derivada" de John Searle) avanço com a atribuição de **racionalidade derivada** ao *background* não cognitivo dos nossos sistemas cognitivos (as nossas protoemoções). Essa atribuição pretende, por um lado, dar conta da importância desse *background* não cognitivo nos nossos processos cognitivos superiores, por outro lado pretende encurtar a distância entre esses dois pólos (mente e corpo) que poderão não estar tão distantes quanto normalmente se pensa.

Para responder à segunda objecção ao critério consequencialista defendo que para falarmos da racionalidade de um agente, de uma acção, ou de um processo cognitivo são necessários dois níveis de condições: **condições de ocorrência** e **condições normativas de avaliação**. Nas condições de ocorrência coloco tanto os componentes

proposicionais do sistema cognitivo do agente (i.e., o seu conjunto de normas, crenças, desejos, intenções e emoções), como o seu *background* pré-racional e não cognitivo, condição necessária para a ocorrência de fenómenos racionais. Neste artigo defino todos esses fenómenos não cognitivos que constituem o *background* sob o termo **protoemoções**. Nas **condições normativas de avaliação** coloco a rede de crenças e de atitudes proposicionais necessárias para que determinado fenómeno seja interpretado, comparado e avaliado quanto à sua racionalidade.

Quanto às condições normativas de avaliação defendo a necessidade de se encontrar uma norma pela qual se possa avaliar a racionalidade dos agentes e avanço um pequeno esboço de como essa norma poderá ser encontrada.

Aqui é importante frisar que estou a usar o termo racionalidade em dois sentidos. No sentido cognitivo, quando me refiro à racionalidade dos processos cognitivos e não cognitivos dos processos de raciocínio do agente (protoemoções, emoções, crenças desejos e intenções), e no sentido normativo, quando me refiro à avaliação da racionalidade das acções do agente. Ao longo do artigo procurarei ir deixando bem claro quando estou a falar de uma e quando estou a falar de outra, referindo-me à primeira como racionalidade cognitiva e à segunda como racionalidade normativa.

Secção I - Critérios de racionalidade: dois problemas levantados ao critério consequencialista

No seu livro "The Fragmentation of Reason" Stephen Stich apresenta-nos o seu critério pragmatista e consequencialista para a avaliação da racionalidade dos agentes e dos seus processos cognitivos. Segundo a teoria da racionalidade defendida por este autor os nossos processos cognitivos são instrumentos que os agentes desenvolveram ao longo da evolução natural e que servem para ajudá-los a atingir determinados fins. A ideia de Stich é que devemos encarar os nossos processos cognitivos como ferramentas mentais que devem ser avaliadas em função do sucesso que tenham em fazer cumprir os objectivos que as pessoas normalmente valorizam. Neste sentido, a racionalidade, a justificação e a verdade, tomadas de um ponto de vista normativo, ou deontológico, não têm qualquer valor para o agente cognitivo, dado que quando age, o agente não procura atingir crenças verdadeiras, ou ser racional, mas antes procura atingir determinados

objectivos, e é relativamente a esses objectivos (i.e. confirmando se foram ou não alcançados) que a racionalidade dos agentes deve ser avaliada. Para Stich, portanto, "todo o valor cognitivo é instrumental" (Miguens, 2004). Como vemos, Stich parece não fazer qualquer distinção entre a racionalidade cognitiva e a racionalidade normativa, a última determina a primeira, e parece-me ser esta a fonte dos problemas que mais à frente levanto ao critério consequencialista de Stich.

Para esta secção baseei-me em dois artigos escritos por Stich (Samuels, Stich & Tremoulet e Stich & Sripada?) onde este procura, no primeiro artigo, justificar a pertinência de um critério consequencialista de racionalidade, por oposição a um critério deontológico e, no segundo artigo, partindo desse critério consequencialista, atribuir racionalidade e irracionalidade a emoções. De seguida apresentarei duas objecções ao critério consequencialista, nomeadamente **a**) o facto de ao seguirmos estritamente esse critério consequencialista (ou seja, atribuindo racionalidade ou irracionalidade a um fenómeno **somente** em função das suas consequências), sermos obrigados a atribuir racionalidade a estados não cognitivos fundamentais para o normal funcionamento do nosso sistema cognitivo (reflexos, impulsos, motivações, o sistema de regulação metabólica e emoções de fundo como o prazer e a dor – agrupo todos estes fenómenos não cognitivos sob o termo geral de **protoemoções**) e **b**) a necessidade de um meta-critério normativo que nos dê um conceito de normalidade contra o qual avaliar a racionalidade do agente, em termos das consequências das suas acções e processos de raciocínio.

Em *Reason and Rationality* (Samuels, Stich e Faucher, 2004) Stich defende que as descrições de racionalidade podem ser divididas em *descrições deontológicas* e *descrições consequencialistas.* As descrições deontológicas avaliam a racionalidade dos agentes, dos seus processos de raciocínio e tomadas de decisão em função dos chamados "cânones de racionalidade", como os princípios da lógica e da teoria da decisão. Quanto às avaliações consequencialistas a mesma avaliação é feita em função das consequências que essas acções ou processos de raciocínio produzem num determinado ambiente.

Neste artigo Stich rejeita a concepção deontológica de racionalidade, implícita àquilo que chama de Posição Standard, de acordo com a qual "ser racional é raciocinar de acordo com princípios derivados de teorias formais" (Samuels, Stich e Faucher, 2004), em favor da concepção consequencialista de racionalidade de acordo com a qual raciocinar correctamente é raciocinar de forma a tornar provável a

realização de certos objectivos ou resultados (Samuels, Stich e Faucher, 2004). Assim, segundo Stich, um processo é racional quando é um meio eficaz para atingir determinados fins, é irracional quando não o é. Um bom raciocínio, para a concepção consequencialista de racionalidade, é aquele que nos põe no caminho da obtenção daquilo que valorizamos (por exemplo, crenças verdadeiras, a realização dos nossos objectivos, etc.). Assim, o fiabilismo de Goldman (Goldman, 1986) é uma forma de consequencialismo de acordo com a qual um bom processo de raciocínio é aquele que conduz à obtenção de crenças verdadeiras. Já para Stich, contudo, aos agentes reais não interessa para nada se as suas crenças são ou não verdadeiras (Stich, 1993) pelo que este autor defende outra concepção de consequencialismo – o pragmatismo – segundo a qual um bom processo de raciocínio é aquele que eficazmente nos ajuda a obter o objectivo pragmático de satisfazermos os nossos fins e desejos pessoais: "*um bom raciocínio é aquele que tende a resultar na obtenção das coisas que valorizamos."* (Samuels, Stich, Faucher, 2004)

Contra a posição standard (ou concepção deontológica de racionalidade) Stich diz que não se sabe concretamente por que razão é que devemos valorizar aqueles raciocínios que estão de acordo com determinados princípios normativos quando estes não nos levam aos fins e objectivos que pretendemos atingir. Um defensor da posição deontológica, segundo Stich, terá de defender que as suas regras de raciocínio são as mais correctas mesmo quando lhe é demonstrado que outras (deontologicamente incorrectas, como algumas heurísticas "rápidas e sujas") têm mais sucesso em fazer com que alcancemos os nossos objectivos.

Uma vez que os agentes racionais humanos estão sujeitos a diversas limitações cognitivas (de tempo, de informação disponível, de processamento dessa informação, de memória, etc.), diferentes agentes cognitivos podem ter diferentes regras e processos para se atingirem os mesmo objectivos. Da mesma forma, diferentes ambientes em que o agente se move (diferentes tipos de informação disponível, tempos diferentes para processar essa informação, etc.) podem proporcionar diferentes resultados (adequados ou inadequados) aos mesmos processos de raciocínio. Desta forma rejeita-se mais uma vez a posição standard, uma vez que os mesmos cânones de Bom Raciocínio não se aplicam a todos os ambientes.

Stich avança ainda outro forte argumento contra a posição standard, nomeadamente aquilo a que chama de "problema da

derivação", ou seja o problema de se saber como é que os princípios normativos daquilo em que consiste um bom raciocínio derivam de sistemas formais como a lógica e a teoria da decisão: "os axiomas e os teoremas do cálculo de probabilidades não implicam logicamente que **devemos** raciocinar de acordo com eles. Eles simplesmente afirmam umas verdades acerca de probabilidades." (Samuels, Stich e Faucher, p.37). Resumindo, para Stich é a racionalidade no sentido consequencialista que verdadeiramente nos interessa.

Stich serve-se então destes argumentos para abandonar a pretensão de se avaliar a racionalidade dos agentes e dos seus processos cognitivos contra um determinado padrão normativo standard; por outras palavras, torna-se bastante difícil saber se os nossos raciocínios são contra-normativos, uma vez que se torna extremamente difícil encontrar uma norma – como disse, mais à frente procurarei uma forma de se encontrar essa norma.

Noutro artigo escrito com o psicólogo Chandra Sripada, "Evolução, cultura e a irracionalidade das emoções" (Sripada e Stich, 2004), Stich serve-se desta concepção consequencialista e pragmatista de racionalidade para justificar a atribuição de racionalidade e irracionalidade às nossas emoções.

Neste artigo os autores avançam a hipótese de as emoções estarem ligadas a um conjunto de normas, objectivos e valores mentalmente representados, a que chamam de "estrutura de valores" (value structure). Segundo Stich e Sripada essas estruturas de valores de um sujeito são os "antecedentes cruciais das emoções". E nesta posição dizem-se acompanhados por influentes autores como Nico Frijda, Klaus Scherer, Keith Oatley e Andrew Ortony, para quem essas normas, valores e objectivos mentalmente representados **desempenham, de facto, um papel fundamental nos processos psicológicos que estão na origem das emoções**. Assim, quando estas estruturas de valores num organismo ou sistema são inadequadas ao meio ambiente em que se inserem conduzem naturalmente a emoções e a comportamentos também eles inadequados, no sentido de não serem propícios à obtenção dos fins e resultados pretendidos. Stich dá o exemplo de medos inatos, fobias várias, sentimentos inadequados de honra e orgulho, reacções negativas a alimentos inofensivos e tabus culturais de comida. Como tal, e segundo o critério consequencialista de racionalidade defendido por Stich, essas estruturas de valores inadequadas **conduzem a emoções e comportamentos irracionais no agente.**

É aqui que, quanto a mim, se levantam dois problemas ao critério consequencialista:

a) o facto de ao seguirmos à risca o critério consequencialista (ou seja, atribuindo racionalidade ou irracionalidade a algo **somente** em função das suas consequências), podermos de facto atribuir racionalidade a estados cognitivos como algumas emoções, mas também podermos alargar essa atribuição a alguns estados não mentais como as protoemoções, que é o que Stich faz quando fala da irracionalidade de certos medos e fobias inadequadas como o sentirmos pânico ao ficarmos fechados num elevador, de aversões corporais como o nojo perante alimentos inofensivos e de certos padrões de reacção emocional como a raiva perante uma afronta inofensiva. Como ficou dito, Stich e Sripada falam-nos de uma estrutura mental de normas e valores que conduzem a emoções racionais quando essas normas e valores são adequados ao nosso ambiente actual e a emoções irracionais quando não o são. **No entanto, e aqui está a minha principal crítica a Stich, o artigo não deixa bem claro se entre o sentimento de uma emoção (racional ou irracional) e a estrutura de valor que o despoleta existe uma avaliação cognitiva da estrutura de valores por parte do agente.** Parece-me que Stich não afirma isso, e que no artigo fica implícito o contrário (como quando Stich fala de uma *reacção irracional de nojo*), ou seja que a passagem das normas, objectivos e valores da estrutura de valores para a reacção emocional consequente é feita de forma directa ou não cognitiva, i.e., biologicamente causada, num processo que o neurocientista Joseph LeDoux define por Low-Road Pathway (Le Doux and Phelps) - daqui para a frente, sempre que me referir a processos não cognitivos estarei a fazer referência a estímulos que são processados ao nível da amígdala e do hipocampo não passando pelo neocortex, dando assim lugar a respostas mais rápidas e imediatas a esses estímulos.

O exemplo que Stich avança do nojo parece-me ser uma reacção demasiado rápida e imediata para ser considerada cognitiva, e como tal não a incluiria na categoria das emoções, mas sim na das protoemoções.

Sendo assim ficamos sem saber onde parar a atribuição de racionalidade. Se o critério é apenas o das consequências das acções a que as estruturas de valor conduzem, por que não falarmos da racionalidade e irracionalidade daquilo que causa essas mesmas estruturas de valor, ou seja, a própria estrutura biológica do agente cognitivo de que as nossas estruturas de valor são a consequência?

Stich fala dos genes (além do ambiente e a cultura) como uma das fontes da nossa estrutura de valores.

b) a segunda objecção que coloco ao critério consequencialista é a da necessidade que este critério tem de um meta-critério normativo que lhe indique quais as boas e as más consequências, ou seja, a necessidade de um conceito de normalidade contra o qual avaliar a racionalidade de uma emoção. Na segunda secção procurarei abordar estes dois problemas.

Secção II - Racionalidade derivada e avaliação de racionalidade

Nesta secção, e num primeiro momento procuro encontrar uma resposta ao primeiro problema levantado ao critério consequencialista defendendo a atribuição de racionalidade derivada (no sentido cognitivo) ao que atrás defini como protoemoções. Para este primeiro ponto inspirei-me na obra "Intencionalidade" de John Searle, mais concretamente nos seus conceitos de *rede de intencionalidade, background pré-intencional e intencionalidade derivada*. Num segundo momento procuro enfrentar o segundo problema do critério consequencialista sondando muito superficialmente aquilo que considero serem as condições normativas para a avaliação de racionalidade dos processos cognitivos dos agentes e arriscando um modelo ou método para que se encontre o tal meta-critério normativo de racionalidade.

O trabalho de John Searle em filosofia da mente evoluiu a partir do seu trabalho em filosofia da linguagem e o tratamento que dá ao problema da intencionalidade é ilustrativo disso mesmo, no sentido em que na sua abordagem a esse problema se inspirou na forma como compreendemos o significado das frases de uma linguagem. Essa compreensão, diz-nos Searle, "requer um *Background* (...) de suposições pré-intencionais que não fazem parte do significado literal da frase." (Searle, p.188). Isso é demonstrado quando, ao mudar-se o dito *Background* (que pode ser entendido como o contexto em que a frase se insere), o sentido da frase (a mesma frase) se altera.

Searle fala assim da forma como facilmente atribuímos intencionalidade a uma série de fenómenos não mentais como palavras, figuras e símbolos que adquirem significado apenas em função do *background* em que se inserem. Nestes casos, diz-nos Searle, a intencionalidade da mente é invocada para dar conta da intencionalidade destes fenómenos não mentais. Searle atribui a estes

fenómenos não mentais **intencionalidade derivada**.

A intencionalidade é para Searle uma propriedade intrínseca da mente, por meio da qual esta é **acerca de**, **se refere a**, ou **se dirige a** algo no mundo exterior. O que Searle nos diz é que, analogamente ao que se passa com o reconhecimento do significado literal de uma frase, um estado mental intencional requer uma *rede de intencionalidade* em cuja base, ou *background*, está uma série de "capacidades, posturas, suposições e pressuposições pré-intencionais, práticas e hábitos" (Searle, p.199) que são o fundamento dessa *rede de intencionalidade*. Ou seja, um *background* pré-intencional que nos permite formar uma intenção superior. É essa intenção superior que, segundo Searle, torna intencionais todos os actos intermédios que conduzem à prossecução dessa intenção maior. Isto é, podemos atribuir intencionalidade a esses actos intermédios mas apenas em função da intencionalidade de uma acção superior. Assim, quando formo a intenção de me levantar e sair porta fora, faço-o devido a um *background* não intencional de tendências, posturas e capacidades pré-intencionais. Por exemplo, tenho tendência a sentir-me mal em locais fechados, tenho vontade de sair desta sala, sou capaz de me movimentar andando. Agora, a nenhum destes fenómenos (o meu mal estar, a minha vontade em procurar uma situação em que me sinta mais confortável, a minha capacidade de andar) Searle atribui **intencionalidade**, no entanto todos eles são necessários para que eu forme a minha intenção de sair por aquela porta.

Esta posição de Searle é obviamente contestável e julgo que pelo menos em relação aos dois primeiros exemplos (mal estar e vontade de procurar uma situação mais agradável) um autor como Tim Crane diria serem exemplos de estados intencionais, uma vez serem estados mentais, mesmo que inconscientes, dirigidos a estados do corpo. Na minha opinião, um estado para ser considerado intrinsecamente intencional deverá ser um estado mental consciente.

Para Searle, um estado intencional exige ainda algo mais que um *background* pré-intencional. Para que algo (um estado mental, uma acção) seja considerado intencional tem de estar inserido numa *rede de intencionalidade*, isto é, tem de fazer parte de um conjunto de atitudes proposicionais como crenças, desejos e intenções várias que lhe confiram intencionalidade. Assim, ao formar uma intenção de sair por aquela porta, só posso ter essa intenção dentro de uma rede de outras crenças, desejos e intenções. Devo acreditar que atrás daquela porta está um buraco pelo qual posso sair desta sala, devo desejar levantar-

me desta cadeira e dirigir-me para a porta, devo ter a intenção de atravessar a pé todo este espaço entre a cadeira e a porta, etc. Ou seja, um estado intencional como pretender sair desta sala só é um estado intencional quando ligado a um número imenso de outros estados intencionais, ou seja, quando inserido numa *rede de intencionalidade*, caso contrário seria um mero evento físico. Ou seja, a meu ver, e resumindo grosseiramente o que atrás ficou dito, podemos compreender melhor o que entendo por um evento intencional se o virmos como um **evento físico com significado para o agente**.

O ponto que aqui pretendo focar é que para que se possa falar tanto de intencionalidade como de racionalidade de um agente, de uma acção ou de um processo de raciocínio (tanto de racionalidade no sentido cognitivo como no sentido normativo), são necessários dois níveis de condições a que chamo: **condições de ocorrência** e **condições normativas de avaliação**.

Quando falamos de algo como a racionalidade de um agente temos de falar, num primeiro nível, das **condições de ocorrência** de racionalidade. Como vimos atrás, segundo Searle, para que uma acção, um agente, ou um processo de raciocínio sejam intencionais precisam de um "conjunto de condições capacitantes que tornam possível o funcionamento de formas particulares de intencionalidade" (Searle, p.202) - o *Background* pré-intencional. **Da mesma forma, para que uma acção, um agente, ou um processo de raciocínio possam tornar-se racionais também necessitam de todo um conjunto de sentimentos, vontades e atitudes avaliadoras do corpo, ou seja, de estados puramente físicos, não cognitivos mas com alguma forma de intencionalidade, motivação, volição e com um determinado "ponto de vista" sobre o mundo, ou seja, com alguma forma de racionalidade**. A esse conjunto não cognitivo de sentimentos e vontades do corpo dei o nome de **protoemoções** (Secção I). Na linha de António Damásio (Damásio, 1994) e Keith Oatley (1992) considero que essas protoemoções (como os reflexos, impulsos e motivações, o sistema de regulação metabólica, as emoções de fundo como o prazer e a dor, a nossa "caixa de ferramentas heurísticas" (Zilhão, 2006), as nossas diferentes formas de pensamento indutivo, etc.) constituem o *background*, ou a estrutura biológica da cognição, sendo responsáveis pela diminuição do número de considerações relevantes para uma deliberação racional dos custos e dos benefícios de uma acção. Inclino-me neste ponto para a interessante proposta de Gianmateo Mameli (Mameli, 2004) para quem uma tomada de decisão nunca é puramente

racional e são esses sentimentos inconscientes que determinam do princípio ao fim o processo de tomada de decisão. Isto é, devido à precedência evolutiva do nosso sistema emotivo sobre o nosso sistema deliberativo, Mameli acredita que são as nossas emoções (positivas ou negativas) que nos indicam que linha de acção devemos seguir (emoção positiva) e qual devemos evitar (emoção negativa).

Assim, e este é um dos pontos centrais do artigo, defendo que esses eventos físicos e não cognitivos, por constituírem o *background* pré-racional da racionalidade, **são condições necessárias para a ocorrência de estados cognitivos como crenças, desejos, intenções e devem por isso ser consideradas cognitivamente racionais, mesmo que num sentido fraco (derivado) de racionalidade.** Da mesma forma que para Searle a intencionalidade de certos fenómenos não mentais é derivada de uma intencionalidade superior - a intencionalidade da mente - a racionalidade cognitiva das protoemoções é também ela derivada de uma racionalidade superior, a racionalidade de um agente inserido numa rede de atitudes proposicionais. Como vimos, para compreendermos o significado das frases, figuras e símbolos atribuímos-lhes uma forma artificial de intencionalidade. Da mesma forma acredito que, para realmente compreendermos o modo como raciocinamos e como o nosso *background* pré-racional é determinante para aquilo que entendemos como a nossa racionalidade (no sentido cognitivo), esta atribuição de **racionalidade derivada** às nossas protoemoções, apesar de artificial, é uma estratégia metodologicamente útil uma vez que nos ajuda a compreender e aceitar o contributo e a importância das ferramentas não cognitivas nos nossos processos de raciocínio.

Quanto ao segundo nível de condições que considero essenciais para se falar de racionalidade (agora no sentido normativo) temos aquilo que apelidei de **condições normativas de avaliação**. Da mesma forma que a atribuição de intencionalidade exige uma inclusão de todo um *background* de atitudes pré-intencionais do agente numa *rede de intencionalidade* (crenças, desejos e intenções), também a atribuição de racionalidade exige a inclusão de todo um conjunto de processos (cognitivos e não cognitivos) do agente numa **rede de normas de conduta** (o que deve e não deve ser valorizado) e numa **rede de atitudes proposicionais** (crenças, desejos e intenções). Ou seja, um agente, uma acção ou um processo de raciocínio só são avaliados quanto à sua racionalidade normativa quando inseridos num contexto linguístico e num contexto social complexo, com todas as normas,

regras e valores a eles associados, pelo que não faz sentido falar de racionalidade de forma abstracta, da mesma forma que não faz sentido falar de justiça sem falarmos em normas, regras e valores que a concretizem. Assim, para aferir a racionalidade de um agente, uma acção ou um processo cognitivo é necessária uma **norma**, e essa norma é-nos dada pela rede de crenças e atitudes proposicionais em que o agente se insere, ou seja, pelas condições normativas de avaliação. **O que eu aqui defendo, e este é o segundo ponto central deste artigo, é que é preciso algo mais do que apenas as condições normativas de avaliação para aferir a racionalidade ou irracionalidade de fenómenos cognitivos (agentes, acções, processos), e isto tanto no sentido cognitivo como no sentido normativo de racionalidade.**

Ou seja, por um lado, a racionalidade de um agente ou sistema (i.e., as componentes proposicionais do seu sistema cognitivo – a sua rede de crenças, desejos, intenções e emoções superiores) é condicionada *a priori* por um conjunto de processos não racionais, ou *background* pré-racional – **condições de ocorrência -** , por outro lado, essa mesma racionalidade só é avaliada *à posteriori* – **condições normativas de avaliação** – contra um conjunto de normas e atitudes proposicionais que rodeiam o agente. O que acontece é que estas avaliações normativas de racionalidade não costumam levar em consideração o papel daqueles processos não cognitivos que constituem as condições de ocorrência, o que leva os avaliadores a afirmar que a forma como as pessoas raciocinam e tomam decisões viola sistematicamente os cânones familiares de racionalidade (para uma amostragem de algumas investigações em psicologia cognitiva que parecem apoiar estas "interpretações pessimistas" sobre a racionalidade humana ver "*Repensando a Racionalidade: de Implicações Pessimistas a Módulos Darwinianos"* de Samuels, Stich e Tremoulet, 2003).

A minha opinião é que qualquer avaliação da racionalidade deve ser uma **interpretação normativa dos resultados das acções de um conjunto de condições de ocorrência**. Ou seja, uma avaliação da racionalidade que leve em linha de conta a contribuição do *background* pré-racional (as **condições de ocorrência**) para a racionalidade (e irracionalidade) do agente. Assim, ao defendermos a atribuição de **racionalidade derivada** às protoemoções estamos ao mesmo tempo a reconhecer a importância desse *background* pré-racional enquanto possibilidade mesma de qualquer acontecimento racional e a defender que uma futura teoria filosófica da racionalidade se deverá tornar mais biológica, no sentido em que deverá procurar incorporar tanto os

aspectos psicológicos e pessoais (crenças, desejos, intenções e emoções), como os aspectos fisiológicos/cerebrais e sub-pessoais dos nossos processos cognitivos (protoemoções, heurísticas, módulos mentais, intuições, etc.) avaliando-os contra uma rede de normas de conduta socio-culturais que necessariamente rodeia os agentes reais. Por outras palavras, a intuição que procurei aprofundar neste artigo é que **não podemos falar de uma avaliação normativa de racionalidade sem compreendermos a racionalidade cognitiva dos agentes reais.**

Seria essa *futura teoria filosófica da racionalidade* que nos providenciaria o tal meta-critério normativo que falei em cima na *segunda objecção ao critério consequencialista*. Na verdade não tenho bem claro como poderemos alcançar esse critério – e uma tentativa de o desenvolver faz parte de uma possível futura investigação caindo, por isso, fora do escopo imediato deste artigo -, no entanto, procurando levantar um pouco o véu, julgo que essa norma de racionalidade deverá ter em linha de conta os seguintes pontos:

1. a mecânica cognitiva do agente, i.e., tanto os componentes proposicionais do seu sistema cognitivo (i.e., o conjunto de normas, crenças, desejos, intenções e emoções superiores do agente), como também o seu *background* pré-racional e protoemocional (sentimentos avaliadores, pontos de vista emocionais, "caixa de ferramentas heurísticas", formas de pensamento indutivo, etc).
2. os cenários em que o agente se insere e as suas acções se desenrolam, ou seja, o seu meio ambiente, sociedade e cultura – aquilo que Stich chama o *ambiente actual* do agente (Sripada e Stich, 2004).
3. as consequências das acções do agente.

Defendo que **1**) e **2**) dar-nos-ão pistas para a formulação desse meta-critério normativo contra o qual avaliar **3**).

Ou seja, como se pode ver, não rejeito o critério consequencialista de Stich para a avaliação da racionalidade do agente, mas julgo que o mesmo deve ser complementado por um meta-critério normativo que tenha em conta não somente o meio ambiente em que os agentes se inserem mas também a sua estrutura bio-cognitiva. Seria *contra* esse meta-critério normativo que se avaliariam as consequências das acções dos agentes.

Resumindo um pouco o que ficou para trás, são as nossas capacidades, tendências e habilidades biológicas (o meu *background* pré-racional) que constituem a base do meu aparato cognitivo e é essa base biológica que me permite lidar com o mundo da forma que me é mais apropriada (módulos mentais, heurísticas rápidas e sujas, formas de raciocínio indutivo etc.). São essas capacidades avaliadoras, sub-pessoais e pré-racionais que me permitem sentir o mundo e agir em conformidade com ele, e é este agir pragmático e natural do meu organismo que confunde aqueles que acham que o agir racional deve de alguma forma estar de acordo com determinadas leis formais e artificiais (os chamados cânones de racionalidade como a lógica, a teoria da decisão, a teoria das probabilidades) que não dão conta da forma como os agentes reais realmente raciocinam e agem.

Conclusões

a) Julgo que muita da resistência que temos em atribuir racionalidade a algo como um sentimento emocional tem origem na vetusta dicotomia que separa razão de emoção, colocando a primeira no lado da **mente** e a segunda no lado do **corpo**. Como tal, aquilo que dissermos acerca da racionalidade cognitiva das emoções e das protoemoções reflectirá certamente aquilo que pensamos acerca do problema **mente-corpo**. Nesse sentido acredito que nada nos diz, ou antes, só a nossa intuição comum nos diz, que um pensamento tem de ser um estado mental abstracto e separado do corpo, e a minha intuição é de que essa intuição comum está errada. Nesse sentido defendo que os seres humanos de alguma forma sentem aquilo que pensam e como tal um pensamento deverá ser entendido como uma espécie de sentimento avaliador corporizado. Tal não implica que se possa falar da racionalidade intrínseca desse sentimento avaliador, pois para isso teríamos de entrar no campo das *condições normativas de avaliação*, ou seja, das atitudes proposicionais e da rede de crenças e normas do agente. E essas condições de avaliação de racionalidade no sentido normativo só se aplicam intrinsecamente aos agentes como um todo e não a cada um dos seus processos de raciocínio em particular, que apenas podem ser avaliados quanto à sua racionalidade derivada. Ou seja, **a racionalidade dos processos de raciocínio dos agentes é <u>derivada</u> de uma avaliação normativa da racionalidade dos agentes**. Como tal, a atribuição de **racionalidade derivada** (no sentido cognitivo) às emoções e atitudes proposicionais assim como às

protoemoções, reflecte a opacidade do termo emoção, ou seja, a difusão alargada do fenómeno emocional pelo nosso sistema cognitivo (de estados neurológicos não cognitivos a atitudes proposicionais cognitivas) e pretende encurtar um pouco a distância entre os dois extremos desse sistema, o biológico e o racional, o corpo e a mente. Ao encurtar essa distância estou a tentar eliminar do vocabulário comum (e filosófico) uma vetusta metáfora que identifica o corpo com uma máquina estúpida e cega, sendo o espírito ou a mente aquilo que daria alguma sabedoria e inteligência ao homem. Uma máquina é, de facto, estúpida e cega, no entanto o corpo, mesmo sem ser animado por algum espírito vital, uma *anima*, é algo orgânico que evoluiu, no âmbito da espécie, por selecção natural ao longo de milhões de anos e, no âmbito do indivíduo, no decurso da sua formação socio-cultural ao longo de algumas dezenas de anos. O nosso organismo possui, como tal, informação e conteúdo. E é exactamente aqui que o corpo se aproxima da mente pois o que distingue os processos orgânico-mentais de outros processos físico-maquínicos (como a digestão e o funcionamento de um termostato) é o facto de aqueles possuírem conteúdo e informação (i.e., alguma forma de semântica) e estes não.

b) Quanto ao que ficou dito acerca do meta-critério normativo de racionalidade julgo que uma crítica que se pode fazer a esse critério é que, devido à sua necessária confirmação empírica em agentes reais, dificilmente será um critério perene e estável, mas antes volúvel, permeável a infirmação científica, tendencialmente individualizante e dependente do contexto, o que torna bastante trabalhosa a avaliação da racionalidade dos agentes e dos seus processos cognitivos e, além disso, torna problemático o próprio termo *critério normativo* que temos vindo a utilizar.

Uma resposta possível seria que a questão da avaliação da racionalidade é uma questão que teremos de deixar à nossa própria mecânica cognitiva sub-pessoal para responder. Talvez tenhamos um módulo mental para detectar instintivamente racionalidade e irracionalidade em agentes e acções! Ou tenhamos talvez uma heurística qualquer, ou um tipo de pensamento indutivo mais fiável que qualquer outro juízo ou pensamento dedutivo mais "racional"? Na verdade, tendo em conta a forma como parece que realmente raciocinamos e tomamos decisões, o que teria isto de estranho?

No entanto inclino-me para uma segunda resposta possível. Julgo que a semelhança biológica entre os sistemas com que uma avaliação normativa de racionalidade normalmente se preocupa - os seres

humanos - assim como a forma semelhante como esses sistemas evoluíram e os ambientes relativamente estáveis em que estes actualmente se inserem, deve servir-nos de base segura para algumas generalizações normativas quanto à sua racionalidade e irracionalidade. Assim, um agente com um *background* totalmente díspar do de outros seres humanos e que, como tal, conduza a acções e reacções completamente diferentes dos outros seres humanos e, além disso, a acções e reacções completamente inadequadas ao seu meio ambiente actual, esse agente, dizia, será um agente irracional e as acções e reacções a que o seu *background* conduz serão também elas, em alguns agentes mais do que outros, frequentemente irracionais.

Resumindo, e procurando articular **a)** e **b)**, quando falamos da racionalidade dos processos de raciocínio de um agente (cognitivos ou não cognitivos) estamos a falar de racionalidade num sentido cognitivo, e aqui apenas podemos falar da racionalidade derivada desses processos, uma vez que aferimos a sua racionalidade apenas em função de uma avaliação normativa da racionalidade do agente. Quando falamos da racionalidade do agente falamos de racionalidade num sentido normativo e aqui já podemos falar da racionalidade (ou irracionalidade) intrínseca do agente e para isso temos que ter em conta os seus processos de raciocínio (cognitivos e não cognitivos).

Referências

Damásio, António R., 1994, *O Erro de Descartes. Emoção, Razão e Cérebro Humano*, Mem-Martins, Publicações Europa-América, Trad. Dora Vicente e Georgina Segurado.

Damásio, António R., "William James and the modern neurobiology of emotion", in Evans, Dylan and Cruse, Pierre (eds.), *Emotion, Evolution and Rationality*, Oxford, Oxford University Press, 2004.

Damásio, António R., The Feeling of What Happens. Body and Emotion in the Making of Consciousness, New York, Harscourt Brace and Company, 1999.

Evans, Dylan and Cruse, Pierre (ed), 2004, *Emotion, Evolution and Rationality*, Oxford, Oxford University Press.

Goldman, Alvin, *Epistemology and Cognition,* Cambridge, Mass., Harvard University Press, 1986.

LeDoux, Joseph, E. and Phelps, Elizabeth A., "Emotional networks in the brain", in Lewis, Michael and Haviland-Jones,Jeannette M., *Handbook of emotions,* New York, The Guilford Press.

Lewis, Michael and Haviland-Jones, Jeannette M., *Handbook of emotions,* New York, The Guilford Press.

Mamelli, G., "The role of emotions in ecological and practical rationality", in Evans, Dylan and Cruse, Pierre (eds.), *Emotion, Evolution and Rationality*, Oxford, Oxford University Press, 2004.

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

Oatley, K., *Best laid schemes: The psychology of emotions*, Cambridge, Cambridge University Press, 1992.

Samuels, Richard, Stich, Stephen and Tremoulet, Patrice, D., "Repensando a Racionalidade: de Implicações Pessimistas a Módulos Darwinianos", Intelectu nº 9 www.intelectu.com (Rethinking Rationality: from bleack implications to darwinian modules, in Lepore, E. and Pylyshyn, Z., eds., *Invitation to cognitive science,* 2003, tradução portuguesa de Tomás Magalhães Carneiro).

Samuels, Richard, Stich, Stephen and Faucher, Luc, "Reason and Rationality", in *Handbook of epistemology* ed. I. Niiniluoto, M. Sintonen and J. Wolenski, Dordrecht, Kluwer, 2004. Pp. 1-50.

Searle, John R., *Intentionality : an essay in the philosophy of mind*, Cambridge, Cambridge University Press, rep. 1997.

Searle, Jon. Rationality in Action. Cambridge: MIT Press, 2001.

Sripada, Chandra S. & Stich, S., "Evolução, cultura e a irracionalidade das emoções", Intelectu nº 11 www.intelectu.com (Evolution, Culture and the Irrationality of the emotions, in Evans, Dylan and Cruse, Pierre, eds., *Emotion, Evolution and Rationality*, Oxford, Oxford University Press, 2004, tradução portuguesa de Tomás Magalhães Carneiro).

Stich, Steven, *The Fragmentation of Reason – Preface to a pragmatic Theory of Cognitive Evolution,* Cambridge, MA, MIT Press, 1990.

Zilhão, António, "Heurísticas Rápidas e frugais, encontro de probabilidades e incontinência", 2006, (no prelo).

A teoria da acção de Donald Davidson e o problema da causação mental

Susana Cadilha[1]

Resumo: O presente trabalho pretende ser uma análise crítica de alguns aspectos da filosofia de Donald Davidson, nomeadamente da sua teoria da acção e da proposta ontológica com ela intimamente relacionada. Ainda que, formando a obra de Davidson uma visão de conjunto integrada e coerente que abrange praticamente todos os problemas filosóficos, tais considerações possam ter implicações noutros domínios que não os explicitamente tratados, o nosso principal problema será o da causação mental – se é possível defendê-la no interior do esquema davidsoniano.

Abstract: This paper is a critical survey of some issues of Donald Davidson's philosophy. I focus primarily on Donald Davidson's theory of action and his ontological proposal. Although these considerations may have implications in domains other than the ones directly related to them, my main here problem will be that of mental causation – whether it is possible to coherently sustain it within the framework of Davidson's philosophy, or not.

[1] Membro e investigadora do *Mind Language and Action Group* – MLAG – do Instituto de Filosofia da Universidade do Porto. Bolseira de doutoramento da FCT.

A teoria da acção de Davidson

Começarei por expor a teoria da acção davidsoniana, baseando-me, para tal, sobretudo no artigo de referência *Actions, Reasons and Causes*, onde Davidson introduz a terminologia de que pretendemos fazer uso.

No artigo em causa, o autor trata da questão da explicação da acção e tem como objectivo esclarecer qual a relação que é possível estabelecer entre razões e acções. De acordo com Davidson, uma razão explica uma acção apenas se constituir a razão pela qual o agente levou a cabo essa acção. Uma tal explicação ele designa por *racionalização*. Isto significa que Davidson admite a possibilidade de alguém ter uma razão para realizar uma acção, levar a cabo a acção, e essa não ter sido a razão pela qual ele fez o que fez[2]. Nesse caso, a razão não explica a acção.

Davidson visa mostrar que quando essa relação entre razões e acções se verifica – isto é, quando a razão de facto explica a acção – essa explicação é uma forma de explicação causal. Isto quer dizer que as razões explicam as acções na medida em que são as suas causas, ou melhor, é precisamente porque a razão é a causa da acção que esta pode ser explicada explicitando aquela.

Por aqui se vê como Davidson atribui um papel positivo à causação mental – se as razões são causas é porque ele acredita que a mente intervém no mundo. De que forma isso é possível e se uma tal tese está em conformidade com os pressupostos de Davidson é o que pretendemos averiguar.

Prosseguindo com a apresentação da terminologia proposta no artigo em questão, Davidson descreve com mais pormenor a situação na qual uma razão racionaliza a acção. Isso acontece quando ao explicitar a razão estamos a dar conta daquele aspecto da nossa acção que

[2] Para ilustrar um tal caso, Davidson apresenta o exemplo de um montanhista que segue à frente de um outro, sentindo-se incomodado pelo peso e pelo cansaço que essa situação lhe provoca, e pelo perigo que ela representa. Ele pode querer livrar-se do peso e pode saber que largando a corda conseguiria isso. Pensar em tal coisa pode enervá-lo a tal ponto que acaba por soltar a corda e o segundo montanhista cai. O que Davidson pretende mostrar é que existiriam crenças e desejos que racionalizariam essa acção, mas só no caso de a terem causado de forma apropriada, ou seja, no caso de ter sido essa realmente a razão pela qual a acção é feita. Neste exemplo, tal poderia não ter sucedido, e o ocorrido pode nem ser considerado uma acção, se não foi intencional (cf. Davidson, Donald, "Freedom to act", in *Essays on Actions and Events*, Oxford, Oxford University Press, 1980).

valorizamos, consideramos desejável ou obrigatório. Precisamente porque para uma dada acção podem, em princípio, ser encontradas muitas explicações possíveis, é necessário saber qual é a que é de facto o caso. É neste sentido que Davidson afirma que racionalizar a acção é apresentar as *atitudes pro*[3] (talvez possamos traduzir por atitudes positivas) que um agente mostra relativamente a ela (desejar, valorizar, considerar imperativo algum dos seus aspectos, etc), assim como a crença que o agente tem de que essa acção vai cumprir aquela que é a sua intenção ao realizá-la.

Racionalizar a acção é, portanto, descrevê-la de um ponto de vista mentalista, explicando-a através da atribuição de crenças e desejos a um agente. Indicar a atitude pro e a crença relativa a essa atitude é apresentar a *razão primária* pela qual o agente realizou a acção. E é esse processo, a que o agente chega por introspecção, que Davidson designa por *racionalização*.

De notar que o esquema apresentado por Davidson permite dar resposta aos dois principais problemas com que uma teoria da acção se debate – não apenas ao problema da explicação da acção como também ao problema metafísico de saber em que consiste uma acção (o que distingue uma acção de um mero acontecimento físico ou de uma qualquer ocorrência que simplesmente nos acontece?). Dadas as teses anteriormente defendidas, o que distingue uma acção é o facto de poder ser descrita de um ponto de vista mentalista ou intencionalista, recorrendo a razões que envolvem crenças e desejos (estados intencionais de um agente). É o facto de ser causada por razões que distingue uma acção de outros eventos no mundo. Uma acção consistirá, então, numa descrição intencional de um evento.

O nosso principal alvo de crítica, nesta fase, será o conceito de racionalização e suas virtualidades, pelo que será conveniente explanarmos o melhor possível o que entendia Davidson por tal noção.
É de salientar, antes de mais, que racionalizar uma acção *não é* mostrar que essa acção é racional segundo a teoria canónica da racionalidade na acção – a teoria instrumental. Uma teoria da acção ocupa-se com as questões da natureza da acção e da sua explicação; pode ser de carácter psicologista, como o modelo crença-desejo, definindo os estados intencionais como crenças e desejos como a marca da acção e a via para a sua explicação (constituirá aquilo que alguns designam como uma

[3] Neste conceito incluem-se coisas tão diversas como desejos, objectivos, valores, convenções sociais, princípios estéticos, etc.

perspectiva *inward-looking*[4] em teoria da acção), ou pode ser de carácter não-psicologista e considerar que aquilo que define a acção é a sua adaptação ao mundo exterior e não é o facto de acreditarmos em alguma coisa o que nos incita a agir[5] (essa abordagem em teoria da acção será então *outward-looking*). Já uma teoria da racionalidade na acção tem como tarefa essencial definir o que faz com que uma acção seja considerada racional. Neste âmbito, a teoria tida como mais consensual é a que apresenta uma definição instrumental de racionalidade – ser racional será mobilizar os meios mais adequados com vista a um determinado fim. Uma teoria da acção não é necessariamente uma teoria da racionalidade na acção, apesar de ambas se encontrarem ligadas.[6]

Neste caso específico, a teoria da acção de Davidson pretende tratar a questão da explicação da acção e não a questão de saber quais os critérios que tornam uma acção racional.

Por exemplo, de acordo com a teoria instrumental, se eu quero matar alguém, realizarei uma acção racional se mobilizar os meios adequados para atingir esse fim; se, por exemplo, disparar contra essa pessoa. Mas a minha acção será irracional se, em vez disso, fizer algo como oferecer um animal inofensivo a essa pessoa, digamos, um rato. No entanto, se de acordo com o modelo canónico da racionalidade na acção, esta pode ser considerada uma acção irracional, de acordo com a teoria da acção de Davidson, ela pode ainda ser racionalizada. A razão primária que, nesse caso, explicaria a minha acção seria constituída pelo meu desejo de matar essa pessoa e pela minha crença de que oferecendo um rato a essa pessoa, ela morrerá, na medida em que disponho de uma outra crença relacionada com essa – a de que essa pessoa tem um medo terrível de ratos e sofrerá um ataque cardíaco se estiver em contacto directo com um.

Racionalizar uma acção também *não* pretende ser uma sua justificação, como alguns parecem entender. Com o intuito de criticar o modelo crença-desejo, que tem em Davidson um dos seus mais célebres defensores, Stout apresenta o exemplo de uma vigilante de um exame que deveria dar por terminado o mesmo passadas três horas, só que, passadas apenas duas horas, ela forma por engano a crença de que já havia acabado o tempo regulamentar, e pára o exame. A pergunta é: "tem ela justificação para ter feito o que fez pelo simples facto de

[4] Cf. Stout, Rowland, *Action*, Acumen, 2005.
[5] Cf. Dancy, Jonathan, *Practical Reality*, Oxford: Oxford University Press, 2000.
[6] Cf. Madeira, Pedro, "O que é o modelo crença-desejo?", in *Intelectu*, nº9, 2003.

acreditar que o tempo já tinha passado?"[7]. O autor defende que *acreditar* simplesmente não pode contar como justificação para agir, porque se assim fosse ela não teria agido mal mas apenas "acreditado mal", isto é, sustentado uma crença errada. E não poderia nesse caso ser responsabilizada pela sua acção, na medida em que ela acreditava estar a fazer o que era suposto.

Ora, parece-me que uma tal crítica está mal dirigida, porque o que Davidson afirma é que é possível apresentar as razões que explicam porque é que ela fez isso, o que não significa que a sua acção seja justificável. A racionalização não é, pois, uma questão de justificação mas de tornar inteligível um acto. O objectivo desta crítica apontada a Davidson e ao modelo crença-desejo é mostrar que não são as nossas crenças e desejos que explicam o nosso comportamento; para saber o que fazer temos antes de olhar para fora, para o mundo exterior, e comportarmo-nos em conformidade (nas palavras do autor, "a vigilante devia consultar o relógio, não o seu estado mental"[8]). No entanto, é sempre necessário que acreditemos, por exemplo, que as horas do relógio estão certas para que consigamos agir.

O alcance do conceito de racionalização
A noção de akrasia

Aclarado o conceito de racionalização na acepção davidsoniana, podemos perguntarmo-nos quais as suas implicações. Mais concretamente, trataremos de apontar aquilo que Davidson acredita ser possível conseguir através de um tal processo, e as insuficiências que ele nos parece comportar.

Por meio da racionalização, será possível chegar às razões das nossas acções, e avaliando as nossas acções à luz das crenças e desejos que formam essas razões será possível caracterizar uma tal relação como racional ou não. Que ligação pode, então, ser estabelecida entre o processo de racionalização e o conceito de racionalidade? Se racionalizar não é mostrar que uma acção é racional, em que sentido fala Davidson de racionalidade, neste contexto particular?

Racionalizar uma acção é torná-la inteligível, apresentando os desejos e crenças que são a sua causa, não é mostrar que essa acção é racional. No entanto, uma vez encontradas as razões que tornam essa

[7] Stout, Rowland, *op. cit.*, pp. 37-39.
[8] Ibidem.

acção inteligível (que explicam porque é que ela ocorreu), é possível imputar racionalidade ou irracionalidade a um agente se se verificar que a sua acção está, ou não está, em concordância com o seu particular conjunto de crenças e desejos.

A relação que se estabelece entre razões e acções é, por isso, uma relação racional se a minha acção estiver em conformidade com o par crença-desejo que forma a razão primária da minha acção. O ponto a reter é que os nossos desejos e crenças não precisam de ser racionais para servirem de razões para a minha acção, mas a acção deve ser racional, dados os meus particulares desejos e crenças.[9] Racional no sentido de estar em coerência com, e não no sentido de ser o melhor meio para obter um fim (nesse caso nada teríamos adiantado à teoria instrumental). Davidson di-lo numa passagem de um outro célebre artigo, *Psychology as Philosophy* – "duas ideias são construídas a partir do conceito de agir por uma razão: a ideia de causa e a ideia de racionalidade. (…) Uma forma através da qual a ideia de racionalidade é construída é óbvia: a causa [razão] deve ser [composta por] uma crença e um desejo à luz dos quais a acção é razoável."[10] Uma acção será racional, portanto, se for desejável dados os meus particulares desejos e crenças. Ou seja, considerando aqueles que são os meus desejos e crenças, a minha acção revela-se, ou não, coerente com eles.

É por esta via que é possível imputar racionalidade ou irracionalidade a um agente. Irracional seria aquele que numa dada situação não agisse em conformidade com os seus desejos e crenças. Trata-se de uma inconsistência interna relativa às crenças e desejos do agente, e não uma inconsistência relativamente a padrões externos, pois nesse caso não seria possível decidir quais os critérios que serviriam de standard para aferir da irracionalidade da acção, e não seria possível falar de "irracionalidade objectiva"[11], como pretende Davidson neste âmbito.

Um exemplo de uma acção irracional seria o caso da akrasia – um caso em que o agente age contrariamente ao seu melhor juízo. Perante uma determinada situação, surgem-lhe várias alternativas de acção, e ele, considerando todos os aspectos em jogo, decide por uma delas. No

---

[9] Cf. Davidson, Donald, "Intending", in *Essays on Actions and Events*, Oxford, Oxford University Press, 1980.
[10] Davidson, Donald, "Psychology as Philosophy", in *Essays on Actions and Events*, Oxford, Oxford University Press, 1980, p. 233.
[11] Cf. Davidson, Donald, "Incoherence and Irrationality", in *Problems of Rationality*, Oxford, Oxford University Press, 2004.

entanto, acaba por agir em sentido contrário. Nesta situação, o que se passa, nas palavras de Davidson, é que o agente viola o princípio da continência, segundo o qual o agente deve agir de acordo com o seu melhor juízo, e não age em conformidade com os seus desejos e crenças mais relevantes.

Em *How is the weakness of the will possible*, o autor apresenta esse mesmo princípio de forma mais detalhada. Ele é formado pelo princípio (P1) de acordo com o qual se um agente é livre de fazer x ou y, e quer fazer x mais do que quer fazer y, então ele intencionalmente fará x, e o princípio (P2) de acordo com o qual se um agente julga que seria melhor fazer x do que fazer y, então ele quer fazer x mais do que quer fazer y.[12] Na perspectiva de Davidson, estes dois princípios são absolutamente evidentes, e atestam a irracionalidade das acções akráticas. Pois, se um agente é livre de fazer x ou y, e quer fazer x mais do que quer fazer y, então está a agir irracionalmente se não fizer x. Há irracionalidade porque o agente não atribui às considerações relevantes que estão em jogo o peso que elas de facto teriam. De facto, ele encontra razões que tornam desejável tanto um como o outro curso de acção, mas decide-se por um por considerar ser o mais aliciante; apesar disso, age em sentido inverso. Há irracionalidade porque a razão para fazer o que fez suplanta o próprio princípio da continência – essa não é uma razão contra o princípio em si (não o põe em causa), mas é usada como tal.

Como acabámos de ver, a pressuposição de racionalidade ou irracionalidade na acção depende, em parte, do processo de racionalização – do facto de encontrarmos em nós as crenças e desejos relevantes que nos incitam a agir, e da existência ou não de coerência entre eles. Há irracionalidade se a nossa acção não está em concordância com as crenças e desejos, não estabelecendo uma conexão racional com eles.

A nossa proposta é a de que tal processo de racionalização não será suficiente para dar conta das razões em jogo quando agimos, e portanto não é suficiente para imputar irracionalidade a um agente. Consideramos que o agente é o único que pode dar conta de tais razões, mas sustentamos que apesar de se tratar de um acesso privilegiado por parte do agente, não é um acesso suficiente para alcançar os efeitos pretendidos. E, neste contexto, o próprio processo de racionalização

---

[12] Cf. Davidson, Donald, "How is weakness of the will possible", in *Essays on Actions and Events*, Oxford, Oxford University Press, 1980.

fica fragilizado.

De acordo com Davidson, acções são descrições de eventos mentalistas e conscientes – parte-se então do pressuposto que o agente pode fornecer a explicação para a sua acção. Mas que dizer das acções inconscientes, naquelas situações em que apresentamos uma certa razão que conscientemente acreditamos ser a razão pela qual agimos como agimos, mas inconscientemente fazemo-lo por uma outra razão à qual não temos acesso directo? Temos realmente acesso às verdadeiras razões das nossas acções? É a introspecção um método fiável para acedermos a elas? Até que ponto?

A nossa sugestão é a de que a partir de um certo patamar de complexidade, o agente, sendo a única autoridade para dizer quais foram as razões da sua acção, deixa de ter autoridade suficiente. Se acções mais básicas, motivadas pela satisfação de necessidades biológicas, podem ser explicadas apontando o par crença-desejo que as racionaliza (por exemplo, a minha acção de tomar um copo de água é explicada pela meu desejo de beber água, motivado pela minha necessidade biológica de saciar a sede, conjuntamente com a minha crença de que bebendo o conteúdo do copo mato a minha sede), nas acções mais complexas o agente pode não ter acesso àquelas que são as reais razões pelas quais fez o que fez.

De facto, parece-nos plausível assumir a possibilidade de, na realidade, não sabermos quais são os desejos e crenças que motivam a nossa acção, apesar de conscientemente declararmos que são tais e tais. O próprio agente, em não raras ocasiões, pode duvidar da verdade das intenções declaradas, e essa situação de dualidade é, parece-me, uma prova de que o processo de racionalização não é tão simples e linear como aparece em Davidson.

E se se dá o caso de não podermos saber quais são as crenças e desejos que constituem a verdadeira razão da minha acção, também não poderemos averiguar da irracionalidade entre essas crenças e desejos e a minha acção.

Vejamos novamente o caso da akrasia. A argumentação davidsoniana parte do pressuposto de que nós temos acesso àquelas que são as nossas razões, e somos irracionais precisamente porque não seguimos aquelas que consideramos serem as melhores, indo contra nós próprios. Aceitamos que, teoricamente, é racional afirmar-se que devemos seguir o nosso melhor juízo, mas o problema é que é perfeitamente possível alguém afirmar que uma razão é a melhor, e de facto não ser isso o que pensa. Podemos conscientemente decidir que

um dado curso de acção é o mais acertado porque é o socialmente aceite como mais conveniente, ou porque moralmente nos parece o mais adequado. Mas isso não é motivo suficiente para que queiramos de facto optar por esse curso de acção, nem que o consideremos realmente o melhor. Assim, se o que acabamos por fazer se revela o exacto oposto, devemos ser considerados irracionais por estarmos a ir contra o nosso melhor juízo, ou estaremos simplesmente a fazer o que na verdade queríamos? O ponto é que não podemos saber, a introspecção não é um processo fiável para nos dar as razões (quaisquer que elas sejam) para fazer o que fazemos, pelo menos a partir do patamar de complexidade no qual as acções não constituem a mera satisfação de necessidades básicas.

Neste exacto momento, eu poderia estar a passear em vez de estar a trabalhar. Deparei-me com as duas alternativas (entre outras possíveis) e decidi que tinha melhores razões para optar por trabalhar, pois quero terminar o trabalho que me comprometi a fazer. Em outras ocasiões, porém, tendo feito exactamente o mesmo juízo, acabei por realizar a acção inversa, e fui passear. Nessa altura, deveria ser acusada de me ter comportado de forma irracional, porque fui contra o meu melhor juízo? O meu juízo pode indicar-me o que eu acho que devo fazer, o que não significa que seja mesmo o que quero fazer, logo, ao contrariá-lo, posso não estar a ir contra mim mesma, nem contra o principio de continência, e não devo por isso ser considerado um agente irracional.

Nem sempre aquilo que entendemos ser o melhor a fazer é o que de facto queremos fazer, e mesmo quando nos parece que queremos de facto agir num certo sentido, nem sempre o fazemos necessariamente. E nem por isso devemos ser considerados agentes irracionais. Neste ponto, estamos a defender que os princípios P1 e P2 formulados por Davidson não são, de modo nenhum, evidentes, mas discutíveis.

De salientar que as acções acráticas não constituem uma prova de que as crenças e desejos que formam as nossas razões não são as causas das nossas acções, porque quando o agente age contrariamente ao seu melhor juízo, ainda assim, ele age em virtude de uma razão; o que as acções acráticas atestam é somente a possibilidade de nós não estarmos em situação de saber quais são essas razões. É o processo de racionalização, e não a causação que está em causa.

Auto-engano e confabulação e a possibilidade de dar conta das razões das nossas acções

O auto-engano (*self-deception*) é um caso paradigmático de uma situação em que estão em jogo crenças e desejos inconscientes que tornam o processo de racionalização problemático. Numa tal situação, o que se passa é que alguém acredita que *não-p* apesar de todas as evidências lhe indicarem que *p*. Dado que *p* surge ao agente como tratando-se de algo desagradável, que ele deseja que não fosse real, ele desenvolve a crença, apresentando razões nesse sentido, de que tal não é de facto o caso. E age de acordo com essas razões. O exemplo típico é o do marido enganado que, dados todos os indícios que apontam para a infidelidade da esposa, inventa razões que expliquem esses sinais, e desenvolve a crença de que a mulher lhe é de facto fiel, agindo em conformidade com essa crença. Ele (inconscientemente) sabe que a mulher o engana e (inconscientemente) deseja que tal não seja verdade, pelo que conscientemente acredita na sua fidelidade. Como racionalizar as suas acções nesse contexto? Tem ele acesso às reais razões que as explicam?

Como explica Davidson o auto-engano?[13] Nessas situações, o que se passa é que o agente acredita na ocorrência de *p*, o que o incita a formar a crença de que *não-p*. Essa não é uma crença racional, na medida em que o desejo do agente de alterar a sua crença inicial não é uma razão para tomar a crença derivada como verdadeira (não racionaliza o facto de chegarmos a acreditar que *não-p*). E como é possível que coexistam essas crenças contraditórias, que o agente ao mesmo tempo aceite e rejeite uma proposição? Davidson afirma que as duas crenças opostas só podem coexistir se de alguma forma se mantiverem separadas, não conjuntamente acessíveis à consciência.

Temos, portanto, uma crença e um desejo inconscientes que são a causa de outras crenças e a explicação para algumas das minhas acções, sem que, no entanto, tenhamos acesso a eles e possamos portanto saber quais as reais razões dessas acções.

Outro caso que nos parece pertinente ter em conta é o da confabulação (*confabulation*). Em psiquiatria, a confabulação é tida como uma desordem mental acompanhada de uma certa anomalia a nível neurológico. Não é esse naturalmente o caso que nos preocupa,

[13] Cf. Davidson, Donald, "Who is fooled?", in *Problems of Rationality*, Oxford, Oxford University Press, 2004.

mas antes os episódios de confabulação nas ditas pessoas "normais".

Um caso clínico de confabulação consistiria, por exemplo, numa situação em que um paciente fantasiasse acerca da sua condição, e, sendo-lhe perguntado pelas suas actividades mais recentes, ele relatasse de forma coerente toda uma história onde descreve uma ida a Paris ou uma longa reunião de trabalho, apesar de, na realidade, se encontrar internado há já vários dias. E ele genuinamente acredita na sua história. Acresce ainda que todas as restantes faculdades mentais se mantêm inalteradas – ele sabe perfeitamente quem é, reconhece as pessoas que lhe são próximas e mostra-se perfeitamente lúcido.

Mas a confabulação não ocorre apenas em circunstâncias patológicas – alguns estudos apontam para a ocorrência de tais fenómenos também em pessoas mentalmente sãs, em circunstâncias específicas. Podem ser afectadas por ela tanto crianças de tenra idade, como pessoas sujeitas a hipnose ou mesmo pessoas que tentam justificar as suas escolhas ou descrever estados mentais,[14] inventando histórias plausíveis. De acordo com esses estudos, a confabulação seria frequente, por exemplo, nas ocasiões em que nos perguntam alguma coisa para a qual não sabemos a resposta, mas não nos permitimos afirmar tal coisa por acharmos que é algo para o qual devíamos ter uma explicação. Então, inconscientemente "confabulamos", e inventamos histórias acerca de nós próprios e dos nossos putativos estados mentais, preenchendo a lacuna. O autor do livro a que nos reportamos apresenta o exemplo daquelas situações em que nos perguntam porque é que gostamos de alguém – é algo para o qual podemos não ter uma justificação (pelo menos consciente), mas normalmente tendemos a justificar tal comportamento recorrendo a razões que nos parecem viáveis. O ponto é que essas podem ser razões que nos soam convenientes no momento, mas podem não ser as reais razões que explicam muitas das nossas acções.

É possível sustentar, dados esses episódios banais, que nem sempre temos "acesso fiável ao que vai nas nossas mentes; isto é, que a introspecção não é para ser entendida segundo o modelo da visão, e que relatos introspectivos não são similares aos relatos de eventos vistos"[15]. O próprio Dennett, citado pelo autor em causa, diz que "há circunstâncias nas quais as pessoas estão simplesmente enganadas

[14] Cf. Hirstein, William, *Brain Fiction – Self-deception and the Riddle of Confabulation*, Cambridge: The MIT Press, 2005, Capítulo 1.
[15] Hirstein, William, *Brain Fiction – Self-deception and the Riddle of Confabulation*, Cambridge: The MIT Press, 2005, p. 14.

acerca do que estão a fazer e como o estão a fazer. (…) Elas não têm nenhuma maneira de «ver» (através de um olho interior, presumivelmente) os processos que governam as suas asserções, mas isso não as impede de exprimir as suas sentidas opiniões".[16]

Aparentemente, os casos de confabulação em pessoas mentalmente saudáveis parecem dar azo à ideia de que podemos julgar ter feito algo em virtude de determinadas razões, e apresentá-las, mas de facto nada nos garante que tenha sido por elas que agimos como agimos. Na realidade, podemos mesmo não saber que razões causam as nossas acções, mas continuamente criamos a ideia de que sim, para que tais acções façam para nós sentido.

Defendo a plausibilidade de uma tal ideia, mas defendo igualmente que o facto de nem sempre sabermos o que nos leva a agir de certa forma não implica que a nossa vida mental consciente não tenha nenhum papel causal na produção do nosso comportamento; simplesmente não conta a história toda, e sustenho que a parte de nós à qual não temos acesso cognitivo, mas que influi no nosso modo de agir, deve ser levada em conta se o nosso propósito é analisar a acção de um agente real.

Passagem da Teoria da Acção à Ontologia

Esse é o tópico que nos propomos, em seguida, analisar – o papel da causação mental e se ele é salvaguardado no âmbito da filosofia davidsoniana.

No que toca à sua filosofia da acção, Davidson claramente quer defender o papel activo que a nossa mente exerce sobre o mundo físico, ao sustentar que as razões são causas da acção. De facto, como tivemos oportunidade de ver, o autor deixa absolutamente claro que, para explicar uma acção, é necessário encontrar as razões pelas quais a acção foi realizada, ou seja, é preciso que as crenças e desejos que formam essas razões estejam causalmente implicadas na produção da acção. É neste contexto que se encerra o aceso debate entre wittgensteinianos e davidsonianos acerca de saber se as razões podem ser causas – se a causação mental é possível. Aqui não me ocuparei com esse debate mas tão-só com a questão de saber se o objectivo de Davidson – a defesa da causação mental – é realmente atingido.

É verdade que Davidson explicitamente assume a existência de

[16] Dennett, Daniel, *Consciousness Explained*, Boston: Little, Brown, 1991, p. 94.

uma interacção entre o mental e o físico. É no artigo *Mental Events* que ele apresenta o princípio que designa por Princípio da Interacção Causal, no qual sustém que pelo menos alguns eventos mentais interagem causalmente com eventos físicos. Isso significa que a interacção pode ter os dois sentidos: alguns eventos mentais são causados por certos eventos físicos (exemplo de Davidson: o apercebimento de que um navio se aproxima é causado pela aproximação real, física, do navio), e, similarmente, assume-se que alguns eventos físicos são causados por certos eventos mentais (exemplo: o facto de alguém ter, por exemplo, afundado esse navio deve ter sido causado por certos eventos mentais tais como juízos, decisões, crenças e desejos, intenções).

Mais, ele considera que seria impossível referirmo-nos a acções intencionais sem recorrermos às razões, às causas mentais que as sustentam. Como vimos, é isso que distingue uma acção de um qualquer evento que ocorre no mundo – o facto de ser susceptível de uma descrição mentalista. Se explicássemos/descrevêssemos o acto de beber um copo de água, por exemplo, em termos puramente físicos, estaríamos a falar de eventos cerebrais ou de movimentos físicos, mas nunca de acções. Só podemos falar de acções quando podemos falar de razões (crenças e desejos) que as expliquem. Por isso é que a proposta davidsoniana é tanto, a meu ver, uma resposta para a questão da explicação da acção como para a questão da natureza da acção.

Mas se a sua filosofia da acção aponta no sentido de salvaguardar o papel da causação mental, a proposta ontológica que servirá de base a essa filosofia torna essa noção problemática, como veremos. No intuito de saber se a teoria davidsoniana no seu todo permite enquadrar a possibilidade da causação mental, tal como ele sugere, será necessário explanar a sua ontologia, e perceber qual o lugar que Davidson efectivamente concede ao mental no mundo físico.

Antes disso, porém, valerá a pena explicar por que razão a possibilidade ou não da causação mental me parece uma questão importante a ter em conta. E é-o por vários motivos. Primeiro, porque dessa ideia depende a própria noção de agência. Classificarmo-nos como agentes e considerar que aquilo que fazemos como acções exige que os nossos estados intencionais, as nossas escolhas e decisões, sejam de facto os propulsores daquilo que acontece, caso contrário não nos podemos considerar autores daquilo que fazemos. Queremos poder continuar a dizer que é o nosso querer ir à Índia que faz com que iniciemos um conjunto de acções que nos conduz até lá, ou, nas

palavras de Fodor, "que o meu querer é causalmente responsável pelo meu ir buscar, o meu sentir comichão causalmente responsável pelo meu coçar e o meu acreditar causalmente responsável pelo meu dizer (…), [pois] se nada disto é literalmente verdadeiro, então praticamente tudo o que eu acredito acerca do que quer que seja é falso e é o fim do mundo."[17] Será difícil considerarmo-nos a nós próprios como agentes se o que nós pensamos não tiver nada a ver com aquilo que ocorre no mundo, se pelo pensamento não pudermos causar coisas no mundo.

Em causa está, por isso também, a própria possibilidade da acção livre e da acção moral – se os movimentos do meu corpo não puderem ser explicados pelos meus estados mentais, pelas minhas intenções e decisões, então como posso ser responsabilizado pelo que faço? E como posso ser livre, de que forma a minha "acção" se distinguiria da de um autómato previamente programado?

Acresce ainda que o problema da causação mental, por envolver a questão da ligação do físico com o mental, é um dos pontos a considerar quando se tem em vista o problema mente-corpo, e um tópico fundamental quando nos interessa estabelecer o lugar da mente na natureza, questão esta essencial na filosofia da mente.

A proposta ontológica de Davidson e a possibilidade da causação mental

Qual é, então, o lugar dos eventos mentais e dos eventos físicos no mundo, como podem ser definidos e como se relacionam entre eles, de acordo com Davidson?

O que significa dizer que um evento é mental ou físico? Saber o que os distingue é importante para a questão de saber como interagem entre si. Para Davidson, um evento será uma entidade concreta particular (singular e irrepetível) passível de ser descrita de diferentes formas. Assim, o que faz com que um evento seja físico (ou mental) é o facto de ser susceptível de ser descrito em termos físicos (ou mentais). Este é um ponto a reter para perceber o esquema davidsoniano, pois se a distinção entre eventos físicos e mentais fosse ontologicamente relevante, estaríamos em presença de um dualismo de substâncias e a questão da interacção causal entre eles seria bem mais intricada (era essa, de resto, a principal dificuldade de Descartes – reconciliar o seu dualismo com a estreita relação que vemos existir entre mente e corpo).

[17] Fodor, Jerry, *A Theory of Content and Other Essays*, MIT Press, 1990, p.156.

Assim, a ideia fulcral em Davidson consiste no seguinte: aquilo que ocorre pode ser descrito como sendo físico ou mental. O mundo é de uma só natureza, e é físico (por isso Davidson é monista, e não dualista), mas pode ser descrito de diferentes formas. É esta ideia que vai resolver o paradoxo que parece existir entre duas suposições de Davidson – o papel causal dos eventos mentais no mundo físico e a anomalia do mental, isto é, o facto de esses eventos mentais não serem susceptíveis de ser capturados numa rede nomológica causal estrita como aquela que existe entre os eventos físicos.

A sua proposta ontológica vai receber o nome de monismo anómalo e pretende ser uma resposta ao problema de saber qual o lugar do mental num mundo fundamentalmente físico.

Monismo precisamente por sustentar que todos os eventos são eventos físicos, e portanto todo o evento que possa ter uma descrição mental, tem necessariamente também uma descrição física. No entanto, enquanto mentais, os eventos não podem ser subsumidos a leis estritas, porque não existem leis causais determinísticas no domínio do mental, daí que seja considerado anómalo. Como é possível, então, defender a possibilidade da causação mental, se para falar em causalidade é preciso falar em leis? Parece haver uma contradição entre o Princípio da Interacção Causal (que atesta a possibilidade da causação mental), o Princípio do Carácter Nomológico da Causalidade (que assume que eventos relacionados causalmente têm necessariamente de ser subsumidos a leis determinísticas) e o Princípio da Anomalia do Mental (segundo o qual não pode haver leis de tipo estrito e determinístico que permitam explicar ou prever eventos mentais).

Ou seja, Davidson defende que as razões são causas que intervêm no mundo físico, fazendo-se aí sentir os seus efeitos, mas ao mesmo tempo sustém que a relação entre dois acontecimentos só é causal se for uma instância de uma lei estrita (*a* é causa de *b* se a ocorrência de *a* garantir a ocorrência de *b*, ou, dito de outra forma, *b* tem necessariamente de acontecer se *a* tiver lugar) e que o domínio do mental é tal que não é possível elaborar leis desse carácter – é impossível dizer que nas ocasiões em que o evento mental M1 ocorre, o evento mental M2 necessariamente terá lugar, assim como é impossível estabelecer uma correlação de carácter determinístico entre eventos físicos e eventos mentais, do tipo: quando M1 ocorre, seguir-se-á obrigatoriamente do evento físico F1. Por outras palavras, não existem leis psicofísicas.

A resposta para um tal dilema é precisamente aquela que

começamos por ressaltar – é a resposta monista de acordo com a qual tudo o que ocorre é físico, mas alguns eventos físicos podem ser descritos de um ponto de vista mentalista. Davidson apresenta, como o próprio afirma, uma versão da Teoria da Identidade entre o mental e o físico, uma teoria da identidade entre *tokens* (exemplares). De acordo com essa versão, cada evento mental é *token-identical* a algum evento físico. Desta forma, é possível falar de leis determinísticas na medida em que, sob a sua descrição física, um evento pode ser subsumido a uma lei causal estrita. Assim, um mesmo par de eventos pode instanciar uma lei (causal, precisa) quando é apresentado sob uma descrição (física), e não quando surge sob uma outra descrição (mental). Ou seja, um evento descrito como mental (uma razão) é também um evento físico que causa algo no mundo – será o mesmo evento, mas sob descrições distintas; enquanto descrito em termos mentalistas, nenhuma lei estrita pode ser estabelecida, enquanto descrito em termos físicos, já é possível chegar a uma tal lei. Tal como podemos ler em *Mental* Events, "a causalidade e a identidade são relações entre eventos individuais independentemente de como estes são descritos. Mas as leis são linguísticas; e por isso os eventos podem instanciar leis, e dessa forma serem explicados ou previstos à luz dessas leis, apenas na medida em que eles são descritos de uma determinada forma. (…) O princípio do carácter nomológico da causalidade deve ser lido cuidadosamente: ele afirma que quando os eventos se relacionam em termos de causa-efeito, eles apresentam descrições que instanciam uma lei. Não diz que cada afirmação singular verdadeira de causalidade instancia uma lei."[18] Davidson distingue entre os eventos e a forma como eles são descritos linguisticamente; a causalidade é uma relação entre eventos que se estabelece não importa como esses eventos são apresentados, mas as leis causais só se estabelecem quando os eventos são descritos da maneira apropriada.

Por esta via pretende Davidson mostrar como a causação mental é uma realidade, apesar de os eventos mentais não poderem ser capturados numa rede causal de tipo determinístico, por formarem um domínio "anómalo". ***Qual é então o lugar do mental num mundo que é físico?***

O mental pode ser descrito em termos físicos, mas não pode ser reduzido ao físico, precisamente porque enquanto os eventos físicos

[18] Davidson, Donald, "Mental Events", in *Essays on Actions and Events*, Oxford, Oxford University Press, 1980, p. 215.

estão relacionados entre si de uma forma que pode ser explicada e prevista recorrendo a leis, os eventos mentais escapam a essas leis. O mental é, pois, nomologicamente irredutível ao físico. Como o próprio atesta: "nós sabemos demais acerca do pensamento e do comportamento para confiar em afirmações universais e exactas a correlacioná-los"[19] e tais considerações "fornecem pelo menos indícios de que não devemos esperar ligações nomológicas entre o físico e o mental"[20]. Davidson considera que o domínio mental é necessariamente holista, isto é, nós atribuímos sentido às crenças de uma pessoa na medida em que elas são coerentes com outras crenças, preferências ou intenções. Nós só conseguimos identificar pensamentos e saber o que eles significam porque conseguimos inseri-los no interior de uma rede de crenças relacionadas que lhe dão sentido[21]. Mas não há leis a governar essa atribuição de crenças a outrem, nem podemos prever o que alguém fará ou pensará com base no que ele nos mostrou até agora. Enquanto que para um qualquer evento físico pode ser traçada a sua história causal e determinística, o mesmo não acontece quando falamos de eventos mentais. "É característico da realidade física que uma alteração física possa ser explicada através de leis que a relacionam com outras alterações e condicionamentos fisicamente descritos. É característico do mental que a atribuição de fenómenos mentais dependa do background de razões, crenças e intenções do indivíduo."[22] As leis físicas não se aplicam aos eventos enquanto estes são descritos como mentais. Conhecer todos os factos do mundo físico não nos permite prever ou explicar qualquer evento mental de qualquer pessoa. O facto de não existirem leis psicofísicas a unir o mundo mental ao físico implica que não possa haver uma redução das descrições mentais às descrições físicas. Daí que a ciência física não possa dar conta dos eventos descritos enquanto mentais.

Há, pois, uma independência nomológica do mental relativamente ao físico. Isto significa que para falar do mental precisamos de razões, crenças, intenções – de noções mentais. De igual modo, as explicações acerca do mental não podem ser dadas recorrendo ao vocabulário das

[19] Idem, p. 217.
[20] Ibidem.
[21] Cf. Davidson, Donald, "Rational Animals", in *Subjective, Intersubjective and Objective*, Oxford, Oxford University Press, 2004.
[22] Davidson, Donald, "Mental Events", in *Essays on Actions and Events*, Oxford, Oxford University Press, 1980, p. 222.

ciências físicas. Os instrumentos conceptuais que usamos para pensar no mental não são os mesmos que utilizamos para pensar no mundo físico. Por isso Davidson distingue a explicação nas ciências físicas (que consiste na subsunção a leis) da explicação da acção humana (que consiste em apresentar a razão primária que racionaliza a acção). Ter um conhecimento completo das causas físicas que explicam o movimento de um corpo não permite explicar porque é que um agente agiu da forma como agiu. Para isso precisamos de causas mentais, precisamos de razões.

Mas isto não significa, como assinalámos anteriormente, que existam dois mundos, ou duas substâncias. O mundo é o mesmo, e é físico. Simplesmente, há coisas que se passam nesse mundo físico que têm que ser descritas e explicadas recorrendo a outro vocabulário, caso contrário não seria possível atribuir-lhes sentido.

De facto, Davidson é monista, porque é partidário de uma versão da Teoria da Identidade segundo a qual cada evento ou exemplar mental particular (*token*) é idêntico a algum evento ou exemplar físico particular (*token*), ou seja, há uma identidade entre acontecimentos individuais, datados e irrepetíveis. Uma tal versão, que se designa por teoria da identidade exemplar-exemplar, diverge duma teoria da identidade tipo-tipo, de acordo com a qual não são apenas os exemplares de eventos mentais que são idênticos a exemplares de eventos físicos, mas são os próprios tipos ou propriedades mentais que se identificam com os tipos ou propriedades físicos. Antes de Davidson, as teorias identitativas da mente comummente defendiam que a identidade entre eventos físicos e mentais dependeria da descoberta de leis que correlacionassem as propriedades mentais às propriedades físicas de uma forma clara. Em directa oposição a essas teorias, o autor vai sustentar que os *tipos* ou *propriedades* mentais são distintos e irredutíveis aos *tipos* ou *propriedades* físicos, precisamente porque aqueles não são governados por leis físicas. Desta forma, Davidson vai sustentar a identidade entre eventos mentais e físicos (*token-identity*) mas argumentando contra a existência de identidade entre propriedades físicas e mentais *(type-identity*). Falar de propriedades para Davidson é falar das maneiras como um evento pode ser descrito, de tal maneira que "certos particulares simples tanto podem ser enquadrados em categorias que configuram um discurso mental como em categorias que configuram um discurso neurofisiológico ou outro"[23]. Assim, se os

[23] Zilhão, António, "Fisicalismo", in Branquinho, J. e Murcho, D. (org), *Enciclopédia de*

exemplares se identificam, não é possível, no entanto, reduzir as propriedades mentais às físicas, porque estas são apenas maneiras distintas de descrever os mesmos exemplares, não são nada que possa ser incluído numa ontologia.

No esquema davidsoniano, a relação que existe entre as propriedades físicas e mentais não é de identidade, mas de *superveniência*. Davidson é o primeiro filósofo a emprestar a um tal termo um significado filosoficamente relevante. De acordo com o próprio, uma tal relação implica que "as características [ou propriedades] mentais são de alguma forma dependentes ou supervenientes relativamente às características [ou propriedades] físicas. Isso significa que não podem existir dois eventos semelhantes em todos os aspectos físicos mas diferindo em algum aspecto mental."[24] Segundo a sua tese, portanto, existe uma identidade entre os eventos mentais e os eventos físicos que lhes correspondem, porque tudo o que é mental é também físico. O meu querer ler um livro, que conduz ao meu acto de pegar num livro para ler, é equivalente a um evento físico que ocorre no meu cérebro e ao subsequente movimento do meu corpo. Mas, dada a irredutibilidade nomológica do mental, não é possível explicar uma tal acção em termos puramente físicos, pelo que precisamos de recorrer às propriedades ou descrições mentais. Há, portanto, uma identidade entre eventos mas não uma identidade entre propriedades ou descrições. O que existe é antes uma dependência do mental relativamente ao físico, expressa nessa relação de superveniência, segundo a qual quando há uma modificação em algum aspecto mental, é porque houve uma alteração em qualquer aspecto físico.

Alguns autores sustentam que a proposta de Davidson conforma um dualismo de propriedades, dada a irredutibilidade das categorias mentais às físicas[25]; no entanto, o dualismo de propriedades é incompatível com o fisicalismo de que Davidson se diz defensor. A ideia de superveniência vem resolver essa questão – de acordo com essa noção, as propriedades mentais não deixam de ser físicas, pois há uma dependência que Davidson supõe ser metafísica das propriedades mentais em relação às propriedades físicas. Apesar de uma descrição

Termos Lógico-filosóficos, Gradiva, Lisboa, 2001.

[24] Davidson, Donald, "Mental Events", in *Essays on Actions and Events*, Oxford, Oxford University Press, 1980, p. 214.

[25] Cf. Kim, Jaegwon, *Mind in a Physical World – An Essay on the Mind-Body Problem and Mental Causation*, Cambridge: The MIT Press, 1998, p.58.

ser nomologicamente irredutível à outra, as descrições mentais não deixam de ser descrições de algo que em última analise é também físico, e é por isso que as propriedades mentais são determinadas pelas físicas – são supervenientes em relação a elas.

A relação que se estabelece entre o mental e o físico é, pois, uma relação de superveniência, tendo no entanto sempre em atenção que o mental é uma descrição de algo que é fundamentalmente físico. Apesar disso, Davidson sustenta igualmente que o mental causa, isto é, tem efeitos no mundo físico.

O monismo anómalo de Davidson é uma proposta ontológica bastante atraente precisamente na medida em que consegue conciliar o monismo (trata-se, portanto, de uma proposta fisicalista) com a anomalia – o facto de as descrições mentais não poderem ser reduzidas às descrições físicas.

Ainda assim, o seu modelo foi alvo de algumas críticas, que apontam tanto para a inconsistência entre algumas das suas propostas como para a incapacidade de salvaguardar um dos tópicos mais importantes no seu pensamento – a causação mental.

Críticas à proposta davidsoniana

A primeira crítica vai no sentido de pôr em causa a compatibilidade entre a noção de superveniência e as outras teses defendidas por Davidson. A tese da superveniência apela para uma dependência do mental em relação ao físico (na exacta medida em que se dois eventos partilham todas as suas propriedades físicas, necessariamente partilharão todas as suas propriedades mentais), mas sem implicar uma redução do mental ao físico, pelo que protege uma certa autonomia do mental. Uma tal posição deu origem ao chamado fisicalismo não-redutivo, uma proposta metafísica muito em voga actualmente.

Alguns autores sustêm que se se diz que não podem existir dois eventos semelhantes em termos físicos que sejam dissemelhantes em algum aspecto mental, então alguma lei do tipo psicofísico está a ser instanciada, pois está-se a afirmar que um estado físico F1 implica a ocorrência de um tipo de estado mental M1, de tal forma que se outro objecto partilhar o mesmo estado físico F1, necessariamente experienciará o mesmo estado mental M1. A resposta de Davidson será a de caracterizar tais correlações como meras generalizações, leis do tipo *ceteris paribus* e nunca leis estritas, precisas, do tipo

determinístico.

Davidson sustentará que o estabelecimento de relações de superveniência incorpora em si um certo poder explicativo, pois permite observar correlações entre estados mentais e físicos, mas não permite prever, tal como acontece nas ciências físicas, que estados mentais dão origem a que estados físicos, e vice-versa. O que se pode afirmar dadas tais relações é que se um evento mental ocorre, deve haver uma qualquer explicação física desse evento (até porque tudo o que é mental, é também físico). Uma mudança das propriedades físicas coincide com uma mudança nas propriedades mentais, e há pois, uma relação de determinação do físico para o mental. Se não se observasse e relatasse qualquer relação de superveniência, essa ocorrência mútua seria em si um facto bruto, sobre o qual nada saberíamos. No entanto, observar tais relações não nos permite estabelecer leis precisas que regulem tais fenómenos.

Em outros autores, porém, a crítica tem o sentido inverso. Jaegwon Kim, por exemplo, argumenta que a noção de superveniência apenas permite identificar um padrão de co-variância entre dois conjuntos de propriedades e portanto não é uma teoria explicativa acerca das relações entre o mental e o físico, como Davidson pretendia. "A superveniência não estabelece uma relação metafisicamente «profunda»; trata-se apenas de uma relação «fenomenológica» entre padrões de co-variação de propriedades"[26]. Será necessário, de acordo com Kim, especificar a relação de dependência entre o mental e o físico que torna possível a superveniência, coisa que Davidson não faz, para que o lugar do mental num mundo que é físico seja esclarecido. O que a tese da superveniência permite atestar é a ideia de que o mental se baseia no físico, mas isso é algo que pode ser defendido tanto por um fisicalista como por um emergentista, contesta Kim.

Outra questão, quanto a mim mais importante, dados os meus propósitos iniciais, é aquela que põe em causa a consistência entre a tese da superveniência de Davidson e a tese da irredutibilidade das propriedades supervenientes relativamente às propriedades que lhe servem de base. É possível duvidar da consistência de uma tal posição, pois se o mental depende do físico, e é determinado por ele, não é essa dependência suficiente para que se possa falar de uma redução entre os dois conjuntos de propriedades? Outra questão ainda – se o mental depende do físico não está o seu poder de causação em risco? Pode o

[26] Idem, p. 14.

mental dar origem a coisas no mundo devido aos seus poderes causais, de forma originária, numa situação em que é determinado pelo seu correlato físico? Se tudo o que eu faço tem uma explicação que pode ser dada em termos físicos, a causalidade mental desempenha ainda algum papel?

Tal questão conduz-nos até ao problema do epifenomenismo, problema que se abate sobre a teoria davidsoniana. O epifenomenismo é a visão segundo a qual os estados mentais não têm quaisquer efeitos no mundo físico. Davidson tem como objectivo a rejeição de uma tal visão – como vimos, ele é um interaccionista, pois acredita que o mental exerce os seus poderes causais sobre o mundo físico. O problema, quanto a mim, é que a sua proposta ontológica parece criar dificuldades à sua própria teoria da acção, onde se compromete com uma posição segundo a qual as razões são causas.

O problema advém do choque entre duas suposições de Davidson. Por um lado, ele é monista e fisicalista. Que significa isso? Ser monista significa que ele sustém que o mundo é um, de uma só natureza. Ser fisicalista não tem um significado tão definido, mas uma proposta bem razoável consiste em afirmar que ser fisicalista é defender que "a física conta toda a história acerca da causação de *eventos físicos*: isto é, eventos que têm traços ou propriedades físicas. De acordo com o fisicalismo, qualquer coisa de físico que aconteça, qualquer coisa que possa contar como um efeito, tem de ser o resultado de causas puramente físicas, em concordância com uma lei física. Esta é uma doutrina acerca de causalidade."[27] É a doutrina da completude da física. E é algo que Davidson assume [28].

Por outro lado, porém, Davidson assume igualmente a posição segundo a qual o mental intervém no mundo físico, e as nossas acções são explicadas explicitando as razões que as causam. Dizer que as crenças e desejos que formam as nossas razões são as causas das nossas acções é dizer que há efeitos físicos (o acto de pegar num livro, por exemplo) que não teriam ocorrido senão fossem essas causas mentais, e isto é negar a completude da física. Davidson não aceitaria, presumo eu, tal objecção, na medida em que sustenta que tais causas mentais são, também elas, físicas, pois o mental é somente uma descrição diferente daquilo que é físico. Mas, se assim for, cai no problema inverso – torna-

[27] Crane, Tim, *Elements of Mind – An Introduction to the Philosophy of Mind*, Oxford, Oxford University Press, 2001, p. 45.

[28] Cf. Davidson, Donald, "Mental Events", in *Essays on Actions and Events*, Oxford, Oxford University Press, 1980, p. 222.

se epifenomenista. Se assim for, então, é legítimo perguntar para que servem as causas mentais, se as causas físicas são tudo o que é preciso para que um certo efeito físico tenha lugar.

Uma saída possível é a afirmação da sobredeterminação – afirmar que um efeito pode ter mais do que uma causa, e que portanto tanto as propriedades físicas como as propriedades mentais seriam causalmente relevantes para que um dado efeito físico ocorra, de tal forma que se faltasse a causa mental, o efeito teria ainda lugar, e o mesmo se falhasse a causa física. Mas será que esta hipótese faz sentido? Mais, será que é mesmo desta forma que o mental e o físico interagem?

A dificuldade que nesse caso surge advém do problema da exclusividade explicativa, que Kim descreve alongadamente no livro acima citado. O princípio da exclusividade explicativa diz-nos que não podem existir duas explicações independentes e completas – individualmente suficientes para que um dado evento tenha lugar. Se um estado mental M1 causa o estado físico F1, mas M1 tem ele próprio uma descrição física (chamemos-lhe o estado físico F2), então podemos dizer que o estado físico F1 é causado tanto pelo estado mental M1 como pelo estado físico F2. Os dois, independentemente um do outro, são considerados suficientes para causar F1. Surge o problema da abundância das causas, ou da exclusividade explicativa, e, quanto a mim, mantém-se a suspeita de epifenomenismo, pois se se diz que um efeito tem mais do que uma causa, e que esse efeito teria ocorrido de igual forma mesmo que a causa mental não ocorresse (como acontece se defendermos a sobredeterminação), como é possível sustentar concomitantemente a relevância causal do mental?

Davidson poderia ainda contra-argumentar dizendo que as duas causas são de facto uma e a mesma, mas apresentadas segundo diferentes descrições, aplicando-se a diferentes níveis. Mas o problema, quanto a mim, consolida-se em vez de se desvanecer, pois se a causa é na verdade apenas uma (e dado o seu monismo, seria física), para quê falar de causação mental? Como pode ele, nesse caso, afirmar que razões são causas? Tratar-se-á então de uma mera forma de falar, de tornar as nossas acções compreensíveis, mas sem que haja nada de verdadeiramente causal quando apontamos razões como sendo causas? Não me parece que fosse esse o intento de Davidson, pois ele de facto acreditava que o mental exercia efeitos sobre o mundo físico, e tal não pode ser teoricamente sustentado se o discurso acerca do papel causal dos nossos estados intencionais for somente uma vã maneira de falar sem consequências relevantes do ponto de vista filosófico.

Se assim é, então, há claramente uma tensão entre duas explicações causais distintas. O ponto é que as duas causas se apresentam isoladamente como sendo causalmente suficientes para que o efeito tenha lugar. Nas palavras do já citado Kim, "o problema da exclusão causal/explicativa surge nos casos em que existem explicações psicológicas de comportamentos físicos, mas simultaneamente acreditamos que o efeito físico tem, ou deve ter também, uma explicação causal física."[29] É óbvio que podemos explicar e descrever aquilo que vemos ocorrer no mundo de formas distintas, sejam elas físicas ou mentais, mas o problema com que nos deparamos tem a ver com causação e não com explicação; e se formos fisicalistas e acreditarmos na completude da física como Davidson acredita, então o físico só pode ser causado pelo físico, e o mental torna-se supérfluo.

Acresce ainda que, como vimos, não existem leis psicofísicas, leis que relacionem os eventos mentais com os eventos físicos, mas tais redes causais determinísticas podem ser estabelecidas entre eventos apenas de tipo físico. Esta suposição pode acarretar mais dificuldades para a conservação da causação mental no interior do esquema davidsoniano, pois é possível sustentar que somente aquelas propriedades susceptíveis de instanciarem leis precisas podem ter relevância causal; as causas das nossas acções seriam de acordo com uma tal posição puramente físicas, pois é nesse domínio que é possível falar de leis de tipo estrito. O mental seria causalmente impotente, no sentido em que seria apenas em virtude das propriedades físicas que um evento mental causa.

Conclusão

Neste trabalho, propus-me problematizar alguns tópicos da filosofia da acção davidsoniana, tanto no que toca à noção central da sua teoria da acção – a noção de racionalização –, argumentando que tal processo introspectivo nem sempre nos permite chegar a perceber quais são as reais razões das nossas acções, como no que concerne à proposta ontológica que serve de base a essa teoria da acção (crítica essa que foi já sobejamente desenvolvida por inúmeros autores).

O meu ponto principal era a tentativa de mostrar que uma das pedras de toque da sua filosofia – a questão da causação mental – não é

[29] Kim, Jaegwon, *Mind in a Physical World – An Essay on the Mind-Body Problem and Mental Causation*, Cambridge: The MIT Press, 1998, p. 66.

suficientemente assegurada no interior do seu esquema de pensamento. Davidson é um acérrimo defensor do senso comum e da psicologia popular – nós, enquanto agentes e seres pensantes, vemo-nos continuamente a fazer coisas em virtude das nossas crenças e desejos, vemos que o nosso pensamento e as nossas razões causam algo no mundo – e pretendeu sustentar filosoficamente este ponto de vista.

No entanto, através do processo de racionalização nem sempre conseguimos chegar a essas razões – na realidade, podemos mesmo não saber que razões causam as nossas acções, o que faz com que percamos um pouco o domínio que Davidson sustenta que o agente tem sobre si próprio e sobre as razões das suas acções. O autor não leva em conta as razões inconscientes, e por isso atribui ao processo de racionalização uma clarividência quanto a mim exagerada e ao nosso acesso interior um estatuto, não apenas privilegiado, mas absolutamente fiável e incorrigível.

Isto não significa, porém, que possamos afirmar a impotência do mental – porque mesmo que não saibamos quais são exactamente as razões que nos levam a agir, isso não implica que elas, sejam quais forem, não causem algo no mundo físico. Acresce todavia que a proposta ontológica que serve de base à sua filosofia da acção, o monismo anómalo, através da qual estabelece o lugar do mental no mundo físico, acaba por impedir a própria possibilidade da causação mental, pois sendo uma proposta fisicalista, reserva um lugar supérfluo e perfeitamente dispensável para as causas enquanto mentais. Davidson parece não atingir, portanto, os seus intentos.

Quanto a mim, a pergunta crucial não é a de saber como a causação mental é possível, pois ela é uma realidade (é por isso que o problema mente-corpo existe). Torna-se uma noção problemática, porém, se inscrita no interior de um esquema fisicalista, como o que está em consideração. A única saída será, então, negar a completude da física? Qual seria o preço de uma tal rejeição? Aqui atenho-me às palavras de Tim Crane – "negar a completude da física não significa voltar a um dualismo de substâncias, a visão de Descartes. Isto porque alguém pode defender uma teoria monista acerca das substâncias – que todas as substâncias têm propriedades físicas, logo, todas as substâncias são físicas – e ainda assim negar a completude da física, ao negar que *todos* os efeitos físicos sejam inteiramente determinados por causas puramente físicas: em alguns casos, as causas mentais são igualmente

necessárias."[30]

Esta não é totalmente a posição que defendo, pois continuariam a existir casos nos quais teríamos duas causas para o mesmo efeito; no entanto, parece-me interessante a ideia segundo a qual negar a completude da física não é retornar a um dualismo de substâncias, mas seria apenas dizer que há coisas no mundo que não têm que ter causas físicas, ainda que o substrato físico esteja lá quando, por exemplo, pensamos em fazer alguma coisa. Não se nega que os nossos estados mentais tenham correlatos neuronais, que quando eu penso em ler um livro, alguma coisa tenha lugar no meu cérebro, nem tão pouco se rejeita que o acto de pegar num livro tenha sido possível em virtude de alguma ocorrência no meu cérebro. Mas isso significa que foi o que se passou no meu cérebro que me levou a querer ler um livro? Porque um ocorre quando o outro ocorre, quer dizer que tenha que haver anterioridade causal? Penso que a noção de causalidade, quando aplicada às relações entre o mental e o físico, deveria ser concretamente explanada. Pois o que significa dizer que os fenómenos neurológicos causam os fenómentos mentais? Os fenómenos mentais são causados no sentido de possibilitados (na medida em que sem substrato físico não existiriam estados mentais), ou causados no sentido em que é o estado cerebral que determina causalmente um determinado estado mental – que faz com que eu pense, deseje ou queira aquilo que penso (desejo e quero) e não outra coisa qualquer?

Neste último caso, encontrar-nos-íamos então numa situação na qual é o evento físico que explica porque é que eu quis ler um livro, isto é, foi o que aconteceu no meu cérebro que causou o meu querer ler aquele livro e não outro livro qualquer. Os eventos neuronais possibilitam os eventos mentais, são o seu suporte físico, mas é por eles que penso o que penso e ajo em conformidade? É assim tão absurdo supor que o mental explica a acção, e que os eventos físicos não – apenas nos dizem o que concomitantemente ocorreu no cérebro quando eu penso em ler um livro?

Mas então qual seria a relação que se estabelece entre o mental e o físico? Classicamente, uma tal posição poderia ser considerada emergentista – as propriedades mentais (que são de facto causalmente relevantes) emergem a partir da matéria física (a partir de um cérebro altamente desenvolvido), mas não são por ela causadas, nem tão pouco

[30] Crane, Tim, *Elements of Mind – An Introduction to the Philosophy of Mind*, Oxford, Oxford University Press, 2001, pp. 62-63.

são, como as propriedades físicas *tout court*, objectivamente observáveis. Tal posição não nos compromete, penso eu, com um dualismo de substâncias *à la* Descartes. Negar a completude da física implica apenas admitir que há coisas no mundo que não aconteceriam se não existissem mentes. Essas coisas seriam por exemplo acções, e acções a partir de um certo nível de complexidade, pois é obvio que acções do tipo instintivo ou que visam a satisfação de necessidades biológicas (a acção de pegar num copo para beber porque tenho sede) são desencadeadas por processos físicos, apesar de as crenças e desejos também estarem lá. Pode dizer-se que eu pego no copo porque desejo beber, mas é algo físico que leva a que se instale em mim o desejo de beber; parece-me estranho, no entanto, que seja algo físico que instale em mim o desejo de ler um livro, de ir à Índia, de escolher um filme no cinema. O problema será definir a partir de que ponto exacto podemos começar a falar de acções complexas.

O meu objectivo neste artigo, porém, reduz-se tão só à tentativa de mostrar algumas dificuldades que a teoria da acção davidsoniana comporta, principalmente o facto de não dar conta da causação mental – ponto que para mim, e também para Davidson, é absolutamente essencial se queremos tratar da acção intencional humana.

Referências

Branquinho, J. e Murcho, D. (org.), *Enciclopédia de Termos Lógico-filosóficos*, Gradiva, Lisboa, 2001.

Crane, Tim, *Elements of Mind – an introduction to the philosophy of mind*, Oxford University Press, 2001.

Davidson, Donald, "Actions, Reasons and Causes", in *Essays on Actions and Events*, Oxford, Oxford University Press, 1980.

Davidson, Donald, "How is weakness of the will possible", in *Essays on Actions and Events*, Oxford, Oxford University Press, 1980.

Davidson, Donald, "Agency", in *Essays on Actions and Events*, Oxford, Oxford University Press, 1980.

Davidson, Donald, "Intending", in *Essays on Actions and Events*, Oxford, Oxford University Press, 1980.

Davidson, Donald, "Mental Events", in *Essays on Actions and Events*, Oxford, Oxford University Press, 1980.

Davidson, Donald, "Psychology as Philosophy", in *Essays on Actions and*

Events, Oxford, Oxford University Press, 1980.

Davidson, Donald, "Rational Animals", in *Problems of Rationality*, Oxford, Oxford University Press, 2004.

Davidson, Donald, "Incoherence and Irrationality", in *Problems of Rationality*, Oxford, Oxford University Press, 2004.

Davidson, Donald, "Paradoxes of Irrationality", in *Problems of Rationality*, Oxford, Oxford University Press, 2004.

Hirstein, William, *Brain Fiction – Self-deception and the riddle of Confabulation*, Cambridge: The MIT Press, 2005.

Kim, Jaegwon, *Mind in a Physical World – an essay on the mind-body problem and mental causation*, Cambridge: The MIT Press, 1998.

Madeira, Pedro, "O que é o modelo crença-desejo?", in *Intelectu*, nº 9, 2003.

Malpas, Jeff, "Donald Davidson", *The Stanford Encyclopedia of Philosophy (Summer 2005 Edition)*, Edward N. Zalta (ed.),
URL = <http://plato.stanford.edu/archives/sum2005/entries/davidson/>.

Mele, Alfred, "Philosophy of Action", in LUDWIG, Kirk (ed.), *Donald Davidson*, Cambridge, Cambridge University Press, 2003.

Miguens, Sofia, *Racionalidade*, Campo das Letras, 2004.

Robb, David, Heil, John, "Mental Causation", *The Stanford Encyclopedia of Philosophy (Spring 2005 Edition)*, Edward N. Zalta (ed.),
URL = <http://plato.stanford.edu/archives/spr2005/entries/mental-causation/>.

Santos, Ricardo, "Acção e Explicação causal", in Miguens, Pinto & Mauro, *Analyses/Análises*, Porto, Faculdade de Letras da Universidade do Porto, 2006.

Searle, John, *Rationality in Action*, Cambridge, MIT Press, 2001.

Searle, John, *Mente, Cérebro e Ciência*, Edições 70, 1997.

Stout, Rowland, *Action*, Acumen, 2005.

Yalowitz, Steven, "Anomalous Monism", *The Stanford Encyclopedia of Philosophy (Winter 2005 Edition)*, Edward N. Zalta (ed.),
URL = <http://plato.stanford.edu/archives/win2005/entries/anomalous-monism/>.

Davidson on Irrationality and Division

Miguel Amen[1],

Abstract: Donald Davidson in 'Paradoxes of Irrationality' (1982) claims that to understand irrationality one has to postulate a divided mind, absent which one could not make sense of the phenomenon. Here I want to defend his position against some objections advanced by John Heil in 'Divided Minds'. Heil has two complaints against Davidson's theory; first he seems to believe that when we cash out the metaphor of a divided mind to give an account of irrationality, the result is an implausible picture; secondly that there are other models to explain irrationality that do not rely on Division. The idea is that even if the concept of a divided mind could be sufficient to explain irrationality it is not necessary one, and in view of the cumbersome nature of Davidson's explanation it would seem altogether superfluous. This is a serious attack, but to my mind entirely misguided. In this Paper I will show why. First by showing that Heil seems to develop an erroneous account of division and the function of partitioning. Secondly by showing that Heil's model is not consistent with important doctrines that he seems to accept and are central to Davidson. In the end Heil's counterexample is a failure at explanation.

Resumo: Em 'Paradoxes of Irrationality' Donald Davidson defende que para compreendermos a irracionalidade temos que postular uma mente divida, sem o que não conseguiremos dar sentido ao fenómeno. Neste artigo quero defender a posição de Davidson contra algumas objecções avançadas por John Heil em 'Divided Minds'. Heil tem duas grandes objecções à teoria davidsoniana: 1) ele parece acreditar que quando procuramos 'converter' a metáfora da mente divida para termos uma concepção da irracionalidade, o resultado é um quadro implausível, 2) existem outros modelos que explicam a irracionalidade e que não se apoiam na divisão da mente. A ideia é que mesmo que o conceito de mente dividida seja suficiente para explicar a irracionalidade, ele não é necessário, e levando em consideração a natureza complicada da explicação de Davidson ele parece pura e simplesmente supérfluo. Este é um ataque sério, mas na minha opinião totalmente mal dirigido. Neste artigo procurarei mostrar o porquê. Primeiro mostrarei que Heil parece desenvolver uma descrição errónea da divisão e da função da partição. Em segundo lugar mostrarei que o modelo de Heil não é consistente com doutrinas importantes que ele parece aceitar e que são centrais para Davidson. Em última análise, o contraexemplo de Heil falha na tentativa de explicação.

[1] Membro e investigador do *Mind Language and Action Group* – MLAG – do Instituto de Filosofia da Universidade do Porto. Bolseiro de Doutoramento da FCT (Bolsa SFRH/BD/24582/2005).

Donald Davidson in 'Paradoxes of Irrationality' (1982) claims that to understand irrationality one has to postulate a divided mind, absent which one could not make sense of the phenomenon. Here I want to defend his position against some objections advanced by John Heil in 'Divided Minds'. Heil has two complaints against Davidson's theory; first that the conception of a divided mind "for all its apparent straightforwardness…when pushed, seems gratuitously complex" (p.580). He seems to believe that when we cash out the metaphor of a divided mind to give an account of irrationality, the result is an implausible picture; secondly that "there are already available to us simpler, far less cumbersome accounts of the phenomena, accounts that leave room for irrational thoughts and deeds within a framework broadly constrained by charity" (p.580). The idea is that even if the concept of a divided mind could be sufficient to explain irrationality it is not necessary one, and in view of the cumbersome nature of Davidson's explanation it would seem altogether superfluous. This is a serious attack, but to my mind entirely misguided. First Heil seems to develop the wrong model for our understanding of division, and the cumbersome nature already mentioned is thus a result of his conception and interpretation of Davidson, and not, in my view, a proper characterization of Davidson's work on irrationality and division. As we will see Davidson's account does not have the characteristic that makes Heil's reading of him implausible. In fact I think that Heil's account is not only a failure to provide the best model, but is inconsistent with important doctrines that he seems to accept and are central to Davidson. Not only do I think that Davison's account provides a clear model to explain irrationality, it seems to me to provide a necessary one. I will not defend this latter claim in a completely satisfactory way, because I will settle with the weaker claim that Heil's counterexample is a failure at explanation.

Lets begin with the second objection, where it is claimed that "there are already available to us simpler, far less cumbersome accounts of irrationality". What is being asserted is that Davidson's claim that division is a necessary ingredient in certain accounts of irrationality is wrong, since there are available to us other ways of explaining it. In the following we will analyse a different model that Heil takes to be not only sufficient, but also, as we will see later on, a better model to explain irrationality. It will, I hope, be clear why I disagree.

Heil discusses a case of Akrasia, and says that we do not need the model of a partitioned mind to account for irrationality. To take his

example: Wayne acts akratically when he insults a student, because he goes against his better judgment that tells him not to act that way. It is worth quoting Heil in full at this point, he say that

In such cases it may simply be that a certain desire, here the desire to insult Wayne, enjoys motivational clout disproportionate to my assessment of it. In judging what I have most reason to do, then, I assign the desire a relatively low ranking. The desire in question turns out, however, to possess strength disproportionate to its standing and, as a result, I acquire an intention to insult Wayne on its basis and subsequently act on that intention. My action is irrational, not because I fail to act on my strongest desire, but because I act against my considered better judgement, a better judgement that assigns a diminished ranking to that desire. (Heil 89, p.581)

Here we are given a model that, according to Heil, makes sense of Wayne's behaviour. What is not clear is why does he think that he has given an account that makes *sense*[2] of irrationality. Remember however, that what is required in this moment of the dialectics is a different model that is sufficient to *explain* irrationality. Be that as it may, I think that in a way he does not even address the problem. The way that he seems to think that irrationality enters the picture clearly makes the problem invisible. Heil says that an action is irrational because it goes against his best judgement. However this is a description of akrasia and not a description of the source of irrationality. The irrationality enters the picture only because akratic acts lead the agent to hold on to contradictory judgements. However an action, or an intention to perform an action that goes against one's best judgement does not lead one into contradiction. The contradiction enters the picture because one acts against the principle of continence.
Maybe I am going a bit too fast here, so it is convenient now to go into a brief exposition of Davidson's views on the source of the problem of irrationality and his conception of practical reason that is adjacent to it. First to see what needs to be explained and why, because Heil seems to miss it; secondly to see why it is the principle of continence and not acting against one's best judgement that leads to a contradiction.

Why does irrationality pose a problem? Because it leads to paradox. This is explained by Davidson this way

[2] This is not so much a question of explaining irrationality, but as we will se in a moment, of giving an account of its paradoxical nature.

> "The underlying paradox of irrationality, from which no theory can entirely escape, is this: if we explain it too well, we turn it into a concealed form of rationality; while if we assign incoherence too glibly, we merely compromise our ability to diagnose irrationality by withdrawing the background of rationality needed to justify any diagnosis at all." (Paradoxes of Irrationality, p.184)

Now, what needs to be clear is why by assuming others to be incoherence or given to too many inconsistencies we lose the need to explain irrationality.

Succinctly the problem arises because of the nature of propositional attitudes. The fact that "there is a rational element at its core." This comes out clearly in the fact that mental states and action fit into a pattern of logical relations in a way that, for example, an intention can be explained, i.e. rationalized, by referring to a desire and a belief; a desire by referring to the content of other beliefs and values. These logical relations provide explanations, in the form of reasons that not only tell us why someone would do an action by showing us that it makes sense in the light of the contents of his reasons, but that those reasons cause those actions, desires, beliefs, intentions etc. According to Davidson when we interpret someone we try making sense of him by attributing to him mental states in a way that they fit into a rational pattern, endorsing the principle of charity. The problem of inner inconsistency is that it goes against this pattern. To attribute to someone, tout court, that he believes 'p and not p' is to reach a breakdown in the process of making sense of someone in the light of reason. Now, this should not be seen as meaning that upon deviance from a principle of rationality, a rational interpretation is immediately lost. But such stark deviance has holding to 'p and not p' is for Davidson a clear crossing of the shadow line into the domain of the incomprehensible[3]. And once there we lose track of the mental, and once we lose the sense of the mental, questions of rationality have no jurisdiction, and so it makes no sense to pose the irrationality question.
In a way for Davidson, to explain irrationality is to explain how come someone that *prima facie* seems to hold to 'p and not p' in fact holds p and holds not p, but doesn't put the two together. Davidson is explicit –

[3] Compare this with the case of someone that does not make an obvious inference. Here we are much more prepared to accept his failure.

to interpret someone as holding to 'p and not p' is to make a mistake of interpretation.

Now getting back to akrasia and the source of irrationality. If Heil thinks that acting or forming an intention against one best judgment is the source of inconsistency then he would think that somehow, best judgment and action or better, best judgment and intention, present a contradiction. Now, since this, as we will see in a moment, is not Davidson's view, he would have to have a different analysis on matters of practical reasoning. However he does not defend one and seems to accept most of Davidson's views. Heil's explanation might explain, by giving a mechanism, how come one goes against one best judgment, but since that is not the required explanandum it doesn't advance his criticism. To see this we now turn to Davidson's views on practical reasoning, only a brief look, but it will be enough to make matters more clear.

As I said the contradiction consist in one going against the principle of continence, viz, that one should act according to one's best judgement. Davidson claims that irrationality does not enter when we hold reasons for and against an action. The point is that those reasons, or the conclusion of the practical syllogism that support conflicting actions or the formation of conflicting intentions, are not of the form of universal statements, say "all lies are wrong", since every action we perform has positive and negative things to be said for it. The conclusion of practical syllogisms present us with a conditional statement, and so one can have two conclusions, one of which supports a given action and the other that is against it, without leading or having the agent in a contradiction, because being *conditional* judgements, they support an action insofar as certain reasons are given. They are like the conclusions of a piece of statistical inductive reasoning, they are true in so far as the premises are held and cannot be separated from those premises. Basically, conditional judgements do not clash logically speaking. In the same way the akratic is not inconsistent because he is contradicting his best judgement. The point is that there isn't any contradiction there. An action or an intention to act being an unconditional judgement does not clash logically with a conditional judgement. Once again the conditional judgement, and in this particular case we are talking about our best judgement, cannot be separated from the reasons for it, and the unconditional judgement just say that something is good or desirable, period.

Since Davidson seems to think that an intention is an

unconditional judgement, it does not clash with one best judgement.

In the case of Akrasia *the inconsistency enter*s when one considers one of the principles of rationality, the principle of continence. Since the principle is not conditional it *does* enter in contradiction with the akratic act!

So deeming some act akratic is to present one with a problem of interpretation. Someone, following Heil, would have to provide an explanation of how can the same mind that stands for such contradictory judgement or beliefs be explained without being said to hold an open contradiction like 'p and not p'. Such propositional attitudes in a single, undivided mind, that are "present at once and in some sense in operation" would seem to require the attribution that 'p and not p', since not only would he be aware of the attitudes but they would stand in the same web of beliefs. But as it was said before, Davidson thinks that this would be a mistake of interpretation.

Now we reach a point in this work where some assessment of what we have been doing is useful. We have been analysing an example from John Heil's? paper 'Minds Divided' that supposedly gives an explanation of an irrational action that does not presupposes the model of a divided mind. We have seen that this explanation is defective because it misunderstands the nature of practical reasoning and thus the source of irrationality in an akratic act. At same time we have seen that Davidson's conception of practical reasoning, as exposed here, can accommodate, logically speaking, akratic acts and other irrational items. That is, his theory of practical reasoning is compatible with the existence of weakness of will and other irrational items. But this possibility, which Davidson takes very seriously[4], brings a problem of interpretation, with which we will now deal, by presenting Davidson view on division, as giving a factual account of how irrationality is possible and how this account can save our interpretative efforts of the akratic agent.

It seems to me that any account of irrationality that presupposes a picture of practical reasoning with the preceding features needs to postulate a divided mind if is to make sense of inner inconsistency. Davidson claims that the following three theses, that are taken by him to be important conceptual elements of the Freudian thought on the mind, are necessary to provide an explanation of irrationality.

[4] After all Davidson altered considerably his theory of practical reason over the years mostly in order to be able to account for the existence of akratic acts, which he took from granted.

Three Freudian Theses. ('Paradoxes of Irrationality', p.170, 171)

1. **Partitioning**. The mind contains a number of semi-independent structures, these structures being characterized by mental attributes like thoughts, desires, and memories.
2. **Structure**. Parts of the mind are in important respects like people, not only in having (or consisting of) beliefs, wants, and other psychological traits, but in that these factors can combine, as in intentional action, to cause further events in the mind or outside it.
3. **Causal relation**. Some of the dispositions, attitudes, and events that characterize the various substructures in the mind must be viewed on the model of physical dispositions and forces when they affect, or are affected by, other substructures in the mind.

By postulating a partitioning of this kind Davidson is basically extending the process of interpretation and charity, by attribution to the agent sub-divisions that are more rational than the agent as a whole. This process of division should be seen as a metaphor and not as a step into depth psychology – its justification is a result of the logic of the charity of interpretation. Its function is to provide a conceptual division *needed* to explain irrationality.

So lets see how it can handle the case of the aforementioned akratic agent. The akratic person has a reason R, constituted in part by a strong desire to insult Wayne. However his best judgement, having into account all of his reason[5] tells him not to act on that desire. Still he acts on R; R causes him to insult Wayne. Davidson explains this by saying that his desire to insult Wayne has a double role; first it is a reason to perform the insult, which however, as a reason, it is defeated by the sum of his other reasons not to perform the act. So here enters R's second role, the role that is responsible for the irrationality proper; by going against his best judgement, he his in contradiction with a principle that he holds, the principle of continence. Moreover his reason R is not a reason against the principle, it does not logically overrides the

[5] Or having into account a substantial part of his reasons. Davidson sometimes speaks of the better judgement.

principle. In fact since the principle of continence is a principle of rationality, there are no reasons against it[6]. So in its second role R overrules the principle of continence, but R cannot do it by rational means. Davidson solution then is that R causally overrides the principle by being causally related with action or the intention to act without being a reason for the act or the intention. It causes them on the model of pure physical causation. Davidson puts in different non-overlapping parts of the mind those mental states that cannot enter in rational causation. Each partitioning then can be full rational. Akratic actions and irrationality in general arise from mental causes that are not rational causes, and so do not make reasonable the attitudes they cause.

While partitioning and structure make sense of the idea of different parts of the mind with full-blooded intentionality (the intention to insult is a full blooded propositional attitude and thus is part of a web of propositional attitudes) the third of the Freudian thesis says that non-overlapping parts of the mind interact on the model of non-rational causes.

The function of the division is to show us how the agent can fail to stand for his own principles. So we can explain how one person can fail to follow the principle of continence by putting the principle and the intention that goes against it in different parts of the mind, parts that do not have rational access to each other but still interact causally.

Heil's complaint is that this model is too cumbersome. For him this is important. Recall the argument: First he provides a different model and then he tries to show that his model is simpler than Davidson's. We have seen that his model is a failure, but still this latter imputation is still more than he can properly defend. Here is how he pictures it

The cause of the irrational item – here an intention – cannot be a reason for it. A desire to A remains a reason to form the intention to A even when one judges it best not A, hence the desire to A cannot be the irrational cause. What is irrational, however, is my forming the intention to A on the basis of a desire to A together with (i) my judging that, all things considered, I ought not A, and (ii) my acceptance of the

[6] Davidson is a little ambiguous about it, as he seems to say in 'Paradoxes of Irrationality' that only a person that holds the principle could be said to be inconsistent upon doing akratic acts. However in 'Incoherence and Irrationality' he is clear about the matter. The principle of continence, being one of the principles of rationality is not something that a rational creature might decide not to hold. He claim that it is constitutive of being a rational animal to accept those principles. This seems to be right and more in step with his other writings.

principle of continence…The cause of my forming the intention to A, then, is my being in the complex state comprising all (or most) of what is included in partition A [at least the principle of continence, the best judgement, the desire to A] (p.79)

I too think that this model is too complicated and difficult. In particular it does not seem plausible that my best judgement, that goes against A, and the principle of continence are part of the cause of the intention to A. I think however that this is not Davidson's view. Heil goes wrong on two accounts. First, because he pictures the desire to A in the web of propositional attitudes that constitute the intersection of the partitions, it follows that there is an element of rational cause in the process, whereas the model intends to explain irrationality by position a model of non rational cause. There is a partial cause that is also a rational cause, namely the desire A. But then there is still an element of contradiction luring, for he still has a rational element causing something for which it isn't a reason in light of his principle of continence. Moreover, the fact that Heil puts the principle of continence to be part of the cause seems to me somewhat perverse. I see no justification for it.

If the desire to A is to be part of the cause the desire has to be outside of the partitioning that includes the intention. Heil does not see that this is necessary because he thinks that

The relevant cause and effect pair cannot be my desire to insult Wayne and my forming an intention to carry out the insult. The desire is, if anything, a reason for the intention. (p.579)

But the point of partitioning is to find the conceptual boundaries in the mind so that one can make sense of the overruling of a principle of rationality, and the boundary is between the relevant cause and effect. And the relevant cause is what in normal cases of intentional action is *the* reason for the action or formation of the intention. There is nothing else than the desire to A (or the reason that includes the desire A): that is the cause. The point of partitioning is to show how an attitude can cause without being a rational cause. So while or in the process of explaining irrationality the desire to A causes non-rationally. Heil seem for a while to miss this function of partitioning. He seems to look at the logical relation between the desire to A and the intention to A and thus rules them out as being related in a way the one causes the other without being a reason for it. But by being in different partitioning that is what happens, that is the function of thesis 3. I think this view is much more natural. The overruling of the principle of continence is

simple a matter of a reason for A to be a non rational cause of A[7]. The overruling of the principle comes about by a reason that causes non-rationally. It's a matter of blind causality, which does not see reason nor inconsistencies lurking.

Though I am not absolutely sure that this is what Davidson has in mind I find it consistent with what he has to say, and a much better model than that of Heil's.

References

Davidson, D.

1969, 'How is weakness of Will Possible?' In Davidson, 2001a
1970, 'Mental Events' in Davidson, 2001a
1974, 'Psychology as Philosophy' in Davidson, 2001a
1982, 'Paradoxes of Irrationality' in Davidson 2001b
1985, 'Incoherence and Irrationality' in Davidson 2001b
1986, 'Deception and Division' in Davidson 2001b
2001a, *Essays on Actions and Events*, Oxford: Clarendon Press
2001b, *Problems of Rationality*, Oxford: Clarendon Press

Heil, J., 1989, 'Minds Divided' , *Mind*, Vol 98

[7] The paradox is thus resolved because being in different partitions, as in being in different minds, a mental cause can cause without being a rational cause.

Boole e Frege:
matematização da lógica *vs.* logificação

João Alberto Pinto[1]

Resumo: Este artigo contrasta, de um ponto de vista histórico, os modos de pensar a formalização da lógica que estão presentes em G. Boole e G. Frege - e que remetem, respectivamente, para a ideia de matematização da lógica e para a ideia de logificação (da matemática). As diferentes perspectivas destes dois autores têm profundas repercussões no modo contemporâneo de reflectir sobre a natureza formal da lógica. Conclui-se apontando a diferença de posições em que um booleano e um fregeano se encontram relativamente à noção de validade, bem como a dificuldade - para um fregeano - de formular a questão filosófica sobre a prioridade de uma noção semântica ou de uma noção sintáctica de validade tal como estas noções se apresentam no âmbito da metalógica.

Abstract: This article contrasts, from a historical point of view, the attempts at the formalization of logic that can be found in G. Boole and G. Frege, and which are associated respectively with the idea of a mathematization of logic and with the idea of a logification (of mathematics). The different approaches of these two authors have profound repercussions in how the formal nature of logic is conceived today. The article concludes by pointing out the different positions in which a boolean and a fregean find themselves in when considering the notion of validity and also the difficulty - for a fregean - of formulating the philosophical question about the priority of a semantic or syntactic notion of validity as these notions present themselves from a metalogical point of view.

[1] Membro e investigadora do *Mind Language and Action Group* – MLAG – do Instituto de Filosofia da Universidade do Porto e Professor do Departamento de Filosofia da Faculdade de Letras da Universidade do Porto

1

Em meados do século XIX, George Boole teve a ideia de levar a cabo a matematização da lógica – ou da racionalidade tal como ela era *então* encarada na área a que se chama, ainda hoje e no seguimento de Aristóteles, Lógica. A ideia conduziu Boole a escrever essencialmente dois livros. O primeiro, de 1847, tem por título *The Mathematical Analysis of Logic, Being an Essay towards a Calculus of Deductive Reasoning* [MAL]; o segundo, publicado em 1854, *An Investigation of the Laws of Thought on which are Founded the Mathematical Theories of Logic and Probabilities* [LT]. Nos dois casos, a concretização da ideia de matematização da lógica assume uma forma algébrica – de acordo com a qual Boole falou, desde logo, de *uma* álgebra da lógica (LT, pp. 37-38) e, depois, passou a falar-se também de álgebra(s) booleana(s) ou, mais geralmente e hoje em dia, de uma lógica algébrica. Esta concretização da ideia contrasta – parcialmente – com as tentativas de matematização da lógica que assumiam uma forma geométrica em Gottfried Leibniz, Leonhard Euler e Joseph D. Gergonne mas que culminaram, por confluência com o trabalho desenvolvido por Boole, em John Venn e de modo muito mais rico nas investigações lógicas de Charles S. Peirce.

A própria ideia de uma matematização da lógica, porém, contrasta – radicalmente – com a ideia de logificação da matemática que está presente em Gottlob Frege. O primeiro livro de Frege foi publicado em 1879 com o título *Begriffsschrift, eine der arithmetischen nachgebildet Formelsprache das reinen Denkens* [BS], i. e. "notação (ou escrita) conceptual (ou, ainda, ideografia) – uma linguagem formal (ou, mais literalmente, uma linguagem de fórmulas) para o pensamento puro construída como a (ou à imagem da) aritmética (ou, de modo muito mais revelador, segundo o modelo proporcionado pela (linguagem da aritmética)". A tentativa de concretização da ideia de Frege assenta na invenção de *uma nova lógica* – ou numa quase completa renovação da anterior imagem da racionalidade pela articulação desta última com um avanço decisivo na própria Lógica (BS, p. 7). Sob clara inspiração leibniziana (BS, pp. 6-7), Frege propôs-se assim criar uma lógica capaz de aparecer simultaneamente como uma (língua) *characteristica universalis* e como um *calculus ratiocinator* que abrangesse aqueles fenómenos tratados no âmbito da lógica anterior, mas também capaz de revelar (ou, simplesmente, de tornar imaginável) o carácter lógico dos conceitos da matemática pura – tal como sejam, por exemplo e em primeira instância, os conceitos aritméticos de ordem numa série e de

número. Sem pensar nas vicissitudes que atingiram este último aspecto – o aspecto conceptual e propriamente logicista – das posições de Frege, parece razoável considerar que a distinção rígida entre a justificação e a descoberta de uma verdade (BS, p. 5) é solidária em Frege da ideia de uma logificação generalizada do saber (BS, p. 7): «...we can with the greatest expectation of success proceed to fill the gaps in the existing formula languages, connect their hitherto separated fields into a single domain, and extend this domain to include fields that up to now have lacked such a language.» Deste ponto de vista, não é de admirar muito que Frege (BS, p. 7) se revelasse confiante no facto de que a sua ideografia podia ser «...successfully used wherever special value must be placed on the validity of proofs» – em *qualquer* lugar ou contexto, quer se tratasse de lugares ou contextos matemáticos, quer se tratasse de lugares ou contextos não matemáticos. Também se compreende bem que Frege tivesse que pensar sempre na sua lógica como universal – no sentido em que essa lógica devia servir para esclarecer *o que é uma prova quaisquer que fossem os objectos aí em causa*. De acordo com isto, a lógica apresentada por Frege no BS não possui então, para além de três funções (propriamente) lógicas que correspondem à negação, à condicionalização e à quantificação universal, senão variáveis. Estas variáveis podem ser variáveis proposicionais, notacionalmente assimiladas por Frege a nomes (de objectos) arbitrários, variáveis funcionais (para conceitos arbitrários) autorizando uma quantificação pelo menos de segunda ordem ou, claro, variáveis individuais para absolutamente *todos os objectos* – ou para um objecto qualquer da totalidade de objectos que há – e não para objectos deste ou daquele tipo específico.

O modo de pensar de Frege contrasta com a ideia avançada por Boole (LT, p. 42) no momento em que retomava a noção de universo *do discurso* já antes usada por Augustus De Morgan: «In every discourse, whether of the mind conversing with its own thoughts, or of the individual in his intercourse with others, there is an assumed or expressed limit within which the subjects of its operation are confined. The most unfettered discourse is that in which the words we use are understood in the widest possible application, and for them the limits of discourse are co-extensive with those of the universe itself. But more usually we confine ourselves to a less spacious field.» Pode certamente pensar-se que assim como a universalidade (fregeana) – presente, antes de mais, em "para todo o x ..." – faz (pelo menos) sentido para Boole, também acaba por ser relativamente fácil proporcionar um sentido para

a noção (booleana) de universo do discurso no âmbito da lógica de Frege – por meio de "para todo o x, se x está em *u*, então ...". Porém, o ponto é aqui apenas o de que Frege julgou completamente dispensável o modo – relativizado, dir-se-ia – de pensar sobre a racionalidade que a afirmação de Boole exemplifica e que, aliás, Frege também reprovou na concretização de alguns pontos do programa da metamatemática que David Hilbert tentava levar a cabo. De facto, este ponto não é mais que uma das várias maneiras de compreender por que razão Frege pôde afirmar que a sua lógica – ao contrário da lógica de Boole e de Erwin Schröder, um dos seguidores directos de Boole – não era apenas (ou não era meramente) um *calculus ratiocinator*.

Pode também pensar-se que Frege tinha o objectivo de fornecer uma análise da noção de prova – independentemente da universalidade da lógica específica com base na qual essa análise se desenvolveria ou, de um modo similar, independentemente da adequação (ou desadequação) da imagem da racionalidade que a universalidade (ou a falta de universalidade) na própria Lógica poderia então impôr. No que respeita a esse ponto, acontece que os desenvolvimentos posteriores revelaram o quanto Frege estava *próximo* da verdade ao manifestar a confiança que efectivamente manifestou nos poderes da sua lógica – ou, mais exactamente, nos poderes daquela parte da sua lógica que é hoje conhecida como Lógica de Primeira Ordem com Identidade. Ainda assim, parece razoável pensar que uma tal aproximação à verdade prolonga algumas ideias que já estavam presentes em Boole – e pensar, além disso, que uma das mais salientes consequências dessa aproximação à verdade se encontra ligada a uma espécie de falhanço nos propósitos mais estritamente leibnizianos de Frege. Por um lado, parece dever-se a Boole a distinção entre uma perspectiva sintáctica e uma perspectiva semântica do funcionamento das linguagens lógicas (pelo menos) – muito similar à distinção entre essas duas perspectivas que é absolutamente essencial quando se pensa na actual metalógica. Por outro lado, pode-se considerar que qualquer análise da noção de prova – assim como a discussão da relevância de uma qualquer prova que, por exemplo, recorra à Lógica de Primeira Ordem com Identidade – acaba por envolver alguns resultados provenientes da metalógica, mas que são estranhos a uma lógica concebida simultaneamente como uma (língua) *characteristica universalis* e um *calculus ratiocinator*. Deste ponto de vista, o uso de muitos processos cuja natureza deve ser encarada como matemática – no âmbito da metalógica desenvolvida desde a terceira década do século XX – constitui uma espécie de

recuperação (pelo menos) histórica da ideia de matematização da lógica.

2

Em Boole, a ideia de matematização da lógica tem por base – explicitamente e desde o primeiro parágrafo de MAL – a atribuição de uma importância fundamental a determinado princípio cuja origem remonta a dois matemáticos contemporâneos de Boole: Gregory Peacock (o criador da chamada Álgebra Simbólica) e Duncan F. Gregory (que alargou algumas ideias de Peacock ao âmbito da Análise). Na formulação adoptada por Boole (MAL, p. 3), o princípio combina dois aspectos dos processos matemáticos – ou, de modo mais exacto e atendendo ao trabalho desenvolvido por Peacock e Gregory, respectivamente de uma parte importante dos processos algébricos e dos processos analíticos. De acordo com o primeiro aspecto, os processos aí em causa são processos simbólicos – no sentido em que a natureza (ou, como diz Boole, a validade) de tais processos depende das leis que regulam o uso dos símbolos (ditas, por Boole, leis de combinação) e não da interpretação desses símbolos; o segundo aspecto do princípio, por sua vez, consagra a possibilidade de haver várias interpretações de tais processos simbólicos – todas elas igualmente admissíveis na condição de serem conservadas as leis (ou, nas palavras de Boole, a verdade das leis) que regulam os símbolos.

Atendendo ao facto de que os símbolos a que Boole se refere são para ser encarados, de acordo com o primeiro aspecto do princípio, como formas arbitrariamente fixadas – ignorando-se o sentido, em geral, que tais formas tenham – e dada a exigência de invariância incluída no segundo aspecto do princípio, este princípio pode ser dito *princípio de invariância da forma*. Além disso e recorrendo a uma terminologia que na época começava a ser usada no âmbito da linguística, pode-se ainda pensar de imediato que os dois aspectos do princípio suscitam – respectivamente – uma abordagem sintáctica e uma abordagem semântica das linguagens artificiais a que Boole chama simbólicas.[2] De qualquer modo, o problema de que Boole trata – após a

[2] Estas linguagens simbólicas têm um papel crucial na parte final da introdução a MAL quando está em causa a posição de Boole sobre três temas: o tema do progresso científico em geral e não apenas no âmbito da matemática ou – mais incipientemente do que na matemática (MAL, p. 11) – no âmbito da lógica tradicional; o tema dos poderes do intelecto humano, particularmente no que respeita à manutenção de um (relativo) optimismo sobre o seu uso teórico e prático (MAL, pp. 10, 14); o tema das relações entre a matemática, a lógica e a filosofia – de cuja discussão resulta a célebre associação da

formulação do próprio princípio de invariância da forma – é o da importância desse princípio. Para Boole (MAL, p. 3) há aí um problema pela simples razão de que «...the full recognition of the consequences of this important doctrine has been, in some measure, retarded by accidental circumstances.» Eis a especificação destas circunstâncias históricas (ou acidentais) tal como Boole (MAL, pp. 3-4) a efectua: «It has happened in every known form of analysis, that the elements to be determined have been conceived as measurable by comparison with some fixed standard. The predominant idea has been that of magnitude, or, more strictly, of numerical ratio. The expression of magnitude, or of operations upon magnitude, has been the express object for which the symbols of Analysis ["not less than the ostensive diagrams of ancient geometry" (p. 4)] have been invented, and for which their laws have been investigated.»

A passagem é relevante por duas razões principais. Em primeiro lugar, Boole parece reconhecer aí como pouco razoável a suposição de que na origem – histórica – de uma linguagem simbólica não se encontre uma qualquer interpretação específica dessa linguagem. Deste ponto de vista histórico, a ligação entre os dois aspectos do princípio de invariância da forma é evidentemente um facto assinalável. Note-se que, para Boole, o desenvolvimento (ou, mais literalmente, os avanços) – e já não apenas a origem – da matemática envolve a possibilidade de haver várias interpretações das linguagens simbólicas. Isto explica que os exemplos de Boole (MAL, p. 3), logo após a formulação do segundo aspecto do princípio de invariância da forma, assinalem a relação da álgebra com questões aritméticas (sobre propriedades dos números) ou com a análise (com problemas dinâmicos ou ópticos) e a geometria (por via, pelo menos, da geometria analítica). Em segundo lugar, a passagem é relevante por permitir a Boole articular a questão da importância – realmente fundamental, dir-se-ia – do princípio de invariância da forma com a crítica de uma certa concepção acerca da matemática. Escreve Boole (MAL, p. 4): «The consideration of that view which has already been stated, as embodying the true principle [o princípio de invariância da forma] of the Algebra of Symbols, would, however, lead us to infer that this conclusion ["the notion that Mathematics are essentially, as well as actually, the Science of Magnitude", numa formulação imediatamente anterior] is by no means necessary.»

A crítica de Boole (MAL, p. 4) começa por colocar na base da

lógica à matemática, mais do que à filosofia (MAL, pp. 10-13).

concepção estritamente numérica da matemática um determinado raciocínio – ou, de maneira muito mais precisa, um raciocínio de tipo indutivo: «If every existing interpretation is shewn to involve the idea of magnitude, it is only by induction that we can assert that no other interpretation is possible.» Isto explica, desde logo, o facto de Boole não considerar necessária uma tal concepção. O defeito maior do raciocínio é, no entanto, o de que ele recorre – para conceber a matemática como uma ciência estritamente numérica – apenas à experiência histórica. Numa primeira objecção directa a esta componente do raciocínio, Boole (MAL, p. 4) nota simplesmente o seguinte: «The history of pure Analysis is, it may be said, too recent to permit us to set limits to the extent of its application.»

Uma segunda e mais profunda objecção é a de que o raciocínio procede a uma sobrevalorização da ligação entre os dois aspectos do princípio de invariância da forma. A este propósito, Boole (MAL, p. 4) observa então que mesmo que a diversidade de interpretações numéricas seja encarada como uma condição histórica da origem e do desenvolvimento das linguagens simbólicas em matemática (nomeadamente na álgebra, na análise e na geometria), tais interpretações numéricas não correspondem a qualquer estado definitivo da matemática – nem se pode assumir que elas constituam uma autêntica condição de existência (uma condição universal, como diz Boole) da matemática. Regressando ao modo como inicialmente o problema foi colocado, a predominância e o sucesso das interpretações numéricas das linguagens simbólicas apenas conseguiu protelar o reconhecimento de todas as consequências do princípio de invariância da forma – impondo, ao mesmo tempo (MAL, p. 4), «...the notion that Mathematics are essentially, as actually, the science of magnitude.»

Neste preciso momento, a estratégia de Boole radicaliza-se – e acaba por revelar aquela que é talvez a maior novidade do seu pensamento. Boole está disposto a supôr (MAL, p. 4) que o anterior raciocínio indutivo é (provavelmente) legítimo – para, logo de seguida, reafirmar a suficiência teórica de uma certa definição a que o princípio de invariância da forma conduz. Trata-se da definição daquilo que é um cálculo – escrevendo, agora, Boole (MAL, p. 4) que o princípio de invariância da forma por si só assegura «...the definitive character of a true Calculus, that it is a method resting upon the employment of Symbols, whose laws of combination are known and general, and whose results admit of a consistent interpretation.» A novidade em causa é evidentemente a ideia de que diante de uma linguagem

simbólica – ou, melhor, depois de efectuada a identificação do (verdadeiro) cálculo associado ao emprego (metódico) dos símbolos dessa linguagem – nada pode impedir a tentativa de desenvolver uma interpretação alternativa a qualquer uma das interpretações já existentes dessa mesma linguagem. Note-se que esta ideia concretiza uma determinada observação efectuada alguns anos antes por Peacock. De acordo com essa observação de Peacock, as questões de interpretação poderiam ser posteriores – e não preceder – as questões relativas aos símbolos eles próprios.

Segue-se a formulação do objectivo de Boole (MAL, p. 4): «It is upon this general principle [o princípio de invariância da forma], that I purpose to establish the Calculus of Logic, and that I claim for it a place among the acknowledged forms of Mathematical Analysis, regardless that in its object ["the human intellect", um pouco adiante (MAL, 7) e isso ainda que o objecto imediato de exame possa também ser (um)a linguagem natural (LT, p. 24)] and in its instruments [métodos ou procedimentos] it must at present stand alone.» Um pouco após a publicação de MAL – em carta dirigida a Arthur Cayley – Boole (*George Boole–Selected Manuscripts on Logic and its Philosophy*, pp. 191-192) retomou a distinção entre os dois aspectos do princípio de invariância da forma para salientar que ao pretender que certas operações da álgebra (elementar) vigoram no seu cálculo lógico, «...I mean of course the symbolical operations – those which depend upon laws of combination, not upon interpretation.» Numa outra passagem que integraria o seu terceiro livro de lógica (nunca terminado) e antes de realçar outra vez a independência – literalmente e até certo ponto – dos dois aspectos do princípio de invariância, Boole (*George Boole–Selected Manuscripts on Logic and its Philosophy*, p. 148) escreve: «...generally if any system of symbols subject to formal laws express thought under conditions of interpretation ... the conditions of interpretation do not impose any necessary restriction upon the processes [métodos ou instrumentos] which the formal laws sanction.»

Assim e em primeiro lugar, tem-se que aquilo que Boole chama o seu cálculo lógico não pode ser pensado como uma linguagem simbólica criada mais ou menos a partir do nada – ou à parte da matemática. A concretização do objectivo de Boole apenas se deixa apreender a partir da possibilidade, suscitada pelo princípio de invariância da forma ele próprio, de haver uma nova interpretação para a álgebra – ou, mais precisamente, para uma parte (elementar) da álgebra. Por sua vez e dado que a álgebra é já encarada como uma

linguagem simbólica, o cálculo lógico de Boole aparece aos seus próprios olhos como uma particular re-interpretação – historicamente, pelo menos, enquadrada por interpretações anteriores – de um cálculo (no sentido definido por Boole a partir, mais uma vez, do princípio de invariância da forma) cuja proveniência (pelo menos) é algébrica. Para além disso e em completo acordo com o ponto revelado pela expressão usada pelo próprio Boole em LT (pp. 37-38), os desenvolvimentos históricos posteriores tornaram habitual pensar que o cálculo lógico de Boole é (possui, agora, a natureza de) uma álgebra – a qual se diz booleana, por vezes, mas de maneira não muito rigorosa atentando no próprio cálculo usado por Boole. Eis uma caracterização parcial desse cálculo (tal como ele se apresenta em LT) por meio de onze leis – muitas vezes ditas leis do pensamento tanto em LT, como em MAL – ou axiomas:

(i) $xy=yx$;
(ii) $x+y=y+x$;
(iii) $x(yz)=(xy)z$;
(iv) $x+(y+z)=(x+y)+z$;
(v) $x(y+z)=xy+xz$;
(vi) $x(y–z)=xy–xz$;
(vii) $1x=x$;
(viii) $0x=0$;
(ix) $x(1–x)=0$;
(x) $x+(1–x)=1$;
(xi) $xx=x$.

Note-se que (ix), (x) e (xi) não são leis de combinação no sentido anteriormente intencionado por Boole pois dependem da interpretação de 1 como (um) universo do discurso, 0 como nada (a classe vazia) e x como uma classe qualquer (que tem em 1–x a classe sua complementar). Além disso e se quiser pensar-se nos termos da actual álgebra booleana, o mais notório é, talvez e por um lado, a ausência de uma lei da dualidade similar a (xi) para + mas, também ou por outro lado, que disso não se segue que o cálculo de Boole seja incapaz de representar por exemplo a união e a diferença simétrica entre classes.

Em segundo lugar, acontece que o recurso ao princípio de invariância da forma não autoriza apenas a anterior re-interpretação – a qual procede, em termos gerais, do numérico (algébrico, analítico ou geométrico) para a Lógica. O princípio de invariância da forma é

igualmente crucial para uma outra re-interpretação que ocorre já no âmbito da Lógica, mas que só é assinalada por Boole cerca de um ano depois da publicação de MAL no artigo "The Calculus of Logic". Sem considerar os pormenores desta última re-interpretação, o facto é que ela fica concretizada no momento em que os símbolos do cálculo de Boole primeiramente elaborado para as proposições categóricas (na terminologia de MAL e da tradição aristotélica; para as proposições primárias ou concretas, na terminologia de LT) deixam de referir-se a classes de objectos para se referirem aos valores de verdade verdadeiro e falso – de modo a ter-se então um único cálculo capaz de lidar também com proposições hipotéticas (na terminologia de MAL e da tradição que remonta aos estóicos; com proposições secundárias ou abstractas, na terminologia de LT). Escreve Boole em "The Calculus of Logic" (*Studies in Logic and Probability*, p. 140): «When we pass to the consideration of hypothetical propositions, the same laws and the same general axiom which ought perhaps also be regarded as a law [trata-se aqui de uma espécie de meta-axioma – "equivalent operations performed upon equivalent subjects produce equivalent results" (MAL, p. 18) – visto, por um lado, como justificando quer a aplicação (literalmente) da álgebra à lógica, quer o uso do símbolo = enquanto único símbolo relacional em Lógica, e, por outro lado, como uma espécie de princípio bem mais fundamental (para a Lógica) que o tradicional *dictum de omni et nullo*], continue to prevail: the only difference being that the subjects of thought are no longer classes of objects, but cases of truth or falsehood of propositions.» Em LT – no exacto momento em que inicia o tratamento das proposições secundárias ou abstractas – Boole (LT, p. 159) escreve o seguinte: «The investigation upon which we are entering will, in its general order and progress, resemble that which we have already conducted. The two inquiries differ as to the subjects of thought they recognise, not as to the formal and scientific laws which they reveal, or the methods or processes which are founded upon those laws.»[3]

[3] A ideia volta a ocorrer nos manuscritos destinados ao terceiro livro sobre lógica de Boole. Numa determinada passagem destes textos, Boole (*George Boole – Selected Manuscripts on Logic and its Philosophy*, p. 153) escreve que alguns dos seus resultados «…relate to thought as occupied about *things* but there is also a theory or doctrine of thought as exercised upon propositions – just as in the common logic we have the distinction between the logic of categoricals and that of hypotheticals. In all *formal* respects however, these theories are the same. Only it is to be observed that the science of the forms of thought as exercised about propositions, the fundamental conceptions of truth and falsehood take the place of the fundamental conceptions of existence and non-

3

Cerca de cem anos depois da publicação de MAL, começava a tornar-se evidente a diferença entre a chegada à lógica matemática que Boole tinha concretizado em meados do século XIX e aquela outra específica forma de lógica matemática (ou logística, como então era comum dizer-se também) que surgiu apenas com o BS de Frege.

Ainda assim, uma análise da terminologia consolidada precisamente desde meados do século XX para falar de linguagens formais – em vez de linguagens simbólicas – permite realçar a novidade do pensamento de Boole. Escreve Alonzo Church (*Introduction to Mathematical Logic*, p. 48), de modo bastante similar ao usado por Boole para formular o primeiro aspecto do princípio de invariância da forma – aquele aspecto que assegura as características de um verdadeiro cálculo: «...we begin by setting up, in abstraction from all considerations of meaning, the purely formal part of the language [cujo estudo "...is called *syntax*" (p. 58)], so obtaining an uninterpreted calculus or ... system.» A seguir, Church (*Introduction to Mathematical Logic*, pp. 54, 55) complementa a anterior ideia e observa que, diante de um tal «...system ["the purely formal part of the language"] ..., we still do not have a formalized language until an *interpretation* is provided. ... (This lead us to the subject of *semantics*.)» Por fim e um pouco adiante, Church (*Introduction to Mathematical Logic*, p. 56) nota que a parte puramente formal (ou sintáctica) de uma linguagem formal pode ser desenvolvida (ou estudada) com uma ou mais interpretações específicas em vista – designada(s) então interpretação(ões) principal(is) – mas

---

existence in the science of thought as exercised by things.» A terminologia usada por Boole é completamente coerente e atendendo a esta última passagem, bem como à passagem de LT (p. 159) já citada, têm-se

> (1) leis formais que são as leis de combinação dos símbolos, às quais Boole (MAL, p. 3) se refere para formular o primeiro aspecto do princípio de invariância da forma,
>
> (2) leis científicas, agora reveladas no âmbito de um certo inquérito, teoria ou doutrina simbólica que – de acordo agora com o segundo aspecto do princípio de invariância da forma – está dependente de interpretação (*George Boole – Selected Manuscripts on Logic and its Philosophy*, pp. 191-192) ou expressa pensamento sob condições de interpretação (*George Boole – Selected Manuscripts on Logic and its Philosophy*, p. 148)

e

> (3) métodos ou procedimentos – ou instrumentos – fundados (LT, p. 159) sobre essas leis.

sempre «...retaining our freedom to employ any interpretation that may be found useful.»

A distinção básica, para este modo de pensar, é a distinção entre cálculos (ou sistemas formais, numa designação tomada aqui como sinónima de "cálculo") interpretados e cálculos (sistemas formais) não interpretados. Nos próprios termos de Church, um cálculo (sistema formal) é a parte puramente formal ou sintáctica de uma linguagem formal – cuja parte semântica, por sua vez, se deve à(s) interpretação(ões) fornecida(s) para esse cálculo (sistema formal). Note-se que este modo de pensar assegura uma generalidade para as noções de cálculo (sistema formal) e linguagem formal de acordo com a qual nem todos os cálculos (sistemas formais) e linguagens formais se têm de encarar – por definição – como caracteristicamente lógicos. De outro modo: a pretensão de que um cálculo (sistema formal) é uma lógica depende dele ter uma interpretação susceptível de conduzir à resolução de problemas colocados pela validade de certos raciocínios. Segue-se a descrição efectuada por Church (*Introduction to Mathematical Logic*, pp. 48-49) do desenvolvimento (ou estudo) de uma linguagem formal a partir de um cálculo (sistema formal).

Em primeiro lugar, tem de haver um conjunto (finito ou infinito, mas sempre enumerável) de símbolos primitivos – ao qual se chama vocabulário ou, numa designação que é neste ponto preferível, *alfabeto*: «The vocabulary [alfabeto] of the language is specified by listing the single symbols which are to be used. These are called the *primitive symbols*, and are to be regarded as indivisible in the double sense that (A) in setting up the language no use is made of any division of them into parts and (B) any finite linear [ou, melhor, apenas normalmente linear] sequence of primitive symbols can be regarded in only one way as such a sequence of primitive symbols.»

Em segundo lugar, tem de poder-se decidir – de acordo com um conjunto (finito) de regras de formação – que sequências (finitas) de símbolos primitivos são para encarar como fórmulas ou, dispensando agora qualquer abreviatura, como fórmulas bem formadas: «A finite [normalmente] linear sequence of primitive symbols is called a *formula*. And among the formulas, rules ["following [Rudolf] Carnap let us call them the *formation rules* of the system" (p. 50)] are given by which certain ones are designated as *well-formed formulas* (with the intention, roughly speaking, that only the well-formed formulas are to be regarded as being genuinely expressions of the language).»

Em terceiro lugar, pode ser especificado um certo conjunto (finito

ou infinito, mas sempre enumerável) das anteriores fórmulas bem formadas – às quais se chama axiomas: «Then certain among the well-formed formulas are laid down as *axioms*.» As fórmulas bem formadas que são axiomas de um cálculo (sistema formal) podem encarar-se como simplesmente dadas (num qualquer sentido que permanece aqui indeterminado) ou como notáveis – num sentido que envolve a ligação entre tais fórmulas bem formadas e outras fórmulas bem formadas, às quais se dará então o nome de teoremas, desse mesmo cálculo (sistema formal) ou num outro sentido que envolve imediatamente a(s) interpretação(ões) principal(is) do cálculo (sistema formal). Retomando uma compreensão mais tradicional do termo "axioma" e atendendo ao sentido no qual estão envolvidas a(s) interpretação(ões) principal(is) do cálculo (sistema formal), também pode pensar-se em verdades (encaradas como) evidentes ou, atendendo de novo aos dois sentidos distinguidos, respectivamente em fórmulas bem formadas fundamentais e em verdades fundamentais. Note-se, agora e primeiro, a presença de axiomas em cálculos (sistemas formais) associados a teorias (matemáticas) cuja origem e desenvolvimento histórico dependeu precisamente de axiomas ou cuja reconstrução (num determinado momento histórico) se efectua de forma axiomática. Foi nesta direcção que se desenvolveu o trabalho inicial de Giuseppe Peano – contemporâneo do BS de Frege – mas, também e principalmente, de Hilbert. Nesta situação e tal como acontece no caso de Boole, a obtenção de (alguns, pelo menos) resultados *não tem de ser* concebida como caracteristicamente lógica. Note-se, a seguir, que os axiomas também ocorrem em cálculos (sistemas formais) na base de linguagens formais caracteristicamente lógicas, como seja a de Frege no BS onde os axiomas têm a designação de leis ou juízos do pensamento puro (BS, pp. 28-29) – mas que não é necessária a existência de axiomas para se falar de um cálculo (sistema formal), nem evidentemente para se falar numa linguagem formal caracteristicamente lógica. Basta pensar, a este último propósito, nos cálculos (sistemas formais) na base de linguagens formais caracteristicamente lógicas que dispensam axiomas e que são hoje dito(a)s de dedução natural – no seguimento do trabalho desenvolvido por Gerhard Gentzen a partir de 1934.[4]

[4] Richard I. G. Hughes ("On First-Order Logic", p. 276) explica a motivação do trabalho de Gentzen do seguinte modo: «...if one is not a logicist and nevertheless believes that one of the aims of formalizing logic is to make explicit the inner workings of mathematical reasoning, then it seems inappropriate to borrow the customary form of that reasoning to do so. Yet this is precisely what the axiomatic approach does; it presents

Em quarto e último lugar, tem de ser especificado um conjunto (finito) de regras que ligam entre si fórmulas bem formadas – mas não necessariamente todas as fórmulas bem formadas – de um cálculo (sistema formal). Estas regras permitem decidir, para uma dada fórmula bem formada, se ela pode ser produzida (ou derivada) a partir de um dado conjunto de fórmulas bem formadas – sendo que este conjunto é não vazio e que lhe podem pertencer um ou mais axiomas (se algum existir): «And finally (primitive) *rules of inference* (or *rules of procedure*) are laid down, rules according to which, from appropriate well-formed formulas as *premisses*, a well-formed formula is *immediately inferred* as *conclusion*.» A terminologia usada por Church pode voltar a suscitar a ideia de que apenas há cálculos (sistemas formais) caracteristicamente lógicos. Um modo de evitar esta ideia consiste em optar por falar de regras de produção (ou derivação) – em vez de regras de inferência. O próprio Church faz, primeiro, uma observação e, depois, uma nota que contribuem para afastar a ideia de que apenas há cálculos (sistemas formais) caracteristicamente lógicos. No âmbito da observação, Church escreve que quando se lida com um cálculo (sistema formal) – ou, mais literalmente e nesta passagem, com um sistema não interpretado – «...the terms *premiss*, *immediately infer*, *conclusion* have only such meaning as is conferred upon them by the rules ... themselves.» Depois, no âmbito da nota – a propósito do carácter imediato da produção (ou derivação) de uma fórmula bem formada a partir de uma ou mais fórmulas bem formadas – Church salienta que não se deve pensar em inferências imediatas no sentido da lógica tradicional, mas sim no facto de que é requerida uma e só uma aplicação de uma determinada regra do conjunto das regras de produção (ou derivação) do cálculo (sistema formal) para produzir (ou derivar) uma outra fórmula bem formada.

Atendendo a tudo isto, torna-se possível fixar mais dois pontos terminológicos. O primeiro ponto respeita aos teoremas – que são aquelas fórmulas bem formadas produzidas (ou derivadas) apenas a partir de um ou mais axiomas (se algum existir). O segundo ponto

logic in Euclidean clothing. To put this another way, if we are interested in the logic underlying Euclid's reasoning, then, instead of providing him with more axioms, we should look at the way he gets from one line to the next [na qual estará então um teorema, de acordo com a terminologia acima referida].» Depois, Hughes faz remontar esta perspectiva exactamente a Gentzen, «...who turned away from the axiomatic approach used by Frege, Russell, and Hilbert. "In contrast," he wrote … , "I intended to set up a formal system which came as close as possible to actual reasoning. The result was a *calculus of natural deduction*" (his emphasis).»

respeita à produção (ou derivação) de uma fórmula bem formada em geral. Neste caso fala-se, precisamente, de uma derivação e diz-se da(s) fórmula(s) bem formada(s) sucessivamente produzida(s) (ou derivada(s)) – pela aplicação de uma e só uma regra de produção (ou derivação) de cada vez – que essa(s) fórmula(s) bem formada(s) é (são) derivável(is).[5] Ora, nesta altura, pode ver-se que tanto o cálculo (sistema formal) de Boole, como o cálculo (sistema formal) de Frege permitem a obtenção de teoremas e a efectivação de derivações – mas que *apenas no caso de Frege* os teoremas e as derivações são integralmente e sempre concebidos como caracteristicamente lógicos.

O passo seguinte consiste em referir a existência de um outro modo de pensar em linguagens formais que é – pelo menos de acordo com Church – menos prático do que o modo de pensar atrás apresentado. Numa nota a uma passagem na qual admite o uso de "base primitiva" (de uma linguagem formal) para falar de um cálculo (sistema formal), Church (*Introduction to Mathematical Logic*, p. 50) começa por escrever: «An alternative, which might be thought to accord better with the everyday use of the word "language", would be to define a

[5] No que respeita à terminologia usada até este momento, Douglas R. Hofstadter (*Gödel, Escher, Bach–Laços Eternos*, pp. 38-39) observa, em primeiro lugar, a diferença existente entre um uso comum e o uso técnico de "teorema": «Tais sequências, produzidas pelas regras [de produção (ou derivação)] chamam-se *teoremas*. O termo «teorema» tem, evidentemente, um uso comum em matemática que difere bastante deste. Significa uma proposição em linguagem comum cuja veracidade é demonstrada por uma argumentação rigorosa, como o teorema de Zenão sobre a «inexistência» do movimento, ou o teorema de Euclides sobre a infinidade de números primos. Mas nos sistemas formais os teoremas não devem necessariamente ser concebidos como proposições – são simplesmente sequências de símbolos. E, ao invés de *demonstrados*, os teoremas são simplesmente *produzidos*, como se fosse por meio de máquinas, de acordo com certas regras tipográficas. Sendo assim, ... «teorema» terá, como é evidente, não só o significado quotidiano – um teorema é uma proposição em linguagem comum que alguém demonstrou –, como também o significado técnico – uma sequência que pode ser produzida nalgum sistema formal.» Em segundo lugar, Hofstadter (*Gödel, Escher, Bach–Laços Eternos*, p. 39) nota que o mesmo acontece com "axioma", «...cujo significado técnico também difere bastante do significado usual». Note-se que este significado técnico pode permitir contrastar os axiomas com os teoremas pelo facto de os axiomas serem fundamentais num cálculo (sistema formal) no sentido técnico de não poderem ser derivados nesse cálculo (sistema formal) de qualquer outra fórmula bem formada. Nesta situação, diz-se então que os axiomas – ou o conjunto dos axiomas – são independentes, tornando-se aceitável a ideia de que os axiomas são fórmulas bem formadas fundamentais (no sentido especificado que dispensa o apelo à evidência, tal como também a referência a qualquer interpretação do cálculo (sistema formal)). Em terceiro lugar, Hofstadter (*Gödel, Escher, Bach–Laços Eternos*, p. 39) observa ainda que «...a derivação [no âmbito de um cálculo (sistema formal)] é um primo austero da demonstração».

"language" as consisting of primitive symbols and a definition of well-formed formula, together with an *interpretation* ... and to take the axioms [se algum existir] and rules of inference as constituing a "logic" for the language [ou, de outro modo, como constituindo a componente dedutiva ou o aparato dedutivo (para a linguagem formal em causa)].» Um pouco depois, surge a posição do próprio Church (*Introduction to Mathematical Logic*, p. 50) sobre este modo de pensar: «But we reject it here, partly because of reluctance to change a terminology already fairly well established, partly because the alternative terminology leads to a twofold division in each of the subjects of syntax and semantics ... – according as they treat of the object language alone or of the object language together with a logic for it – which ... seems unnatural, and of little use so far as can now be seen.»

Em primeiro lugar e no seguimento destas observações de Church, pode-se ver o que é uma interpretação – para qualquer um dos dois modos de pensar. Uma interpretação é uma atribuição de referência ou de significado a símbolos primitivos de um cálculo (sistema formal) ou de uma linguagem formal – consagrando nesta última alternativa, respectivamente, os modos de pensar preferido e dispensado por Church. Nos termos em que Boole formulou o princípio de invariância da forma (MAL, p. 3) e falou da natureza de um verdadeiro cálculo (MAL, p. 4), importa é que uma (ou várias) interpretação(ões) seja(m) consistente(s) em dois sentidos: no sentido em que a(s) interpretação(ões) conserva(m) – invariantes – certas regras (ou leis, na terminologia de Boole) de combinação dos símbolos; no sentido – adicional ao anterior – em que, se isso ocorrer com várias interpretações, então todas estas interpretações são igualmente admissíveis. De acordo com o princípio de invariância da forma, aquilo que Boole chamou um verdadeiro cálculo é, por um lado, sintacticamente isolável – podendo ser encarado como uma combinação sob determinadas regras (ou leis, como diz Boole) dos símbolos primitivos de um cálculo (sistema formal) ou de uma linguagem formal – e, por outro lado, atendendo agora ao que Boole (*George Boole – Selected Manuscripts on Logic and its Philosophy*, p. 148) chamou condições de interpretação, semanticamente interpretável de um modo capaz de assegurar uma referência ou um significado não só para os símbolos primitivos, como para os resultados da aplicação das regras (ou da conservação da verdade das leis, na terminologia de Boole) a que estão sujeitos esses símbolos primitivos.

Em segundo lugar, deve-se notar que a diferença entre os dois

modos de pensar desaparece – ou desaparece exactamente *na prática* – quando se trata de linguagens formais que já são encaradas como caracteristicamente lógicas.[6] O uso dos termos "regras de inferência" ou "regras de produção (ou derivação)", por exemplo, revela-se indiferente – dado estar em causa uma interpretação (de símbolos primitivos) de acordo com a qual uma linguagem formal está apta a lidar com problemas suscitados pela validade de certos raciocínios. De acordo com o primeiro modo de pensar, a linguagem formal *é* uma lógica porque uma interpretação dos símbolos primitivos de um cálculo (sistema formal) – juntamente com os axiomas (se algum existir) e as regras de produção (ou derivação) desse cálculo (sistema formal) – permite resolver problemas suscitados pela validade de certos raciocínios. De acordo com o segundo modo de pensar, algumas linguagens formais *têm* a lógica que é determinada pelos axiomas (se algum existir) e pelas regras de inferência de uma *outra* linguagem formal[7] cujos símbolos primitivos também se encontram, tal como antes, interpretados de modo a lidar com problemas suscitados pela validade de certos raciocínios. O ponto é o de que apenas esta outra linguagem formal – a única que tem então cabimento pensar como caracteristicamente lógica ou, ainda, a única linguagem formal que seria verdadeiramente uma lógica – pode ter sido elaborada à parte daquelas linguagens formais das quais se diz *terem* (mas não propriamente *serem*) uma lógica. Ora, a própria ideia de que uma lógica se podia elaborar à parte da matemática – e apenas de acordo com uma imagem ou um modelo tal como o proporcionado por uma parte da matemática ou pela aritmética – é exactamente aquilo que separa o modo de pensar de Boole do modo de pensar que terá sido concretizado por Frege.

4

Embora se possa pensar num uso irrestrito do termo "validade" para qualquer linguagem formal, o mais interessante é ver o que sucede

[6] Por vezes e no âmbito do segundo modo de pensar, usam-se os termos "sistema formal" e "cálculo (lógico)" apenas para aquilo que, no seguimento da nota acima citada de Church, é a lógica – ou, ainda e de novo, a componente dedutiva ou o aparato dedutivo – de uma linguagem formal.

[7] Mesmo que se prefiram usar os termos "sistema formal" e "cálculo (lógico)" para esta outra linguagem formal. Nesta situação, torna-se evidente que há apenas cálculos (sistemas formais) lógicos – embora não seja ainda possível dizer, tal e qual como acontece no primeiro modo de pensar, que todas as linguagens formais são caracteristicamente lógicas ou, ainda, que todas as linguagens formais têm uma lógica (um componente dedutivo ou um aparato dedutivo).

quando essa linguagem formal é caracteristicamente lógica. Nesta situação, existe a possibilidade de considerar dois conceitos de validade: um deles é sintáctico e o outro é semântico – isto sem tomar posição sobre a eventual prioridade de um ou de outro dos dois conceitos. Seja uma linguagem formal L – ou uma particular lógica L – cujas fórmulas bem formadas F_1 , ... , F_{n-1} , F_n ($n \square 1$) representam as premissas (F_1 , ... , F_{n-1}, com $n \square 1$) e a conclusão (F_n) de um argumento.

Para explicar o conceito sintáctico de validade, tem-se

> (1) F_1 , ... , F_{n-1} , F_n é válido em L se e só se F_n é derivável de F_1 , ... , F_{n-1} e dos axiomas de L (se algum existir) de acordo com uma aplicação sequencial das regras de inferência de L – podendo dizer-se, neste caso, que há uma prova (em L) de F_n e escrever-se
>
> F_1 , ... , $F_{n-1} \vdash_L F_n$
>
> (F_n é uma consequência lógica sintáctica, em L, de F_1 , ... , F_{n-1}).

No caso em que n=1, tem-se

> (1.1) F é válida em L – ou é um teorema de L – se e só se F é derivável dos axiomas de L (se algum existir) de acordo com uma aplicação sequencial das regras de inferência de L – escrevendo-se, neste caso,
>
> $\vdash_L F$
>
> (F é um teorema de L).[8]

Para explicar o conceito semântico de validade, tem-se

> (2) F_1 , ... , F_{n-1} , F_n é válido em L se e só se F_n é verdadeira em todas as interpretações em que F_1 , ... , F_{n-1} são verdadeiras – escrevendo-se, neste caso,

[8] Note-se que pode ainda ser estabelecido – pelo menos por comodidade de escrita – que os axiomas de L (se algum existir) contam como teoremas de L.

$F_1 , ... , F_{n-1} \models_L F_n$
(F_n é uma consequência lógica semântica, em L, de $F_1 , ... , F_{n-1}$).

No caso em que n=1, tem-se

(2.1) F é válida em L – ou é uma verdade lógica de L – se e só se F é verdadeira em todas as interpretações de L – escrevendo-se, neste caso,

$\models_L F$
(F é uma verdade lógica de L).[9]

A ordem da explicação anterior pode ser invertida. Neste caso, os conceitos sintáctico e semântico de validade (para um argumento com premissas $F_1 , ... , F_{n-1}$ e conclusão F_n) são explicados a partir da condicionalização

(1*) $(F_1 \square ... \square F_{n-1}) \square F_n$

e conforme esta condicionalização seja, respectivamente,

(1*.1) um teorema de L, $\vdash_L (F_1 \square ... \square F_{n-1}) \square F_n$,

ou

(1*.2) uma verdade lógica de L, $\models_L (F_1 \square ... \square F_{n-1}) \square F_n$.

Qualquer que seja a ordem de explicação, a teoria da prova e a teoria dos modelos são as áreas no âmbito das quais actualmente ocorrem, respectivamente, a concepção sintáctica e a concepção semântica de validade – sendo que uma prova é uma derivação no sentido (sintáctico) atrás introduzido, mas que um *modelo* de uma ou mais fórmulas bem formadas de uma linguagem formal L é uma interpretação de L na qual essas fórmulas bem formadas são verdadeiras. Certamente que esta noção de interpretação – cuja

[9] Note-se que pode pretender-se que cada axioma de L (se algum existir) é uma verdade lógica – e que, nesse caso, todos os teoremas de L serão também pensados (ou pelo menos pensados, por oposição a deriváveis) como verdades lógicas.

articulação completa se deve essencialmente a Alfred Tarski – é bem mais específica do que aquela que Boole usou, mas ela remete ainda assim e muito claramente para a noção de universo de discurso que Boole integrou na Lógica. Além disso, a teoria da prova e a teoria dos modelos têm de encarar-se precisamente como as duas partes fundamentais da metalógica. Nesta perspectiva, a questão da prioridade de um ou de outro dos dois conceitos de validade – ou, até mesmo, a defesa da ideia de que a noção de consequência lógica é, em si mesma, apenas semântica ou apenas sintáctica – é uma questão filosófica (ou da filosofia da lógica) e não uma questão metalógica (ou da metalógica). O ponto é o de que esta(s) questão(ões) não podiam ser formuladas pelo próprio Frege. A sua formulação estava-lhe vedada quer pela ideia de auto-suficiência universal da sua lógica, inspirada pela ideia leibniziana de haver uma (língua) *characteristica universalis* que fosse simultaneamente um *calculus ratiocinator*, quer pela fixação de Frege na noção – precisamente e apenas – de prova.

Referências

Boole, George – *The Mathematical Analysis of Logic, Being an Essay towards a Calculus of Deductive Reasoning* [MAL]. Cambridge: Macmillan, Barclay & Macmillan; London: G. Bell, 1847. Reimp. [Bristol]: Thoemmes, 1998.

Boole, George – *An Investigation of the Laws of Thought on which are Founded the Mathematical Theories of Logic and Probabilities* [LT]. Cambridge: Macmillan; London: Walton & Maberly, 1854. Reimp. [New York]: Dover, 1973.

Boole, George – *George Boole - Selected Manuscripts on Logic and its Philosophy*, Ed.: Ivor Grattan-Guiness; Gérard Bornet. Basel; Boston; Berlin: Birkhäuser, 1997.

Boole, George – "The Calculus of Logic". *Studies in Logic and Probability*, Ed.: Rush Rhees. London: Watts, 1952, pp. 125-140.

Church, Alonzo – *Introduction to Mathematical Logic*. Princeton, NJ: Princeton University Press, 1956.

Frege, Gottlob – *Begriffsschrift, eine der arithmetischen nachgebildet Formelsprache das reinen Denkens* / "*Begriffsschrift*, a Formula Language, Modelled Upon That of Arithmetic, for Pure Thought" [BS]. *From Frege to Gödel: A Souce Book in Mathematical Logic, 1879-1931*, Ed.: Jean van

Heijenoort. Trad. Jean van Heijenoort. Cambridge, MA: Harvard University Press, 1967, pp. 5-82.

Hofstadter, Douglas R. – *Gödel, Escher, Bach: Laços Eternos*. Trad. José Viegas Filho, A. J. Franco de Oliveira; rev. e coord.: A. J. Franco de Oliveira. Lisboa: Gradiva, 2000.

Hughes, Richard I. G. – "On First-Order Logic". *A Philosophical Companion to First-Order Logic*, Ed.: R. I. G. Hughes. Indianapolis; Cambridge: Hackett, 1993, pp. 259-290.