Disputatio 10, Maio 2001

PSICOLOGIA POPULAR, TEORIA DA DECISÃO E COMPORTAMENTO HUMANO COMUM

António Zilhão
Universidade de Lisboa

1. DIE GEMEINSAME MENSCHLICHE HANDLUNGSWEISE

O §206 das *Investigações Filosóficas* (P.U.) de Wittgenstein termina da seguinte maneira: «O modo humano comum de agir [die gemeinsame menschliche Handlungsweise] constitui o sistema de referência por intermédio do qual interpretamos línguas estranhas». Esta frase é uma formulação sintética da declaração programática em volta da qual o chamado ponto de vista da interpretação se veio posteriormente a desenvolver.

Na realidade, tanto o conceito de «tradução radical», de Quine, como o conceito de «interpretação radical», de Davidson, podem ser vistos como elaborações do ponto de vista expresso por Wittgenstein nesta secção das P.U. Nela, Wittgenstein antecipa uma estratégia de dramatização do problema da construção de uma semântica empírica que os dois filósofos norte-americanos tornaram célebre. Trata-se da estratégia de pedir ao leitor que imagine ser um explorador acabado de chegar a uma terra na qual os nativos falam uma língua completamente estranha e que pense no que então teria que fazer para tentar entender os nativos.

Como todas as declarações programáticas, porém, esta e outras passagens semelhantes das P.U., ao mesmo tempo que estabelecem um corte radical com a tradição anterior e definem uma abordagem completamente nova dos estudos semânticos, são bastante vagas no que diz respeito aos detalhes do modo como essa abordagem poderia ser sistematicamente desenvolvida.

Com efeito, em que consiste «o modo humano comum de agir»? Como poderemos caracterizá-lo de forma independente, i.e., sem ser pela simples indicação extensional do conjunto dos seres humanos acompanhada da caracterização indexical «é o modo de agir *destas* criaturas»? Para além de uma pouco clara menção à ideia de «regularidade», não encontramos nas P.U. os meios para responder a esta questão. Do mesmo modo, quando confrontado com ela, Quine limita-se a apelar para a empatia. Ora, a empatia

que nós sentimos quando nos confrontamos com comportamentos humanos semelhantes aos nossos pode, de facto, ser um auxiliar precioso na realização de trabalho de campo; mas ela nem constitui nem substitui uma caracterização substantiva da especificidade da estrutura desses comportamentos.

Um dos méritos da filosofia de Davidson consiste precisamente no facto de que ela tenta apresentar uma caracterização independente e substantiva do modo humano comum de agir. A estratégia básica seguida por este filósofo no tratamento desta questão consiste em tomar a sério a velha definição aristotélica do Homem como sendo o animal racional e partir daí para uma caracterização independente daquilo em que consistiria a racionalidade de um agente. Uma vez esta caracterização encontrada, então o modo humano comum de agir poderia ser definido de forma não circular como o modo racional de agir, mais coisa menos coisa, e o trabalho de interpretação linguístico e comportamental poderia ser desenvolvido sobre esta base.

Esta estratégia combina portanto dois momentos. Um primeiro momento, de determinação conceptual *a priori* do conceito de racionalidade agencial, e um segundo momento, de carácter empírico, de explicação e previsão comportamental e linguística. O seu sucesso ou insucesso encontra-se claramente dependente da adequação ou inadequação do resultado alcançado no primeiro momento às exigências e constrangimentos que caracterizam o segundo momento.

Ora, é minha convicção que esta estratégia se encontra presentemente num impasse e que esse impasse é causado precisamente pelo facto de ela não conseguir superar a discrepância que existe entre o resultado da determinação *a priori* do conceito de racionalidade agencial e os constrangimentos empíricos impostos pela necessidade da sua aplicação prática à explicação e previsão dos comportamentos humanos. No ensaio que se segue, apresentarei alguns dos argumentos nos quais esta convicção se fundamenta.

2. RACIONALIDADE ARISTOTÉLICA E RACIONALIDADE TEÓRICO-DECISIONAL

Comecemos por tentar determinar em que sentido o predicado «é racional» é usado em contextos agenciais.

«É racional» aplica-se tanto a acções como a agentes. Temos, portanto, que discriminar entre estes dois géneros de atribuições de «é racional». Para o fazer, vou chamar à propriedade que é atribuída a agentes *racionalidade dinâmica*, e à propriedade que é atribuída a acções *racionalidade enérgica.* Deste modo, o problema da determinação dos critérios que regulam o uso adequado de «é racional» em contextos agenciais pode ser, *prima facie*, subdividido em dois subproblemas, nomeadamente, o problema da determi-

nação dos critérios que regulam a atribuição de racionalidade enérgica e o problema da determinação dos critérios que regulam a atribuição de racionalidade dinâmica.

O segundo destes problemas parece, porém, ser subsidiário do primeiro. Isto é, *prima facie*, um agente deverá ser considerado como racional se e somente se as suas acções forem, em geral, acções racionais. Se este é o caso, então a compreensão de o que faz com que uma acção seja racional tem prioridade epistémica. Vamos então concentrar-nos na análise do problema da racionalidade enérgica.

Quando aplicado a uma acção, «é racional» é em geral tomado como uma forma de referir uma relação particular que obtém entre, por um lado, os objectivos e a base informativa subjacentes à acção e, por outro lado, ou a própria acção ou as suas consequências. Os objectivos e a base informativa atrás mencionados são, em geral, considerados como sendo dados pelos desejos e as crenças do agente. Se queremos então clarificar em que consiste o aspecto enérgico da racionalidade temos que determinar o carácter desta relação.

No uso comum, a expressão «é racional» parece pertencer à mesma família de expressões que «é eficaz», «é adequado», etc. Tendo este facto semântico em mente, vejamos então como é que «é racional» tem sido tornado mais preciso na literatura filosófica. Nesta, encontramos basicamente dois modelos de caracterização da relação por meio da qual a ideia de racionalidade enérgica se deixaria capturar. Estes são os seguintes.

Em primeiro lugar, o modelo aristotélico. Este é o modelo de acordo com o qual a racionalidade enérgica deve ser caracterizada em termos do chamado «silogismo prático». Em segundo lugar, o modelo da Teoria da Decisão. Este é o modelo de acordo com o qual esta relação deve ser caracterizada nos termos da teoria matemática da tomada de decisões, ou tal como esta foi apresentada por Ramsey, em 1926, ou em qualquer uma das suas variantes.

Vamos começar por considerar o primeiro destes modelos. Segundo este, considera-se que uma acção A é racional se e somente se ela é o resultado de um processo representável por meio do seguinte género de inferência prática:

O agente S tem um desejo D cujo conteúdo é F;
O agente S tem uma crença C cujo conteúdo é o de que fazer a acção A é o melhor que ele tem a fazer para obter F;

∴ O agente S faz A

Como modelo da acção racional, este parece-me ser um algoritmo inadequado. A minha principal objecção contra ele é a de que, se, intuitivamente, a racionalidade de uma acção é algo aparentado com a sua eficácia

ou a sua adequação, então este algoritmo falha num aspecto essencial. Nomeadamente, o da conexão que obtém entre a crença C que um agente tem acerca de qual é a melhor acção que ele tem a fazer para obter um determinado objectivo F e aquela acção que, dadas as crenças que ele tem acerca do mundo e os outros objectivos que ele também tem, é realmente a melhor.

Para clarificar a questão, imagine-se a situação seguinte. Um viajante chega a Constantinopla (a parte europeia de Istanbul) e tenta descortinar qual é a melhor maneira de chegar a Skutari (a parte asiática de Istanbul); de acordo com o modelo do silogismo prático, se ele engendrar a crença de que o melhor modo de chegar a Skutari é viajando à volta do Mar Negro, então a sua acção será racional (e, portanto, normal) se e somente se o viajante se puser a viajar à volta do Mar Negro com o objectivo de finalmente chegar a Skutari. É todavia intuitivamente óbvio que, na maioria das circunstâncias, ele deveria ter engendrado a crença de que o melhor modo de chegar a Skutari seria empreendendo a travessia do Bósforo.

Assim, ou a sua crença acerca de qual é a melhor forma de agir é justificável em termos das suas crenças acerca do modo (anormal) como o mundo se lhe apresenta naquela ocasião e dos seus outros objectivos, e portanto a racionalidade da acção apenas emergirá quando o seu objectivo e a sua crença acerca de o que é o melhor a fazer para o alcançar forem confrontados com essas outras crenças e objectivos que ele igualmente tem, ou não parece ter muito sentido considerar como racional uma acção baseada numa crença acerca de qual a melhor forma de agir que é simplesmente estúpida. Em ambos os casos, o algoritmo revela-se inadequado, uma vez que não nos fornece quaisquer critérios para avaliar da razoabilidade da crença em função da qual a acção é efectivamente levada a efeito.

O conceito de racionalidade enérgica foi, todavia, consideravelmente refinado pela teoria matemática da tomada de decisão. Neste contexto teórico, uma acção é considerada como racional se e somente se for tal que, independentemente da avaliação do próprio agente, origina as consequências que são as melhores para ele, de acordo com os seus objectivos e a informação que ele tem disponível acerca do mundo. Uma tal definição não nos fornece ainda um critério que nos permita determinar qual é a melhor de duas acções possíveis; é apenas uma definição que determina o que significa dizer-se de uma dada acção que ela é racional. Em particular, trata-se de uma definição que equaciona claramente a ideia de racionalidade com a ideia de optimalidade. Ela precisa por isso de ser associada a um critério que especifique como vamos determinar dentro de um leque de acções possíveis qual é ou quais são aquela acção ou aquelas acções que, em cada situação específica, é a melhor ou são as melhores. Esse critério é-nos também fornecido pela teoria. Trata-se do princípio da maximização da utilidade esperada. A compreensão deste princípio exige todavia alguma elucidação.

No contexto desta teoria, as diferentes acções que o agente tem à sua disposição são pesadas não apenas em termos da sua capacidade de produzir o objectivo pretendido mas também em termos da aceitabilidade ou desiderabilidade dos efeitos colaterais a que ela também pode dar origem. Chama-se então ao objectivo pretendido juntamente com os seus efeitos colaterais o desfecho total de uma acção. As representações dos diferentes desfechos totais possíveis em cada conjunto de circunstâncias são, por sua vez, consideradas como estando organizadas numa escala determinada pela atribuição de valores numéricos aos mesmos, de acordo com a sua utilidade ou desiderabilidade respectiva. Chama-se a esta escala a função de utilidade; a teoria pressupõe por conseguinte que cada agente possui uma função de utilidade bem definida.

Por outro lado, a teoria pressupõe também que cada agente tem um conjunto de crenças acerca do mundo que têm pesos diferentes; estes podem ser representados pela atribuição de probabilidades aos conteúdos proposicionais que referem a obtenção ou não obtenção de qualquer um dos estados do mundo que se pressupõe serem relevantes para a determinação do desfecho total da acção do agente.

Considera-se uma acção como racional se e apenas se ela puder ser representada como resultando da seguinte sequência de procedimentos. Primeiro, a probabilidade da obtenção de cada um dos estados do mundo relevantes é multiplicada pela utilidade de cada um dos desfechos totais possíveis de uma acção. Segundo, os produtos assim obtidos são adicionados; à soma assim obtida chama-se a utilidade esperada de empreender uma dada acção. Terceiro, toma-se uma decisão para agir de um certo modo por meio da selecção daquela acção cuja utilidade esperada é máxima.

Podemos assim considerar que este modelo empresta um conteúdo rigorosamente definido à ideia vaga de acção racional, clarificando-a como aquela acção que implementa a escolha da melhor opção disponível dados os objectivos do agente e o modo como este representa o mundo. Não é por isso de admirar que alguns filósofos tenham defendido a ideia de que o modelo da teoria da decisão deveria substituir o modelo do silogismo prático como o modelo adequado da racionalidade enérgica. O mais conhecido dos proponentes desta ideia é Donald Davidson.

Mas será este modelo na realidade um modelo psicologicamente adequado?

3. TEORIA DA DECISÃO E PSICOLOGIA

Suponhamos que a Teoria da Decisão, tal como caracterizada acima, capta adequadamente o conceito abstracto de racionalidade enérgica. Será que, partindo deste pressuposto, poderemos continuar a aceitar como verdadeira a definição aristotélica do Homem como o animal que age racio-

nalmente? Para responder a esta pergunta temos que tentar determinar se a Teoria da Decisão é, de facto, empiricamente verdadeira acerca dos seres humanos. Como é então possível imaginar-se (se é que é de todo possível imaginar-se) que a Teoria da Decisão poderia ser posta perante o tribunal da experiência?

Prima facie a experiência confirmará a teoria se e somente se as seguintes duas condições se encontrarem satisfeitas. Primeira, uma vez identificada uma dada acção e as circunstâncias nas quais ela ocorre, as crenças e os desejos, e os seus respectivos pesos relativos, que a teoria atribui aos agentes nessas circunstâncias são aqueles que eles efectivamente têm. Segunda, a relação que obtém entre as crenças e os desejos dos agentes e as suas acções é adequadamente capturada pelo princípio da maximização da utilidade esperada; isto é, este princípio constitui efectivamente uma lei de carácter empírico por meio do uso da qual é possível explicar e prever as acções humanas.

Comecemos por examinar a primeira condição. Como se propõe a teoria determinar em cada caso que crenças e que desejos cada agente tem e quais são os seus pesos relativos? Comecemos pelos desejos. De acordo com o quadro conceptual delineado pela teoria, identificar os desejos de um agente e qual a sua importância relativa é determinar a escala de utilidades do agente. Vejamos como Ramsey propunha que se determinasse a escala de utilidades de um agente.

De acordo com ele, a estratégia para alcançar uma tal determinação consiste em encontrar uma condição P tal que dela se possa inferir que o agente atribui uma probabilidade subjectiva de 1/2 tanto à sua ocorrência como à sua não ocorrência. O ponto de partida dessa inferência consistiria, por sua vez, na observação de uma reacção de indiferença manifestada pelo agente quando instado a escolher entre 2 apostas nas quais a obtenção de uma consequência claramente preferida pelo agente a uma outra seria, num caso, feita depender da verificação da condição P e, no outro, feita depender da não verificação da condição P. Esta indiferença deveria, então, ser interpretada como um efeito de um facto psicológico: a atribuição pelo agente da mesma utilidade esperada a cada uma das apostas.

Portanto, de acordo com Ramsey, o comportamento de indiferença do agente seria o resultado de ele saber que nenhuma das opções disponíveis lhe permitiria escolher uma consequência cuja utilidade esperada fosse superior à outra. Seria precisamente este pressuposto que nos autorizaria a representar uma tal indiferença por meio de uma igualdade aritmética relacionando duas expressões que interpretariam cada uma das opções disponíveis como uma soma de produtos, os factores dos quais representariam uma probabilidade subjectiva de que uma certa condição obtivesse e a utilidade subjectiva de uma certa consequência.

Isto quer então dizer que, para Ramsey, o pressuposto de que o agente age de acordo com o princípio da maximização da utilidade esperada se encontra inscrito na interpretação das suas respostas comportamentais desde o início. Se esse não fosse o caso, a indiferença revelada pelo agente quando instado a escolher entre as duas apostas em causa não teria que ser interpretada por meio de uma igualdade relacionando essas duas expressões; mas sem a possibilidade de estabelecer uma tal equação nenhuma atribuição de probabilidade subjectiva 1/2 a qualquer condição poderia ter sido derivada.

Isto significa que um dos princípios básicos da teoria da decisão é ele próprio usado com vista à determinação de qual é a crença do agente numa dada circunstância e de qual é o modo como ele organiza a sua escala de utilidades, isto é, que desejos tem ele e qual é a sua importância relativa. Com efeito, os valores atribuídos às utilidades das consequências e das apostas na escala são, também eles, alcançados por meio da solução de novas equações, as quais interpretam novos comportamentos de indiferença como resultados da impossibilidade de escolher uma utilidade esperada máxima. No contexto da teoria, isso só pode significar que as opções têm que ter as mesmas utilidades esperadas.

Vamos agora considerar como é possível identificar as crenças que o agente tem acerca do mundo e quais são os seus pesos respectivos. Trata-se portanto de identificar as probabilidades subjectivas diferentes de 1/2 que o agente atribui a diferentes estados possíveis do mundo. De acordo com Ramsey, estas são para ser extraídas das utilidades esperadas que ele atribui às diferentes apostas que ocorrem na sua escala de utilidades. No entanto, para se poderem extrair tais probabilidades das utilidades conhecidas das apostas, é necessário ter em consideração que, de acordo com a aplicação do princípio da maximização da utilidade esperada, a obtenção das utilidades destas últimas é para ser representada por meio de uma soma de produtos, representando, respectivamente, as probabilidades desconhecidas das condições e as utilidades conhecidas das consequências que ocorrem na definição da aposta.

Por conseguinte, a um sujeito S serão atribuídos exactamente aqueles graus de crença que são requeridos tanto pelas utilidades subjectivas que ele revelou anteriormente ter atribuído a certas apostas como pela verdade pressuposta da teoria. Por outro lado, as utilidades subjectivas das consequências, i.e., os seus desejos e respectivos graus, são aquelas utilidades subjectivas cuja atribuição ao sujeito é requerida pela verdade pressuposta da teoria dada a probabilidade subjectiva de uma certa condição que lhe foi anteriormente atribuída.

Por outras palavras, a própria possibilidade de identificar as crenças e os desejos do sujeito, i.e., os seus graus de confiança na verificação ou não verificação de determinadas condições e a sua escala de utilidades, depen-

de, uma vez mais, do pressuposto de que as acções são empreendidas em acordo com o princípio central da teoria, a saber, o princípio da maximização da utilidade esperada.

Deste modo, se queremos testar a teoria, o que temos de testar é a correcção ou incorrecção deste princípio. Mas como poderemos fazê-lo? Supõe-se que este princípio regula de determinada forma a relação causal que obtém entre crenças e desejos, a montante, e a acção deles decorrente, a jusante. Mas como acabou de se mostrar, mesmo supondo que temos a capacidade intuitiva de identificar de forma não problemática os eventos que ocorrem a jusante, i.e., as acções, apenas podemos ter acesso ao conteúdo das crenças e desejos do agente, i.e., à identificação dos eventos que ocorrem a montante, se pressupusermos como verdadeiro o próprio princípio da maximização da utilidade esperada. Quer dizer, a situação é tal que só se pode ter acesso a um dos conjuntos de correlatos à custa dos quais o princípio relacional se deixa definir, se se pressupuser como verdadeiro o próprio princípio relacional cuja verdade se quer testar.

Que fazer então? É aqui que uma análise aos fundamentos da teoria se impõe. Ramsey fundamentou a teoria da decisão numa análise daquilo a que poderíamos chamar «o comportamento racional de um apostador». Isto é, ele mostrou que um agente que queira dedicar-se à prática das apostas ou se comporta de acordo com um certo conjunto de princípios (os axiomas da teoria da decisão) ou perderá dinheiro (ou qualquer outro produto que use para apostar) seja o que for que aconteça. A tese de que a teoria da decisão seria psicologicamente verdadeira é, então, basicamente uma tese analógica. Ela consiste na defesa da ideia de que um agente humano se comporta na sua vida normal do mesmo modo que um apostador racional se comporta num jogo de apostas. Mas comportar-se-ão de facto os seres humanos como apostadores racionais?

Como é óbvio, o facto de ser necessário observar os axiomas da teoria para não se ser um perdedor num jogo de apostas não constitui só por si uma prova de que a teoria é empiricamente verdadeira acerca dos seres humanos. Dois casos diferentes podem invalidar a teoria. Por um lado, pode dar-se o caso que o ser humano comum não se comporte como um apostador racional num jogo de apostas, isto é, pode bem dar-se o caso de que um apostador profissional seja capaz de, com relativa facilidade, colocar o ser humano comum em situações nas quais ele perderá aconteça o que acontecer. Por outro lado, mesmo que, por hipótese, o ser humano comum se comporte como um apostador racional num jogo de apostas, ainda fica por demonstrar a tese analógica de que ele se comportará na vida quotidiana como se esta fosse um jogo de apostas. Deste modo, para mostrar que a teoria é empiricamente verdadeira acerca dos seres humanos é necessário mostrar que estes se comportam *realmente* como apostadores racionais nos

diferentes domínios da sua vida, isto é, que o seu comportamento intencional normal concorda *realmente* com os axiomas da teoria.

Esta é a questão crucial, como se pode facilmente constatar pelo facto de que, se for possível responder-lhe afirmativamente, então a verdade do princípio da maximização da utilidade esperada deixa-se derivar por meios puramente matemáticos. Isto é, este último princípio é um teorema da teoria e não um dos seus axiomas.

4. HUMANIDADE E APOSTAS RACIONAIS

Pode conceber-se a avaliação da validade ou invalidade empírica da teoria da decisão como ocorrendo a dois níveis. Num dos níveis tenta-se determinar o valor psicológico actual dos seus axiomas. No outro nível, submete-se a teoria como um todo a um processo de confirmação tentando verificar se a identificação das crenças e dos desejos de um agente de acordo com os procedimentos acima descritos nos permite fazer previsões rigorosas de novos comportamentos. Neste ensaio vou concentrar-me no primeiro destes níveis.

Na literatura relevante têm sido sugeridos diferentes sistemas de axiomas e diferentes versões para os mesmos axiomas. Todavia, dois destes axiomas parecem ser, em quaisquer versões, cruciais para a definição da estrutura comportamental de um apostador racional. São os seguintes.

Primeiro, o axioma a que me referirei como axioma A. Este estabelece que uma determinada relação, nomeadamente, a relação «x é pelo menos tão preferido quanto *y*» (ou, noutra versão, «x é no máximo tão preferido quanto *y*») obtém entre as coordenadas de qualquer par ordenado de desfechos disponíveis e que uma tal relação é uma ordem linear fraca, i.e., que é uma relação binária que satisfaz as seguintes propriedades: transitividade, reflexividade e também conectividade. É este axioma que permite que as utilidades sejam colocadas numa correlação um-um com os números reais, isto é, que permite que aquelas sejam medidas por estes de um modo tal que tanto os seus lugares na escala de utilidades como as suas diferenças intrínsecas de valor sejam adequadamente expressas pela sua representação por valores numéricos.

Segundo, o axioma a que me referirei como axioma B. Apresentado informalmente, este estabelece que se uma opção A é pelo menos tão preferida quanto uma opção B então, se as opções C e D resultarem das opções A e B, respectivamente, por meio de uma mudança comum em desfechos que são comuns às duas opções, então a opção C é pelo menos tão preferida quanto a opção D.

Discutirei primeiro o axioma B. Alguns testes psicológicos foram já realizados com o objectivo de o pôr à prova. O mais conhecido destes é o chamado Problema de Allais. Este consiste no seguinte. Um conjunto de

sujeitos é confrontado primeiro com a necessidade de escolher entre as seguintes opções. Opção A: uma aposta na qual o sujeito ganha 1.000.000$ aconteça o que acontecer; opção B: uma aposta na qual o sujeito tem uma probabilidade 0,89 de ganhar 1.000.000$, uma probabilidade 0,10 de ganhar 5.000.000$ e uma probabilidade 0,01 de nada ganhar. Em seguida, o mesmo conjunto de sujeitos é confrontado com as seguintes opções. Opção C: uma aposta na qual o sujeito tem uma probabilidade 0,11 de ganhar 1.000.000$ e uma probabilidade 0,89 de nada ganhar; opção D: uma aposta na qual o sujeito tem uma probabilidade 0,10 de ganhar 5.000.000$ e uma probabilidade 0,90 de nada ganhar.

Os resultados foram os seguintes. A maioria dos sujeitos escolheu, primeiro, a opção A sobre a opção B e, segundo, a opção D sobre a opção C. Ora, pressupondo que a utilidade de ganhar 0$ é 0, este padrão de escolha pode ser descrito por meio da combinação das seguintes inequações:

i) 0,11U(1.000.000$)>0,10U(5.000.000$)

e

ii) 0,10U(5.000.000$)>0,11U(1.000.000$).

Esta combinação de inequações constitui porém uma violação patente do axioma B, uma vez que as opções C e D resultam das opções A e B pela introdução da mesma mudança a desfechos comuns a ambas. Logo, de acordo com o axioma, se a opção A é preferida à opção B, então a opção C tem que ser preferida à opção D.

A discussão clássica destes resultados foi travada entre Allais e Savage. Allais defendeu que estes resultados teriam um valor não apenas psicológico-descritivo mas também normativo. De acordo com ele, seria mais racional agir como a maioria dos agentes se mostram tentados a agir do que agir de acordo com o axioma B. Deste modo, seria mais racional aquele decisor que, colocado perante um problema com esta estrutura, violasse a teoria da decisão racional do que aquele decisor que agisse de acordo com ela. Savage defendeu o valor normativo da teoria argumentando que apesar de a maioria dos agentes agir de acordo com os resultados obtidos nas experiências, era possível mostrar-lhes que eles tinham feito a escolha errada e que, uma vez que os agentes tivessem compreendido onde se encontrava o seu erro, eles aceitariam ser corrigidos de acordo com a prescrição oferecida pela teoria.

Ambos os autores concordaram, portanto, que, em casos como o do problema de Allais, a teoria não era respeitada pela maioria dos agentes (incluindo eles próprios); por conseguinte, a discussão travou-se sobretudo

em torno da questão de decidir se o axioma B teria ou não um valor normativo. Nesta discussão, cada um dos autores se baseou nas suas próprias intuições de normatividade, as quais todavia pareciam não ser coincidentes. Seja como for, o meu interesse presente não reside na discussão de qual o valor normativo que a teoria da decisão poderá ter na orientação das decisões humanas, mas sim na discussão sobre se os princípios que ela codifica poderão desempenhar um papel explicativo em Psicologia.

Ora, deste ponto de vista, o facto de que não apenas a maioria dos sujeitos testados, mas até o próprio Savage, agiram espontaneamente contra o modo de agir determinado pela teoria parece constituir evidência empírica bastante forte a favor da contenção de que a teoria não descreve de modo apropriado o modo como nós *de facto* normalmente agimos num conjunto não negligenciável de circunstâncias. Isto é, a ideia de que a racionalidade agencial, enquanto característica definitória do nosso modo comum de agir, possa ser adequadamente descrita por meio da formalização do comportamento de um apostador racional parece assim encontrar-se claramente posta em causa pelos resultados dos testes feitos a partir do Problema de Allais. Este parece aliás ter sido o entendimento generalizado acerca do desfecho da discussão entre Allais e Savage.

Uma nota aparentemente dissonante neste consenso foi porém introduzida por Tversky. Num comentário a esta discussão, Tversky manifesta o seu acordo com cada um dos contendores nos pontos em que eles discordam um com o outro e o seu desacordo com ambos nos pontos em que eles estão de acordo. Este pouco ortodoxo comentário merece um escrutínio atento.

Aquilo de que Tversky discorda e com que ambos os contendores concordam é a ideia de que os dados da investigação psicológica configurariam óbvias violações do axioma B. Esta discordância é fundamentada do seguinte modo: estes dados apenas podem ser interpretados dessa maneira na base do pressuposto de que as consequências sob consideração devem ser inteiramente caracterizadas pelos seus valores monetários; todavia, de acordo com Tversky, nada nos obriga a proceder desse modo. Com efeito, se se incluirem considerações não monetárias na interpretação das consequências, pode bem dar-se o caso de que a teoria não esteja a ser violada. É até possível, contra as contenções de Allais, que os resultados psicológicos obtidos nos testes feitos com base no problema que ele definiu sejam tanto normativa como descritivamente compatíveis com a teoria. Por outro lado, se se contender, como o faz Savage, que a interpretação das consequências não deve incluir a consideração de quaisquer consequências não monetárias, então, embora seja de facto o caso que o comportamento espontâneo das pessoas viola a teoria, é argumentável que a teoria tem um valor normativo. Nomeadamente, pode explicar-se às pessoas que o seu comportamento se encontra em contradição com a teoria porque elas incluiram implicita-

mente considerações não monetárias na sua tomada de decisão, coisa que não deveriam ter feito. Se elas aceitarem este preceito, e a aceitação ou não aceitação do mesmo depende basicamente dos valores por referência aos quais as pessoas pretendem orientar a sua conduta e não de quaisquer considerações de carácter teórico, então elas devem, de facto, aceitar o convite de Savage para que revejam a sua decisão no sentido de a pôr de acordo com o axioma B.

Deste modo, e ainda de acordo com Tversky, esta discussão torna claro que quaisquer contenções acerca de observância ou violação dos axiomas só pode ser feita em associação com interpretações particulares dos dados. Desta contenção, Tversky deriva as seguintes duas outras contenções:

> A questão crucial, por conseguinte, não é a da adequação dos axiomas, mas antes a do carácter apropriado ou inapropriado da interpretação das consequências

e

> Na ausência de quaisquer constrangimentos, as consequências podem ser sempre interpretadas de um modo tal que satisfaçam os axiomas

Se nos reportarmos agora ao que foi já dito acerca do modo como a Teoria da Decisão poderia ser empiricamente testada, estas citações de Tversky podem ser lidas como uma afirmação de que o pressuposto de que as nossas identificações intuitivas das acções não são problemáticas é um pressuposto errado. Identificar os dados comportamentais disponíveis com o desempenho de determinadas acções constitui desde logo uma operação de interpretação, operação essa que necessita ainda de ser melhor compreendida.

Apoiando-se nestas contenções de Tversky, que ele cita *verbatim* num artigo crucial acerca deste tema («Hempel on Explaining Action»), Davidson introduz então uma nova forma de encarar o estatuto teórico da Teoria da Decisão. Segundo ele, a determinação do carácter apropriado ou inapropriado do modo como as consequências são, em geral, interpretadas não pode ser feita independentemente do enquadramento do trabalho interpretativo num contexto teórico específico. E não existiria qualquer outro contexto teórico para enquadrar esse trabalho interpretativo para além do que é fornecido pela Teoria da Decisão ela própria. Deste modo, qualquer tentativa para testar experimentalmente os seus axiomas estaria votada ao fracasso, uma vez que os dados experimentais obtidos nos testes necessitariam de ser interpretados e a interpretação dos mesmos teria que ser feita por intermédio do recurso implícito àqueles mesmos axiomas que estariam supostamente a ser o objecto do teste experimental. A identificação dos eventos ocorrendo a jusante da tomada de decisão encontrar-se-ia assim tão dependente de um

qualquer enquadramento teórico prévio como a identificação dos eventos ocorrendo a montante da tomada de decisão.

Usando uma outra terminologia, a posição introduzida por Davidson pode ser caracterizada por meio da ideia de que os axiomas da teoria da decisão seriam verdades sintéticas a priori acerca do comportamento intencional de quaisquer criaturas racionais. Isto é, eles seriam constitutivos da própria ideia de racionalidade enérgica. Por outro lado, nós não seríamos livres de deixar de nos considerar como seres racionais, uma vez que todo o nosso esquema conceptual, tal como ele se encontraria posto em evidência na chamada «psicologia popular», se basearia nesse pressuposto. Assim sendo, a teoria seria necessariamente verdadeira a respeito das nossas acções, isto é, a verdade dos seus axiomas não seria falsificável por qualquer experiência psicológica.

A declaração programática do próprio Davidson a este respeito, tal como aparece em «Hempel on Explaining Action» apenas algumas linhas abaixo das citações de Tversky acima apresentadas, é a seguinte:

> A este respeito, a teoria da decisão é como a teoria da medida para o comprimento ou a massa, ou a teoria da verdade de Tarski. A teoria é em cada caso tão poderosa e simples, e tão constitutiva de conceitos pressupostos por outras teorias satisfatórias (físicas, linguísticas), que temos que nos esforçar por adaptar as nossas descobertas, ou as nossas interpretações, de um modo tal que elas preservem a teoria.

Estou convencido de que as observações de Tversky são extremamente pertinentes. Mas será que, para além de nos alertarem para o facto de que os juízos acerca da validade psicológica dos axiomas estão dependentes da interpretação que dermos às consequências em termos das quais as acções são definidas, elas sustentam o ponto de vista expresso por Davidson? Não penso que esse seja o caso. Do facto de se constatar que, na ausência de quaisquer outros constrangimentos, as consequências podem ser sempre interpretadas de um modo tal que satisfaçam os axiomas não se segue que não existam, em geral, outros constrangimentos interpretativos e que, por conseguinte, nós devamos sempre esforçar-nos por interpretar as consequências de um modo tal que a teoria seja preservada, como defende Davidson.

Na realidade, os comentários de Tversky deixam em aberto três caminhos: o primeiro é o daqueles que contendem que existem, em geral, critérios independentes para interpretar apropriadamente as consequências e que o uso de tais critérios nos permite testar positivamente os axiomas; o segundo é o daqueles que, aceitando que esses critérios existem, contendem que o seu uso nos mostrará que os axiomas são, com frequência, inaplicáveis na descrição do comportamento humano comum; o terceiro é o

daqueles que contendem que não existe qualquer conjunto de critérios externos com mais força interpretativa que a dos próprios axiomas da teoria e que, por conseguinte, as consequências devem sempre ser interpretadas de tal modo que satisfaçam os axiomas. O primeiro caminho parece ser aquele que Papineau tem em mente; o terceiro é o defendido por Davidson. Eu gostaria de defender aqui o segundo. Argumentarei a seu favor na próxima secção, na qual tentarei mostrar que uma outra experiência de carácter psicológico, projectada para testar o axioma a que me referi acima como axioma A e levada a efeito e descrita pelo próprio Tversky, nos dá boas razões para escolher este caminho e evitar qualquer um dos outros.

5. TEMPO, MUDANÇAS DE IDEIAS, ALTERNATIVAS MULTIDIMENSIONAIS E TRANSITIVIDADE

No início da secção anterior referi-me ao axioma A como estabelecendo que a relação de preferência que obteria entre as coordenadas de qualquer par ordenado de desfechos possíveis teria que ser uma ordem linear fraca. Apesar de este ser com efeito o modo como a relação em questão é habitualmente caracterizada em tratamentos axiomáticos da teoria da decisão, é também possível concebê-la como uma ordem linear forte, isto é, como uma relação transitiva, irreflexiva, assimétrica e conectada. Este seria o caso se, em vez de «é pelo menos tão preferido quanto» ou «é no máximo tão preferido quanto», a relação em questão fosse pensada simplesmente como «é preferido a».

Seja como for, o que é essencial é que a relação de preferência seja pensada como transitiva. Se a condição da transitividade não for satisfeita por uma dada relação R definida num conjunto de utilidades, então essa relação não pode ser uma ordem; e se R não for uma ordem, então nenhuma correlação um-um pode obter entre as utilidades a respeito das quais R obtém e os números reais, de tal modo que os últimos possam ser usados para representar os lugares na escala e as diferenças intrínsecas de valor das primeiras.

Ora bem, a experiência feita por Tversky parece mostrar que há circunstâncias nas quais a melhor forma de interpretar o modo como os agentes se comportam é atribuindo-lhes conjuntos intransitivos de preferências. Se este é efectivamente o resultado de tal experiência, então a adequação psicológica da teoria da decisão é realmente posta em causa. Com efeito, a existência de falhas de transitividade acarreta uma de duas consequências: ou a consequência de que os sujeitos testados não são apostadores racionais, no sentido em que a adopção de um padrão intransitivo de preferências os coloca numa posição na qual um apostador profissional pode fazê-los perder aconteça o que acontecer; ou a consequência de que todo o processo de seleccionar aquela acção cuja utilidade esperada seria a maior se tornaria

sem sentido, visto que as escalas de utilidades realmente existentes não seriam representáveis por meio de escalas ordinais de utilidades. Nenhuma destas consequências é compatível com a manutenção da tese de que a teoria da decisão descreveria adequadamente os nossos padrões espontâneos de comportamento.

Ora, apesar de indubitavelmente conhecer os resultados experimentais de Tversky, Davidson não deixa de fazer as seguintes contenções:

> Não penso que possamos dizer claramente o que é que nos poderia convencer de que um homem preferiu num certo momento (e sem ter mudado de ideias) a a b, b a c e c a a. A razão subjacente à nossa dificuldade é que não podemos dar um sentido a uma atribuição de preferência a não ser contra um pano de fundo de atitudes coerentes.

e

> Se o comprimento não é transitivo, qual o significado de usar de todo um número para medir o comprimento? Poderíamos encontrar ou inventar uma resposta a esta pergunta, mas a menos que, ou até que, o façamos temos que nos esforçar por interpretar «maior do que» de tal forma que seja transitiva. O mesmo se passa com «preferido a».

Ora bem, se Davidson conhecia os resultados de Tversky e persistiu em manter estas contenções, então é porque, de acordo com o seu ponto de vista, as atribuições de padrões intransitivos de preferências feitas por Tversky aos sujeitos dos seus testes experimentais deveriam ser substituídas ou por diagnósticos de mudanças de ideias ou por reinterpretações das consequências e, por conseguinte, do conteúdo das acções dos agentes, desenvolvidas de tal modo que o axioma A fosse satisfeito. É minha opinião que nem uma nem a outra destas propostas é credível e que, por conseguinte, as experiências de Tversky revelam que o melhor modo de interpretar certos conjuntos de escolhas humanas é aceitando que existem nos seres humanos, com alguma frequência, conjuntos intransitivos de preferências. Se este é o caso, então tanto Papineau como Davidson estão equivocados acerca de qual o papel que a teoria ramseyana da decisão poderá desempenhar na explicação psicológica.

Tversky realizou diferentes conjuntos de testes empíricos com os seus sujeitos experimentais. Para melhor compreender a ideia subjacente à estrutura desses testes, talvez seja necessário mencionar primeiro brevemente o célebre paradoxo da votação de Condorcet.

Neste paradoxo, Condorcet mostra como é possível que escolhas individuais consistentes originem inconsistências sociais. Em particular, ele mostra como o sistema de voto maioritário gera um padrão intransitivo de preferên-

cias sempre que um conjunto de três indivíduos x, y e z consideram um conjunto de três opções A, B e C de um modo tal que x ordena as suas opções de acordo com a classificação ABC, y ordena as suas opções de acordo com a classificação BCA e z ordena as suas opções de acordo com a classificação CAB e as opções são sujeitas a uma sequência de três votações em alternativa.

Nos seus testes, Tversky confronta os sujeitos individuais com problemas de escolha nos quais as opções são consideradas sob mais do que uma dimensão e a sua ordenação sob uma dimensão entra em conflito com a sua ordenação sob as outras dimensões. No caso do paradoxo de Condorcet, a reunião pelo voto maioritário alternativo das escolhas individuais consistentes resultava numa escolha social inconsistente; no caso das experiências psicológicas de Tversky, a escolha inconsistente é individual e resulta da reunião numa única escolha multidimensional das preferências unidimensionalmente consistentes de um mesmo indivíduo.

O conjunto dos testes de Tversky pode ser resumido por meio do seguinte *Gedankenexperiment*. Suponha o leitor que tem que tomar uma decisão acerca da atribuição de uma bolsa de estudo e que tem perante si diferentes candidaturas. Suponha ainda que há duas características que, de acordo com os regulamentos do concurso, são determinantes para a tomada de decisão, nomeadamente, a inteligência e o *curriculum* académico. Finalmente, suponha também que considera que, destas duas características, a inteligência é a mais importante e que, juntamente com o formulário de candidatura, cada candidato entregou também o seu curriculum académico e os resultados que obteve num teste de inteligência. Ora bem, sabendo que os testes de inteligência não são tão rigorosos quanto seria desejável, imagine que formula a seguinte regra de decisão: se um candidato x for mais inteligente que um candidato y, então o candidato x deve obter a bolsa; considera-se um candidato x mais inteligente que um candidato y se e somente se o seu resultado no teste de inteligência for superior ao de y em pelo menos 5 pontos; se a diferença nos resultados dos testes for inferior a 5 pontos, então x e y serão considerados igualmente inteligentes e a bolsa será atribuída àquele que tiver um melhor curriculum académico.

Esta regra de decisão parece ser bastante sensata. Toma em consideração as duas dimensões consideradas relevantes e combina-as na que parece ser uma maneira razoável. Consideremos agora o seguinte caso. Entre os candidatos existe um conjunto de três A, B e C cujos processos estão relacionados entre si do seguinte modo. C teve mais 4 pontos que B no teste de inteligência e, por sua vez, B teve mais 4 pontos que A no teste de inteligência. Mas acontece que A tem o melhor curriculum académico dos três, B o segundo melhor e C o pior. Ora bem, da conjunção destes dados com a regra de selecção atrás definida segue-se que o leitor deveria preferir A a B, B a C e C a A. Por conseguinte, o seu padrão de escolhas, resultante

de uma regra aparentemente bastante sensata, teria sido intransitivo. Ora, parece-me claro que este resultado não era óbvio desde o início. Que esta evidência intuitiva corresponde na realidade aos factos empíricos é o que é mostrado nos resultados experimentais de Tversky.

Vamos agora tentar determinar se uma estratégia interpretativa desenvolvida de acordo com o ponto de vista Davidsoniano é credível. Para tornar as coisas mais vívidas, ponhamos o problema da seguinte forma. Suponha o leitor que detecta alguém tentando aparentemente implementar esta regra e que já se apercebeu que, de acordo com essa interpretação do comportamento do personagem alvo de observação, o seguimento dessa regra de decisão conduz, sob certas circunstâncias, à geração de conjuntos intransitivos de preferências.

Que faria então? Reinterpretaria o padrão de escolhas desta pessoa de modo a torná-lo compatível com a condição da transitividade e informá-lo-ia em seguida que ele não se encontraria a seguir a regra que pensava estar a seguir? Consideraria que a pessoa em questão tinha começado por escolher de acordo com a regra «curriculum académico primeiro» e em seguida, no meio do processo de selecção, teria mudado o seu padrão de selecção para uma regra de «inteligência primeiro», novamente contra o seu próprio relato do seu procedimento e também contra aquela que parece ser uma das regras mais básicas da selecção de candidatos, nomeadamente, a meta-regra de que os critérios de selecção não devem ser alterados a meio do processo? Ou teria aceite o relato dessa pessoa como adequado aos factos e tê-la-ia tentado convencer de que a regra de escolha que ela tinha escolhido era uma regra intransitiva de escolha e, por conseguinte, era *injusta*?

A última destas três opções parece ser a única sensata. De facto, na ausência de evidência de que a pessoa em questão teria um carácter inconstante teria sido bastante objectável atribuir-lhe um comportamento volúvel. Por outro lado, deve fazer-se notar que a correcção das injustiças potencialmente originadas pelo seguimento dessa regra pelo decisor teria sido bloqueada por uma reinterpretação do seu comportamento de escolha de um modo tal que tornasse a regra *ex post facto* consistente. Um tal bloqueio seria presumivelmente difícil de aceitar por aqueles que tivessem sido vítimas de uma eliminação injusta.

Pretendo portanto argumentar que o facto de a reacção correctiva ser aquela que a maioria de nós se sentiria inclinada a ter constitui evidência a favor da tese de acordo com a qual o comportamento de selecção acima descrito deveria ter sido interpretado como um comportamento de seguir uma regra de decisão que, nas circunstâncias apropriadas, teria gerado padrões intransitivos de preferências.

A esta sugestão pode objectar-se que a mudança de ideias no meio de um processo de selecção é também um comportamento que gera injustiças potenciais e, por conseguinte, teria também despoletado uma reacção

correctiva por parte de um observador agindo de boa fé ou por parte das próprias vítimas do procedimento. Deste modo, a reacção de correcção só por si não discrimina entre duas das possibilidades interpretativas acima mencionadas.

Contra esta objecção gostaria de responder que, apesar de o seu conteúdo ser legítimo, ela baseia-se numa atitude interpretativa muito pouco caridosa; o seu levantamento por parte de um defensor do ponto de vista Davidsoniano teria assim um efeito auto-destrutivo. Deste modo, a acusação de que um tal ponto de vista tem uma tendência para gerar epiciclos interpretativos parece aqui ser perfeitamente adequada.

Com efeito, a explicação por meio da intransitividade parece ser uma explicação muito mais económica e abrangente para dar conta desta situação do que qualquer das suas rivais. Estas ou se baseiam em interpretações com um elevado grau de artificialismo ou se baseiam em atitudes interpretativas muito pouco caridosas. De facto, parece fazer todo o sentido considerar que em situações tendo este género de estrutura complexa a estratégia cognitiva operante é normativamente incorrecta e por conseguinte geradora de intransitividades. Esta estratégia cognitiva faz sentido porque, no exemplo que considerámos, há mais do que uma dimensão relevante em jogo. E agir do modo pelo qual as pessoas aparentemente agem e, por conseguinte, gerar intransitividades, é claramente mais fácil do que manusear algoritmos que integrem as diferentes dimensões em jogo num cálculo normativamente adequado.

Na realidade, os testes psicológicos como aqueles que foram resumidos por meio da apresentação deste *Gedankenexperiment* mostram precisamente duas coisas: que as pessoas tendem a restringir ao mínimo o número de alternativas pertencentes a dimensões diferentes que estão dispostas a comparar simultaneamente num problema de decisão; e que em seguida adoptam rotinas simples para escolher uma de duas opções de cada vez. Embora não gerem necessariamente contradições com o resultado que se obteria com o uso de um algoritmo impecável, estes procedimentos são claramente subóptimos. A um prazo mais longo ou mais curto, dependendo das circunstâncias, eles não podem deixar de gerar inconsistências. Todavia, estes métodos simples de processamento têm a enorme vantagem prática sobre os seus rivais consistentes de serem facilmente manuseáveis quando é necessário lidar com a multidimensionalidade. E na vida real, os problemas são, com assinalável frequência, multidimensionais.

Deste modo, parece-me que uma estratégia Davidsoniana dirigida para a reinterpretação de conjuntos aparentemente intransitivos de preferências de tal modo que estes possam ser tornados *ex post facto* compatíveis com o axioma A contradiz a intuição e o senso comum e origina a produção de epiciclos interpretativos. Por outro lado, se se pressupuser que existem constrangimentos do senso comum exteriores à própria teoria da decisão

que nos permitem identificar consequências e acções e que as preferências por eles reveladas são minimamente constantes ao longo do tempo, então tem que se aceitar que o género de testes psicológicos efectuados por Tversky põe decisivamente em causa a ideia de que o axioma A é, sem mais, empiricamente verdadeiro acerca do comportamento comum dos seres humanos.

6. CONCLUSÃO

Dos argumentos e considerações apresentados neste ensaio penso que é possível extrair a seguinte conclusão disjuntiva. Se o modo mais adequado de caracterizar de forma independente a racionalidade de uma acção em determinadas circunstâncias é pelo estabelecimento de uma analogia com o comportamento que um apostador racional, tal como definido pelos axiomas da Teoria da Decisão, teria nessas circunstâncias, então não parece ser possível manter de pé a velha definição aristotélica do Homem. Neste caso, teremos que nos resignar a incluir as tentativas realistas de caracterização de muitos dos modos humanos de agir que fazem sentido no domínio dos comportamentos irracionais. Mas, se quisermos ser conservadores quanto à definição aristotélica, o que é perfeitamente legítimo, então não podemos usar para toda e qualquer circunstância o comportamento de um apostador racional nessa circunstância como o modelo independente por comparação com o qual estabelecemos o carácter racional de uma acção.

Seja qual for a opção que se escolha, parece-me todavia ser claramente o caso que, ao contrário do que Davidson sustenta, a psicologia popular, tomada no sentido daquele conjunto de procedimentos interpretativos por meio do uso implícito dos quais somos capazes de nos entender uns aos outros, não é uma forma vaga e ainda pouco precisa de representar o nosso comportamento como o comportamento de apostadores racionais. Se esse fosse o caso, não seria possível apelar com plausibilidade para o senso comum para fundamentar a correcção da interpretação de certos comportamentos humanos como constituindo exemplos de violações dos axiomas da teoria. Ora, se isso é possível, como espero tê-lo demonstrado, então esses axiomas não podem ser cristalizações de verdades sintéticas *a priori* acerca da Psicologia humana.

António Zilhão
Departamento de Filosofia
Faculdade de Letras da Universidade de Lisboa
Alameda da Universidade, 1600-214 Lisboa
antonio.zilhao@mail.doc.fl.ul.pt

Referências

Allais, M. 1953: “Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l’ecole americaine” in *Econometrica,* 21, 503-46.

Aristóteles, *Nichomachean Ethics* in *The Basic Works of Aristotle*, ed. by McKeon, New York: Random House, 1941, 927-1112.

Churchland, P. 1970: “The Logical Character of Action- Explanations” in *The Philosophical Review*, 79, 214-36.

Condorcet, “Essay sur l’application de l’analyse a la probabilité des décisions rendues à la pluralité des voix - Discours préliminaire” in his *Sur les Elections et autres Textes 1782-94.* Paris, Fayard, 1986.

Davidson, D. 1970: “Mental Events” in his *Essays on Actions and Events*. Oxford: Clarendon Press 1980, 207-28.

—— 1974a: “Psychology as Philosophy” in his *Essays on Actions and Events*, 229-44.

—— 1974b: “On the Very Idea of a Conceptual Scheme” in his *Inquiries into Truth & Interpretation.* Oxford: Clarendon Press, 1984, 183-98.

—— 1976: “Hempel on Explaining Action” in his *Essays on Actions and Events,* 261-76.

—— 1995: “Could There Be a Science of Rationality” in *International Journal of Philosophical Studies*, 3, 1-16.

Jeffrey, R.C. 1983: *The Logic of Decision* (2nd rev. ed.). Chicago: Chicago University Press.

Kahneman, D. and Tversky A. 1982: "The Psychology of Preferences" in *Scientific American* 246, 160- 73.

Machina, M. 1983: "Generalized Expected Utility Analysis and the nature of the observed violations of the Independence Axiom" in Stigum and Wenstøp (eds.), *Foundations of Utility and Risk Theory with Applications*. Dordrecht: Reidel, 263-93.

McClennen, E. 1983: "Sure-Thing Doubts" in Stigum and Wenstøp (eds.), *Foundations of Utility and Risk Theory with Applications*, 117-36.

Papineau, D. 1978: *For Science in the Social Sciences.* London: The MacMillan Press.

—— 1993: *Philosophical Naturalism*. Oxford: Blackwell.

Quine, W.v.O. 1960: *Word and Object.* Cambridge (MA): The MIT Press.

Ramsey, F.P. 1926: "Truth and Probability" in his *The Foundations of Mathematics and other Logical Essays*, ed. in 1931 by R.B. Braithwaite. London: Routledge & Kegan Paul, 156-198.

Savage, L.J. 1954: *The Foundations of Statistics*. New York: John Wiley & Sons.

Tversky, A. 1969: "Intransitivity of Preferences" in *Psychological Review* 76, 31-48.

—— 1975: "A Critique of Expected Utility Theory: Descriptive and Normative Considerations" in *Erkenntnis* 9, 163-74.

Tversky, A. and Kahneman, D. 1988: "Rational Choice and the Framing of Decisions" in Bell, Raiffa and Tversky (eds.) *Decision Making - Descriptive, Normative, and Prescriptive Interactions*. Cambridge: Cambridge University Press, 167-92.