

Elson Luiz de Almeida Pimentel

**O CONCEITO DE RACIONALIDADE
E OS PARADOXOS DA TEORIA DA DECISÃO –
A PROPOSTA DE ROBERT NOZICK
PARA O DILEMA DO PRISIONEIRO**

Belo Horizonte

Faculdade de Filosofia e Ciências Humanas da UFMG

2005

Elson Luiz de Almeida Pimentel

**O CONCEITO DE RACIONALIDADE
E OS PARADOXOS DA TEORIA DA DECISÃO –
A PROPOSTA DE ROBERT NOZICK
PARA O DILEMA DO PRISIONEIRO**

Dissertação apresentada ao Curso de Mestrado da Faculdade de Filosofia e Ciências Humanas, da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Mestre em Filosofia.

Linha de pesquisa: Lógica e filosofia da ciência.

Orientador: Prof. Ernesto Perini-Santos,
Departamento de Filosofia.

Co-orientador: Prof. Bruno Pinheiro Wanderley Reis,
Departamento de Ciências Políticas.

Belo Horizonte
Faculdade de Filosofia e Ciências Humanas da UFMG
2005

100 Pimentel, Elson Luiz de Almeida
P644c O conceito de racionalidade e os paradoxos da teoria da decisão
2005 – a proposta de Robert Nozick para o dilema do prisioneiro / Elson
Luiz de Almeida Pimentel – 2005

177 f.

Orientador: Ernesto Perini Frizzera da Mota Santos

Co-orientador: Bruno Pinheiro Wanderley Reis

Dissertação (mestrado) – Universidade Federal de Minas Gerais,
Departamento de Filosofia

1. Nozick, Robert. 2. Filosofia americana – Teses.
3. Filosofia – Teses. 4. Ciência – Filosofia – Teses. 5. Racionalidade
– Teses. 6. Processo decisório – Teses. I. Santos, Ernesto Perini
Frizzera da Mota. II. Reis, Bruno Pinheiro Wanderley. III. Universi-
dade Federal de Minas Gerais. Departamento de Filosofia. IV. Títu-
lo

Dissertação defendida e aprovada com a nota 95 (noventa e cinco) pela Banca Examinadora constituída pelos professores:

Prof. Dr. Ernesto Perini Frizzera da Mota Santos (Orientador) – UFMG

Prof. Dr. Paulo Roberto Margutti Pinto – UFMG

Prof. Dr. Edgar da Rocha Marques – UERJ

Pós-Graduação em Filosofia da Faculdade de Filosofia e Ciências Humanas
Universidade Federal de Minas Gerais

Belo Horizonte, 22 de junho de 2005.

Agradeço a todas as pessoas que me ajudaram neste caminho,
em especial à minha família – Bete e Luiza,
aos meus professores do Departamento de Filosofia da UFMG, e
aos meus orientadores:

Prof. Edgar da Rocha Marques – UFRJ, pré-orientador;
Prof. Bruno Pinheiro Wanderley Reis – UFMG, co-orientador;
Prof. Ernesto Perini-Santos – UFMG, orientador.

SUMÁRIO

RESUMO / ABSTRACT	ii
INTRODUÇÃO	iv
1 UMA DISCUSSÃO CONTEMPORÂNEA DO CONCEITO DE RACIONALIDADE	01
1.1 Ramsey e a racionalidade como consistência.....	03
1.2 Sen e a crítica da maximização do auto-interesse.....	16
1.3 Blackburn e o domínio das paixões.....	31
1.4 É possível conciliar a discussão?	43
2 DOS PARADOXOS À TEORIA DOS JOGOS	45
2.1 Paradoxos e dilemas.....	46
2.2 Uma introdução à história da teoria dos jogos.....	51
2.3 Classificação e representação dos jogos.....	57
2.4 Um recorte na teoria para examinar o dilema do prisioneiro.....	66
2.5 Caracterização do dilema do prisioneiro.....	75
2.6 O dilema do prisioneiro e o conceito de racionalidade.....	81
3 NOZICK E A NATUREZA DA RACIONALIDADE	86
3.1 As teorias evidencial e causal da decisão.....	88
3.2 O problema de Newcomb e os dois princípios de decisão.....	95
3.3 Uma questão de princípios.....	108
3.4 A teoria da ação de Nozick e o dilema do prisioneiro.....	118
3.5 De volta à racionalidade.....	128
4 CONCLUSÕES: NORMATIVIDADE E EQUILÍBRIO	138
Apêndice 1 – A regra de Bayes e as teorias da decisão	146
Apêndice 2 – O dilema do prisioneiro é um problema de Newcomb? ...	148
Apêndice 3 – A genealogia da teoria da ação de Nozick	149
Apêndice 4 – Exemplos de curvas de desconto que se cruzam	150
Referências	151
Índice remissivo	158

RESUMO

Nesta dissertação, o tema da decisão está circunscrito ao domínio da ação individual, embora a idéia de indivíduo abrigue entidades tais como uma empresa ou uma nação. A questão central é: *o que está por trás da decisão?* A resposta vai ser procurada nas correntes teóricas que se confrontaram ao longo do século XX. Discute-se se a decisão racional pode ser explicada apenas pela consistência interna das preferências, ou se é necessário considerar o contexto e outras variáveis. Um campo de testes foi fornecido pela teoria dos jogos, que permitiu simular situações em que os indivíduos têm interesses que são em parte comuns e em parte conflitantes, o que leva a um paradoxo: se cada um agir visando o interesse próprio, o resultado será desastroso para todos. A discussão central se dá em torno das opções estratégicas oferecidas pelo *dilema do prisioneiro*, um jogo que tem a pretensão de representar os conflitos de interesse no relacionamento humano. A solução tecnicamente correta desse dilema é instrumental, isto é, ela indica os meios (*a ação*) para a realização de um determinado fim (*a maximização da utilidade esperada*) de cada participante. Vários filósofos buscam uma solução diferente para o dilema, tal que justifique uma ação mais cooperativa. Dentre eles, **Nozick** propõe um novo modelo, o *valor-decisão*, no qual a utilidade instrumental continua sendo a parte fundamental, mas a ela deve ser adicionada a *utilidade simbólica*, que é o valor de uma ação como um símbolo. Essa nova parcela pode marcar uma atitude ética do agente, ao mostrar o tipo de pessoa que ele quer ser ou o estado de coisas que ele quer efetivar – o que seria o mesmo que buscar um outro fim, diferente daquele fim único relacionado à maximização da utilidade esperada. Dependendo do contexto, a nova parcela pode ser alta o suficiente para alterar o resultado que seria obtido por uma utilidade exclusivamente instrumental.

Palavras-chave: *decisão, teoria da decisão, decisão racional, racionalidade, dilema do prisioneiro, utilidade simbólica, valor-decisão, ética, Nozick.*

ABSTRACT

In this work, the theme of decision is bounded to the domain of an individual action, even though the idea of an individual comprises entities such as a company or a nation. The main question is: *What is behind the decision?* The answer will be investigated in theoretical currents that faced each other along the 20th century. It has been discussed if the rational decision can just be explained by an internal consistence of preferences, or if it is necessary to consider the context and other variables. A testing field was supplied by game theory that allows the simulation of situations where individuals have partly common and partly conflicting interests, which leads to a paradox: if each one acts according to his own interests, the result will be disastrous for all of them. The central discussion is performed around the strategic options offered by the *prisoner's dilemma*, a game that intends to represent the conflicts of interest in human relationship. The technically correct solution for this dilemma is instrumental, that is, it indicates the means (*the action*) for the accomplishment of a certain end (*the maximization of the expected utility*) for each participant. Several philosophers look for a different solution for this dilemma, justifying a more cooperative action. Among them, **Nozick** proposes a new model, the *decision-value*, in which the instrumental utility continues to be the fundamental part, but an additional *symbolic utility*, that is the value of an action as a symbol, must be added. This new portion can mark an ethical attitude of the agent, by showing which kind of person he wants to be or the state of things he wants to realize – it would be the same thing as searching for another end, different from that one related to the maximization of expected utility. Depending on the context, the new portion could be sufficiently strong to invert the result that would be obtained by an exclusively instrumental utility.

Key-words: *decision, decision theory, rational decision, rationality, prisoner's dilemma, symbolic utility, decision-value, ethics, Nozick.*

INTRODUÇÃO

“A Verdade é sempre o início do pensamento; pensar é sempre sem resultado. Esta é a diferença entre ‘filosofia’ e ciência: a ciência tem resultados; a filosofia, nunca. O pensar começa depois de sentir o efeito fulminante, por assim dizer, de uma experiência da verdade.”
Hannah Arendt, *Carta a Mary McCarthy*.¹

Uma das piadas mais famosas da história da filosofia conta que **Diógenes**, o Cínico, ao ouvir falar que **Platão** definira o ser humano como um “bípede implume”, depenou um galo e o levou ao local das aulas, dizendo: “Eis o homem de **Platão!**”². A velha anedota ilustra bem o eterno risco de qualquer descrição do ser humano, necessariamente parcial, tornar-se ridícula, quando o exagero da parte a transforma em caricatura. Talvez o autor de uma nova descrição conquiste alguma simpatia de seus leitores, se puder contar essa historinha logo no início do seu texto...

Não há como domar completamente esse risco, primeiramente devido aos equívocos e à ambigüidade da linguagem de que se valem os homens para criar, voluntariamente ou não, novos sentidos e referências para exatamente as mesmas palavras. Mas o problema central não estaria na descrição, e sim na dificuldade de captar o conteúdo desse objeto estranho e fugidio, que foi alvo de um conselho socrático aparentemente simples e destinado ao uso constante e renovado: “conhece-te a ti mesmo”³.

¹ Hannah Arendt, “Carta a Mary McCarthy”, *Folha de São Paulo*, Mais!, p. 5, 18/06/1995.

² Diógenes Laértios, VI.40, p. 162.

³ Platão, *Protágoras*, 343b, e *Fedro*, 230a. O autoconhecimento por meio da reflexão é também o tema principal de *O primeiro Alcibiades*, cuja autoria é duvidosa.

O tema desta dissertação – *a decisão* – não podia ficar imune a esse problema. Ela mesma uma parte do comportamento humano, é definida em associação com outro conceito, o da *racionalidade*, que também tenta descrever esse comportamento, constituindo a chamada *decisão racional*.

A racionalidade é objeto de estudo desde a antiguidade. Em um mito de **Platão**⁴, a razão é o cocheiro que comanda a parelha de cavalos alados, que representam as outras partes da alma, o apetite e o desejo. Em **Aristóteles**, a parte racional da alma tem duas faculdades, uma científica, que busca conhecer as coisas cujos princípios são invariáveis, e outra faculdade calculativa, que delibera a respeito daquilo que admite variações, isto é, sobre o futuro e sobre o que é contingente, como as decisões e as ações⁵.

Essas idéias estão presentes na definição dos três pólos em torno dos quais se aglutinam as diferentes teorias contemporâneas de racionalidade prática, como aos vértices de um triângulo⁶. No primeiro, baseado em **Aristóteles**, a razão reconhece o valor das ações nelas mesmas; no segundo pólo, derivado de **Hume** e contrário ao mito platônico, a razão é que está subordinada aos desejos; e no terceiro pólo, baseado em **Kant**, a razão se impõe, ao modo platônico, por meio da autoridade do imperativo categórico.

Esses pólos, tomados dois a dois, têm elementos comuns que se opõem às características do terceiro⁷:

- Kantianos e aristotélicos acreditam em razões categóricas naturais, que contrariam a visão neo-humana de que todas as razões normativas são hipotéticas.

⁴ Platão, *Fedro*, 246a-249d.

⁵ Aristotle, *Nicomachean Ethics (NE)*, VI.1.1139^a6-12 e VI.2.1139^b7.

⁶ Cullity e Gaut, p. 3.

⁷ Idem, p. 4.

- Neo-humeanos e kantianos, por sua vez, se unem na defesa da idéia de que o valor de uma ação é acrescentado pela decisão, e negam a posição aristotélica de que uma ação tenha valor por si mesma.
- Finalmente, aristotélicos e neo-humeanos admitem a existência de razões universais, mas não que a razão tenha autoridade para impor uma legislação universal.

Um resumo das três posições é apresentado na Figura 0.1:

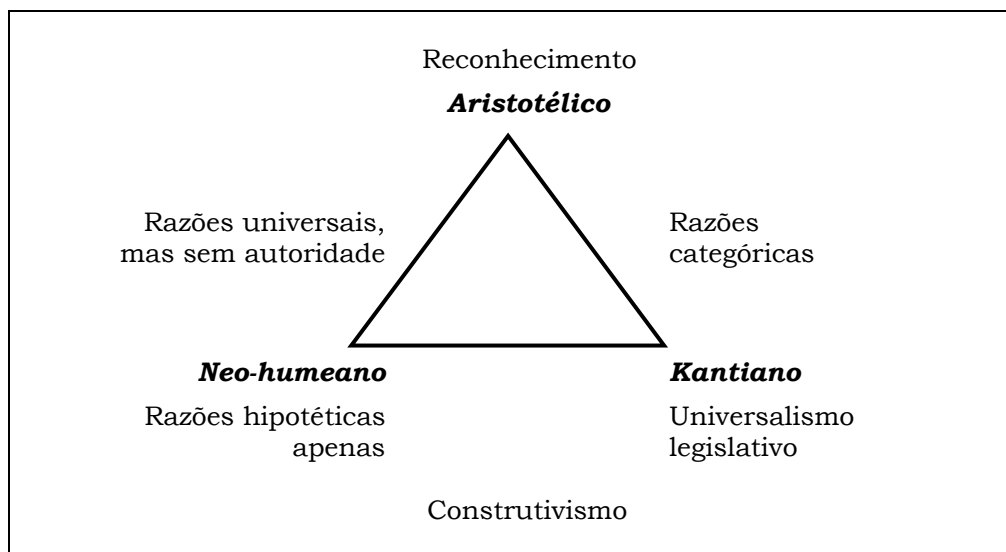


Figura 0.1 – Os pólos neo-humano, aristotélico e kantiano em torno dos quais se aglutinam as teorias contemporâneas de racionalidade prática. ⁸

A primeira dessas oposições, a da visão neo-humana, configura o que foi chamado de “desafio humeano” ⁹, por deixar aos aristotélicos e aos kantianos o ônus de provar a existência de razões normativas categóricas, que sejam eficazes como motivadoras da ação humana. Essa seria uma das questões de fundo deste trabalho: examinar a possibilidade de fins racionais moverem os homens.

* * *

⁸ Idem, p. 5.

⁹ Idem, p. 12. Referência à expressão cunhada por Bernard Williams (1981).

Para verificarmos na prática a influência dessas especulações teóricas, vamos examinar o caso de uma empresa¹⁰, que se considera ter muitas correspondências com um presumido modo de agir do homem, pois em ambos os organismos há uma convergência de processos que são coordenados internamente e que visam a um efeito em algum contexto externo.

Em uma empresa, os processos produtivos podem ser detalhados desde os objetivos mais gerais até o nível operacional mais baixo. As atividades e as estruturas da organização são construídas por necessidade técnica, embora algumas vezes sejam por conveniências pessoais, e sempre podem ser ajustadas conforme as novas situações exijam. Por sua vez, as decisões que a empresa toma a partir de objetivos específicos são perfeitamente localizadas e justificadas, apesar dos erros e ineficiências que podem acontecer.

No caso do homem, há muitos modos pelos quais as ações podem ser vistas, mas em geral as teorias descrevem um processo de decisão único, que representaria em poucas etapas um processo complexo cujos elementos não se conhecem bem e são de difícil detalhamento. As simplificações normalmente adotadas podem ser úteis em alguma finalidade prática, mas uma teoria da ação requer que se especifique exatamente que processos são esses, que fins estão em jogo e como eles se desdobram. Além disso, há necessidade de se compreender os conflitos envolvidos e localizar as suas origens.

¹⁰ Nosso interesse pela empresa não se deve ao fato de que seu comportamento seja uma extensão do comportamento de pessoas em seu interior. Embora as questões pessoais estejam subjacentes e possam gerar desvios, os agentes empresariais têm determinados papéis a cumprir, conforme padrões definidos. Nosso interesse se deve à contribuição que a empresa dá *individualmente* às ações coletivas que ocorrem no mercado, e aí está a analogia com o caso de uma pessoa que age em determinado contexto.

Uma das vantagens dessa analogia pessoa-empresa é de explicitar que também no homem acontecem processos complexos de decisão e ação, cujas características envolvem uma sensibilidade ao contexto e a contribuição de muitas variáveis e princípios, o que faz com que seja importante para o nosso propósito de autoconhecimento.

A escolha do exemplo empresarial trás ainda consigo questões filosóficas que tem a ver com a *minha* história de vida profissional. Ao trabalhar durante muitos anos em uma empresa de energia elétrica¹¹, pude participar de processos de tomada de decisão em diversos níveis, desde a simples escolha da melhor alternativa para ampliação de um elemento do sistema elétrico, passando pela avaliação de projetos de expansão que integravam obras afins, até a coordenação geral do plano de expansão da empresa.

Em todo esse trabalho havia sempre a necessidade de se estabelecer critérios básicos de natureza técnica e econômica para se aplicar no planejamento das ações de expansão, bem como modelos para avaliação de projetos e critérios para compatibilizar as demandas de setores do sistema aos recursos disponíveis. Percebia-se então que as decisões não ocorriam apenas nos níveis mais altos. Mesmo essa escolha de critérios técnicos, considerados “neutros” ou “objetivos”, embutia outras decisões que não eram tão visíveis, mas definiam coisas importantes como, por exemplo, quem seriam os beneficiários finais das medidas.

¹¹ A empresa de nossa referência é a Cemig, de quem o autor foi engenheiro de planejamento de sistemas elétricos durante toda a sua vida profissional. Uma das características principais do serviço público de energia elétrica é que a rede precisa ser expandida continuamente, para atender os novos consumidores que surgem em um ritmo próximo à taxa de natalidade de cerca de vinte anos antes. Até há alguns anos, esse número era de 4% ao ano.

Durante a década de 1980, foi disseminado nas empresas o uso de uma metodologia que incluía no processo de decisão a análise das oportunidades e das ameaças em um ambiente cada vez mais competitivo, consolidada na forma de planejamento estratégico. Um exemplo desse modelo revela certa semelhança com os passos descritos na *Ética a Nicômacos*, de **Aristóteles**, conforme é mostrado na Figura 0.2.



Figura 0.2 – Comparação entre o modelo de decisão aristotélico e um exemplo empresarial ¹⁴

¹² Aristóteles (*NE*, I.4.1095^a16-20) questiona se a felicidade é realmente o maior de todos os bens, porque, mesmo sendo uma opinião comum, há discordâncias sobre o que seja a felicidade. Há ainda uma outra questão. O tradutor H. Rackham (p. 10, n. a) observa que é inevitável traduzir *eudaimonia* por felicidade, mas que talvez o termo *bem-estar* interprete com maior precisão aquilo que seria não um estado de espírito, mas um tipo de atividade.

¹³ Para Sherman (1989, p. 71), se “deliberamos não sobre fins, mas sobre aquilo que contribui para os fins” (*NE*, III.3.1112^b12), então a primeira tarefa da deliberação será uma decisão sobre fins e sobre o modo pelo qual um fim será reduzido a fins intermediários até chegar à ação. Assim, em “uma dada etapa de deliberação, consideramos certos fins como fixos, (...) mas este fim-guia pode ser ele mesmo um produto de deliberações prévias”.

¹⁴ O desenho do modelo aristotélico foi baseado em Aristóteles (*NE*, I.4.1095^a18, II.1.2.1111^b26 e III.3.1112^b12-19), e nos comentários de Sherman, 1989, p. 71. O exemplo de modelo empresarial se refere a um trabalho realizado para a Cemig por Globalconsult, Rio de Janeiro, c1989, citado de memória.

Ao relacionarmos as semelhanças da Figura 0.2 com os pólos explicativos da Figura 0.1, percebemos que o processo de decisão da empresa se aproxima, em suas linhas estruturais, do pólo aristotélico. O que faltaria para que ele também reproduzisse outros pontos de vista como os indicados na Figura 0.1 e que estão ausentes nos modelos de decisão hoje dominantes?

Não se trata de duvidar da eficácia do modelo, mas de se preocupar de que ele, mesmo incompleto, seja eficaz também em seu involuntário papel educativo e de formação de hábitos. Um modelo de decisão é mais do que em geral aparenta, pois cada elemento “esconde” subprocessos que, embora importantes, são aproximações – eles podem falhar em pontos críticos.

É exatamente ao interior dos subprocessos que se volta a nossa questão principal: *o que está por trás das decisões?* Acreditamos que é aí que estão as oportunidades para se representar de maneira mais adequada como de fato decidimos, o que talvez nos faça incluir outros fatores tais como os “entrechoques das paixões” ou os princípios de uma “legislação universal”.

No nosso exemplo empresarial¹⁵, essas oportunidades se manifestavam cotidianamente em pelo menos dois aspectos, cuja consideração era reconhecida por melhorar a qualidade das decisões:

- A participação de pessoas que procediam de diferentes pontos da cadeia produtiva, o que propiciava não só o entrechoque das visões e expectativas, e a crítica permanente dos objetivos da decisão – temas que já faziam parte daquele modelo original mostrado na Figura 0.2 –, mas também:
 - ✓ a busca do conhecimento mútuo dos subprocessos produtivos, o que, por sua vez, era essencial para se conhecer o próprio processo e para se obter o consenso nas decisões que envolviam pesos;

¹⁵ Como mencionado na Nota 10, estamos interessados em ações que têm características individuais, isto é, embora possam ser tomadas por empresas, instituições ou países, integram com outros *indivíduos* em um determinado contexto.

- ✓ a valorização do comportamento ético no âmbito interno e externo à organização.
- O uso de modelos de decisão que considerassem não só as variáveis diretamente envolvidas na cadeia causal, mas também as que influenciavam os resultados de longo prazo, como, por exemplo, as conseqüências coletivas se todos – funcionários, parceiros, fornecedores, etc. – agissem com objetivos voltados exclusivamente para o interesse próprio.

Alguns autores¹⁶ reforçam nessa crítica um aspecto político: os atuais modelos formais de tomada de decisão, tais como a análise custo-benefício, “assumem que os fins são dados externamente, ou pela legislação ou pelas preferências reveladas pelo mercado”¹⁷. Essa falta de discussão sobre os fins pode nos levar a uma descrença no regime democrático e deixar-nos à mercê dos interesses de grupos organizados. A solução talvez seja que se considere toda decisão “uma questão ‘socrática’, (...) onde aquele que delibera é ao mesmo tempo o descobridor e a coisa a ser descoberta”¹⁸.

O caminho desde o “bípede implume” até uma descrição de homem que seja imune à crítica é longo e improvável. Mas há uma questão permanentemente atual que talvez mereça nossa atenção: trata-se de conhecer melhor o que o homem está fazendo consigo mesmo e, por extensão, com o seu contexto social imediato, ao se educar à semelhança do próprio instrumento falível que construiu. Este trabalho tem o pressuposto de que, ao examinar o *modo como* decidimos, estamos examinando quem *somos*, ou melhor, quem *queremos* ser, ou melhor, quem *devemos* ser.

Os resultados obtidos nesse caminho de autoconhecimento estão apresentados em quatro capítulos, resumidos a seguir.

¹⁶ Sherman (1989 e 1993) e Richardson, entre outros.

¹⁷ Richardson, p. xii.

¹⁸ Idem, p. 29. Referência a John Rawls (1971).

O Capítulo 1 relata uma discussão do conceito de racionalidade no âmbito da teoria hegemônica nos campos da economia e da tomada de decisão, mas que sofre críticas por adotar reduções em pontos considerados sensíveis por muitos. São confrontadas três concepções filosóficas contemporâneas diferentes, uma constitutiva do modelo (**Ramsey**) e duas (**Sen** e **Blackburn**) que defendem o seu alargamento.

O Capítulo 2 trata da teoria dos jogos ou da decisão estratégica, e é a complementação do contexto que começou a ser desenhado no capítulo precedente, nesse que é um subproduto da chamada guerra-fria e que tem aplicações, entre outras áreas, na biologia, na teoria econômica, nas ciências sociais, na psicologia e nas relações internacionais. A teoria dos jogos fornece elementos para a continuidade do debate sobre a racionalidade, agora com o foco dirigido ao *dilema do prisioneiro*, um problema que retrata o conflito entre o interesse individual e a possibilidade de cooperação coletiva.

O Capítulo 3 entra no âmago da teoria da decisão, para levar a discussão sobre a racionalidade a alguns pontos básicos, como o princípio da utilidade esperada, a taxa de desconto, os tipos de probabilidade envolvidos, etc. A ênfase desse capítulo é a contribuição que **Nozick** dá à teoria da decisão, em especial a solução do *problema de Newcomb* e sua aplicação ao *dilema do prisioneiro*. O resultado desse enfoque é um novo modelo, chamado de *valor-decisão*, que incorpora aos modelos existentes, em adição às variáveis causais e evidenciais, já incluídas em sua função instrumental, a dimensão dos valores éticos que o agente quer simbolizar por meio da ação.

O Capítulo 4 apresenta algumas conclusões do autor desta pesquisa.

CAPÍTULO 1 – UMA DISCUSSÃO CONTEMPORÂNEA DO CONCEITO DE RACIONALIDADE

- 1.1 *Ramsey e a racionalidade como consistência*
- 1.2 *Sen e a crítica da maximização do auto-interesse*
- 1.3 *Blackburn e o domínio das paixões*
- 1.4 *É possível conciliar a discussão?*

O campo de pesquisa que vamos privilegiar é definido pelo *utilitarismo*, filosofia de origem anglo-saxã que domina a cena econômica e constitui a base da economia do bem-estar desde o século XIX até os nossos dias. O conceito-chave dessa doutrina, o da *utilidade*, foi prefigurado¹ por David **Hume** (1711-1776), formulado por Jeremy **Bentham** (1748-1832) e desenvolvido por John Stuart **Mill** (1806-1873) e Henry **Sidgwick** (1838-1900).

O conceito de utilidade visa “maximizar o *bem-estar coletivo*, definido como a soma do bem-estar (ou da utilidade) dos indivíduos que compõem a coletividade considerada”². Essa medida coletiva é uma mera soma e não considera a maneira como o bem-estar se distribui entre os membros da coletividade. Surge então uma questão de *justiça social* que, embora importante como crítica ao utilitarismo, não será examinada neste trabalho.

Mas há ainda uma segunda questão, pressuposta no conceito de utilidade, que é a da *decisão racional*, e esta, sim, será o tema principal desta pesquisa. Para nos conduzir à discussão de como os conceitos de racionalidade e de utilidade estão associados no interior do utilitarismo, escolhemos

¹ Van Parijs, p. 29.

² Idem, p. 30.

três autores do século XX, **Ramsey**, **Sen** e **Blackburn**, que representam correntes de pensamento distantes em alguns pontos, mas próximos em outros. Longe de esgotar o assunto, a pretensão é que o retrato parcial a ser obtido nos revele os principais elementos do contexto filosófico em que se desenvolvem os temas abordados nos capítulos seguintes – quais sejam, a teoria da decisão e a contribuição que a ela deu Robert **Nozick** (1938-2002), e que são, enfim, os objetivos maiores deste trabalho.

Os três autores que formam o ambiente filosófico que vai ser descrito neste Capítulo 1 têm o seguinte perfil:

Frank Plumpton **Ramsey** (1903-30), inglês, contemporâneo de Bertrand **Russell** e **Wittgenstein**. Apesar da morte prematura, lançou alguns *papers* que tiveram grande influência nos estudos de probabilidade, indução e decisão. Defende a idéia de que a racionalidade seja vista como a capacidade de ordenar preferências de modo consistente.

Amartya **Sen** (1933), economista e filósofo indiano, Prêmio Nobel de Economia em 1998, trabalhou com **Nozick** em vários projetos acadêmicos. Aponta que a escolha não é apenas uma questão de consistência e de maximização da utilidade, mas depende do contexto da ação, e com isso outras variáveis, em especial as considerações éticas, entram em jogo.

Simon **Blackburn** (1944), inglês, autor de *Ruling Passions*, compartilha com **Hume** a posição de que a escolha é função dos desejos, e cabe à razão prover os meios para isso. Mas, além disso, no domínio da ética, a racionalidade tem um papel importante na aprovação intersubjetiva de novos padrões de comportamento.

1.1 Ramsey e a racionalidade como consistência

Quando **Ramsey** atinge o que parece ser o ponto alto de sua argumentação em *Truth and Probability* (1926), *paper* que lançou as bases da teoria da probabilidade subjetiva e da moderna teoria da decisão³, ele claramente dá importância e ênfase ao conceito de consistência: em três parágrafos⁴ cita sete vezes essa palavra ou alguma derivada. Naquele momento em que completava sua teoria de graus parciais de crença, **Ramsey** estabelecia a necessidade de consistência para qualquer conjunto de crenças parciais, pois a relação entre elas deveria obedecer às leis da probabilidade. Se alguém violasse essas leis, em especial a da transitividade das preferências, a decisão estaria comprometida, pois, sendo inconsistente, tal agente estaria nas mãos de um apostador esperto, que poderia fazer contra ele um *dutch book*, levando-o a perder em qualquer aposta.

O que seria exatamente a *consistência*? Em geral, esse conceito indica a ausência de contradição entre as frases verdadeiras de um sistema de crenças. Mas a consistência da ligação entre duas proposições pode ser interpretada de muitas maneiras, seja por uma relação necessária ou uma relação empírica, seja por uma descrição ou uma norma. No caso da consistência ao modo de **Ramsey**, vamos considerar inicialmente uma hipótese

³ Segundo Mellor, o *paper* de Ramsey mostra como nossas escolhas entre ações envolvendo risco podem ser usadas para medir tanto nossos desejos (utilidades subjetivas) quanto nossas crenças (probabilidades subjetivas) (Mellor, p. xvii). Ramsey antecipou com as teorias da utilidade e da probabilidade subjetiva os trabalhos, respectivamente, de von Neumann e Morgenstern (*Theory of Games and Economic Behavior*, 1944, desenvolvido independentemente de Ramsey) e de Savage (1951) (Mellor, p. xviii).

⁴ Ramsey, pp. 78-79.

que lhe atribui um caráter normativo, do tipo “não se deve aceitar uma proposição sem também aceitar a outra”⁵, vinculado a dois princípios:

- (1) não devemos aceitar crenças incompatíveis – e
- (2) não devemos aceitar uma crença sem também aceitar suas conseqüências lógicas.⁶

No primeiro princípio está implícita a exigência do conhecimento das relações lógicas envolvidas, como também das possíveis divergências de referência (“choveu e não choveu”, referindo-se a fatos separados no tempo) e de significado (“somos livres e não somos livres”, significando que somos livres para certas coisas e não somos livres para outras)⁷.

Já a normatividade do segundo princípio de consistência poderia ser pensada como um sistema de crenças dedutivamente fechado, que nos levaria a reiterar indefinidamente para considerar todas as possibilidades, o que seria uma tarefa impossível de ser cumprida por um ser finito.

Uma segunda interpretação para esse segundo princípio tem a ver com a distinção que J. **Cohen** faz entre crença e aceitação: “a crença é uma disposição para sentir, a aceitação uma atitude frente ao raciocínio”⁸. Além disso, a crença é involuntária e sensível à verdade, enquanto a aceitação é um processo que ocorre na vontade, no qual “aquele que aceita escolhe quais proposições tomará como premissa”⁹. Dessa forma, “quando uma crença produz outra crença, o resultado é concebido como involuntário e não como

⁵ Gensler, pp. 8 e 9 n.1.

⁶ Idem, p. 15.

⁷ Idem, p. 21.

⁸ Cohen, Jonathan, p. 5.

⁹ Idem, p. 22.

Se tomarmos a aceitação como um tipo de escolha racional, a observação de J. Cohen implica uma circularidade e uma petição de princípio, pois aquele que escolhe faz uma escolha prévia dos critérios de escolha. Entretanto, o regresso pode ser interrompido em um ponto em que a escolha de critérios recair em alguma crença, de natureza involuntária.

a manifestação de obediência a um princípio”¹⁰, não havendo, portanto, uma cadeia de comprometimento entre uma crença e as crenças derivadas. Bem diferente é o caso da aceitação, em que “a pessoa também aceita indiretamente ou inconscientemente todas as conseqüências deduzidas da proposição que ela aceita explicitamente”¹¹. As distinções propostas por J. **Cohen** estão resumidas na Figura 1.1. Da mesma forma que na primeira interpretação, chegamos também a um sistema dedutivamente fechado de conseqüências, com a diferença que agora estamos tratando da aceitação.

Características da Crença	Características da Aceitação
É uma disposição para sentir	É uma atitude frente ao raciocínio
Tem uma sensibilidade à verdade	Há uma escolha dos princípios e premissas
É um processo involuntário	É um processo que ocorre na vontade
Não há uma cadeia de comprometimento entre uma crença e as crenças derivadas	A pessoa aceita todas as conseqüências deduzidas das proposições já aceitas
↓	↓
Não há um sistema de crenças dedutivamente fechado	Chega-se a um sistema de aceitação dedutivamente fechado

Figura 1.1 – As características da crença e da aceitação¹²

As características involuntárias da crença nos levam a pensar em uma terceira forma de entender a normatividade daquele segundo princípio de consistência, qual seja: a que relaxa, a favor de crenças parciais, a exigência de crença total nas conseqüências lógicas da crença inicial.

Tanto a análise de J. **Cohen** quanto a de **Gensler** negam a possibilidade dos graus de crença¹³ e nos levam ao ponto onde **Ramsey** começou, isto é, na crítica ao conceito de probabilidade, de **Keynes**. Este sustentava que

¹⁰ Idem, p. 23.

¹¹ Idem, p. 27.

¹² Figura baseada em J. Cohen, pp. 5-27.

¹³ J. Cohen (pp. 110-6) não nega diretamente tal possibilidade, mas tem dificuldades em conciliar sua distinção de crença e aceitação com a concepção de crenças parciais.

entre duas proposições quaisquer, tomadas como premissa e conclusão, existe uma e apenas uma relação de um certo tipo, chamada de relação de probabilidade; (...) e se essa relação for de grau α , a partir da crença total na premissa, deveríamos, sendo racionais, chegar a um grau de crença α na conclusão.¹⁴

Contra essa relação objetiva, que não temos como conhecer¹⁵ e para a qual não temos como estimar graus de probabilidade, **Ramsey** propõe uma outra relação: “admitimos que nossa evidência é verdadeira, e então atribuímos um grau de crença à nossa hipótese, [que é uma visão ao mesmo tempo] subjetiva e com orientação prescritiva”¹⁶. Em vez de uma relação objetiva entre duas proposições, tomadas como premissa e conclusão, há uma relação subjetiva, que varia em grau, entre uma evidência, tomada como verdadeira, e uma proposição, tomada como hipótese. A Figura 1.2 resume a comparação entre os dois conceitos:

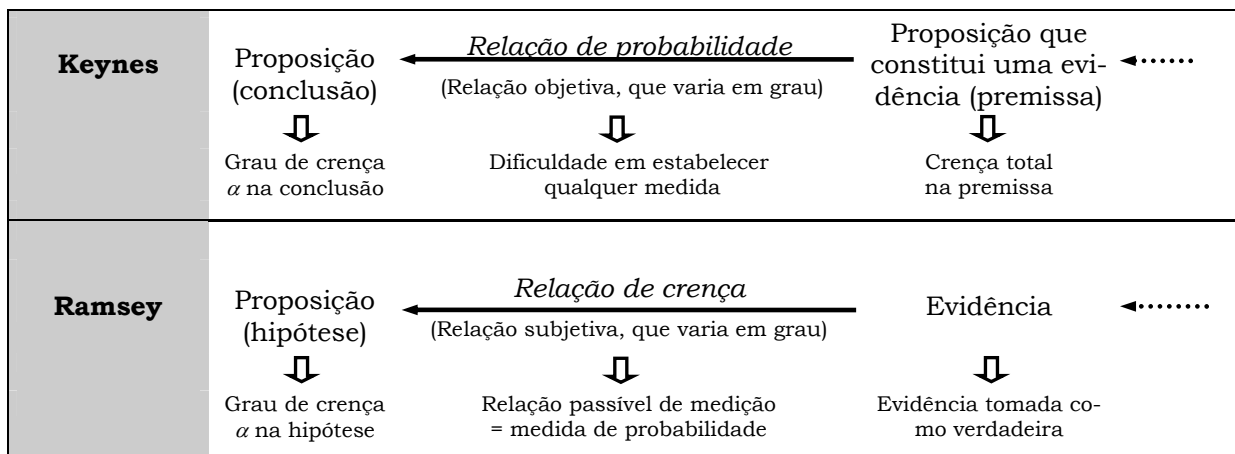


Figura 1.2 – A crítica de Ramsey ao conceito de probabilidade de Keynes.¹⁷

¹⁴ Ramsey, p. 56. Citação de J. M. Keynes (*A Treatise on Probability*, 1921).

¹⁵ A probabilidade objetiva apresentaria não apenas um problema epistêmico, mas também ôntico. De Finetti, por exemplo, “não acredita que haja tais probabilidades objetivas” (Jeffrey, p. 199), acreditando apenas em medidas subjetivas de probabilidade.

¹⁶ Sahlin, p. 13. Aproximadamente, é o que diz Ramsey, pp. 59-ss.

¹⁷ Figura baseada em Ramsey, pp. 56 e 66-67, e Sahlin, p. 13.

Ramsey acha que todas essas relações subjetivas, ou crenças, podem ser medidas, “algumas com maior precisão que outras, (...) dependendo da exatidão de como a medição é conduzida”¹⁸, tal como ocorre também na física. Ele mostra que seria possível construir uma série ordenada de graus, atribuindo a esses graus números que estariam entre 1 (crença plena na proposição) e 0 (crença plena no contraditório); e $\frac{1}{2}$ seria o índice para crenças iguais na proposição e em seu contraditório.

O método de **Ramsey** parte da hipótese de que “o grau de crença é uma propriedade causal dessa crença, e que pode ser expresso vagamente como o ponto até o qual estamos preparados para agir com base nela”¹⁹. O que diferencia a crença, de um lado, de outras atitudes proposicionais²⁰, de outro, é a sua eficácia causal: a crença “levaria à ação em circunstâncias favoráveis, tal como um torrão de arsênico é venenoso não porque matou alguém, mas porque mataria se alguém o comesse”²¹.

¹⁸ Ramsey, p. 63.

¹⁹ Idem, p. 65. Segundo Mellor, além de outras antecipações já citadas (ver Nota 3), Ramsey intuiu, com sua teoria de graus de crença, o mesmo paradigma que foi adotado pela explicação funcionalista da mente, ao tomar “as crenças, os desejos e outras atitudes como definíveis por seus efeitos em nossas ações, por suas causas (p. ex., nossas percepções) e por suas interações (p. ex., a crença que *p* satisfazendo o desejo que *p*)” (Mellor, p. xviii).

²⁰ Ramsey considera a crença como *idéia* em duas passagens do texto original, ambas na p. 66. Entretanto, depois de Bertrand Russell (Bäuerle e Cresswell, pp. 491 ss.) a crença passou a ser considerada uma das coisas “mentais” que, junto com esperanças, desejos e medos, formam as *atitudes proposicionais*. Tais expressões, embora tenham descrições funcionais diferentes, têm a característica comum de pedirem um complemento (“que ...”) que vai ser exatamente o objeto da atitude.

²¹ Ramsey, p. 66. Tanto “crença” como “venenoso” apresentam propriedades *disposicionais*, que diferem das propriedades *categoriais* (forma e massa, p. ex.) por terem uma tendência a se realizarem como ato, dependendo do contexto (ver Dancy e Sosa, verbete “*disposition*”, assinado por John Heil).

Essa definição de crença também se assemelha a uma análise alternativa da causalidade, que toma os efeitos como “‘dependentes contrafactualmente’ das causas” (Honderich, verbete “*causality*”), da seguinte forma: se minha crença em X causa minha ação Y, então se não acreditasse em X, não faria Y. Apesar dessa análise contrafactual não mencionar explicitamente uma sensibilidade ao contexto, ela parece existir, tal como na propriedade disposicional, pois a crença exige “algum pensamento ou experiência relevante para disparar uma disposição” (J. Cohen, p.1), ou seja, as “circunstâncias favoráveis” citadas por Ramsey.

Este método de medir crenças a partir de possíveis ações tem como base uma teoria da ação segundo a qual

agimos do modo que pensamos ser o mais adequado para concretizar os objetos de nossos desejos, de tal forma que as ações de uma pessoa são completamente determinadas por seus desejos e opiniões. Esta teoria não pode ser adequada a todos os fatos, mas parece-me uma aproximação útil à verdade.²²

Além de tomar o grau de crença, ou probabilidade subjetiva, como base para ação, essa teoria nasce vinculada à noção de *utilidade máxima*²³: uma pessoa busca coisas que deseja para si e para os outros (os “bens”) e “sempre escolhe o curso de ação que levará, em sua opinião, à maior soma desses bens”²⁴.

Seu comportamento é governado pelo que é chamado de esperança matemática. Isto quer dizer que, se p é uma proposição sobre a qual ele tem dúvidas, quaisquer bens para cuja realização, na visão dele, p é uma condição necessária e suficiente, entram em seu cálculo multiplicados pela mesma fração, que é chamada de 'grau de sua crença em p '. (...) Suponha que seu grau de crença em p é m/n ; então sua ação seria tal como ele escolheria para ser caso tivesse que repetir isto exatamente n vezes, em m das quais p seria verdadeira, e nas outras falsa²⁵.

Ramsey ilustra essa definição de crença parcial: estou em uma encruzilhada e não sei qual caminho seguir, mas acho que um deles é o certo. Va-

²² Ramsey, p. 69. O original “*opinions*” talvez pudesse ser melhor traduzido por “*crenças*”.

Como vamos ver ainda neste capítulo, e depois no Capítulo 2, esse é um enunciado primitivo da teoria causal, ou bayesiana, da decisão, uma teoria em que a ação racional é aquela que maximiza a utilidade esperada do agente.

²³ Ver Nota 3. No texto acima, o conceito de *utilidade* é apresentado como o resultado do desejo do agente. Entretanto, ao longo deste trabalho, vamos conhecer opiniões divergentes quanto à abrangência desse termo.

Em um sentido técnico, o termo *utilidade* tenta “quebrar qualquer suposta conexão entre racionalidade e egoísmo (...). Um benefício ou ‘utilidade’ consiste em qualquer situação que você deseje obter. Se você é altruísta, você pode desejar o bem-estar de outra pessoa, e então um aumento nesse bem-estar conta como uma utilidade para você” (Sainsbury, p. 54).

Dito de outra forma, o desejo de uma pessoa “significa o que ela quer, todas as coisas consideradas. (...) Entretanto, o que uma pessoa quer pode também significar qualquer coisa a que ela atribui algum valor positivo” (Sainsbury, p. 56).

²⁴ Ramsey, p. 70. No item 1.2 deste trabalho vamos ver uma oposição das idéias de *máxima utilidade* e de *consistência* como explicações concorrentes para o conceito de racionalidade. Na teoria da crença de Ramsey, entretanto, elas aparecem como complementares.

²⁵ Ibidem.

le a pena sair da estrada para perguntar a alguém que está ao lado, 400 m distante, no meio do campo? Depende da confiança que eu tiver na minha opinião: “quanto mais confiante estiver, menor será a distância que estarei disposto a percorrer para conferir minha opinião” ²⁶.

Para medir o grau de crença como *valor* ²⁷, **Ramsey** abandona um modelo baseado em apostas, adotado inicialmente, e trabalha com a hipótese de que, se dermos opções a uma pessoa, “podemos descobrir como ela põe em ordem de mérito todos os possíveis cursos do mundo” ²⁸. Se a pessoa

for capaz de ter dúvidas, então podemos testar seu grau de crença em diferentes proposições, fazendo-lhe ofertas do tipo: ‘o que você preferiria: ter o mundo α em qualquer situação; ou o mundo β se p for verdadeiro, ou o mundo γ se p for falso?’ ²⁹

Um exemplo de proposição p seria “a situação econômica vai melhorar” e as opções para a ação seriam estabelecidas em função da validade de p e de seus efeitos no mundo – no caso, descrições macroeconômicas. A questão que **Ramsey** coloca é que “se a pessoa estivesse certa da verdade de p , bastaria comparar α e β e decidir como se não houvesse qualquer condição; mas, havendo dúvida, a decisão entre β e γ não seria tão simples” ³⁰.

O próximo passo de **Ramsey** para atribuir valores de utilidade às ações é adotar oito axiomas que tratam “dos princípios que governam as escolhas desse tipo” ³¹.

²⁶ Ramsey, pp. 70-1.

²⁷ ‘Valor’, referido a um *mundo*, indica uma preferência que será medida em *utilidades*.

²⁸ Ramsey, p. 72.

²⁹ Idem, p. 73.

³⁰ Ibidem.

³¹ Ibidem.

O primeiro desses axiomas estabelece que “existe uma proposição eticamente neutra³² p em que se acredita com grau $\frac{1}{2}$ ”³³. Nas palavras de **Sahlin**, enquanto esse primeiro axioma é ontológico, os outros sete são axiomas de racionalidade, “que uma pessoa ‘racional’ deve satisfazer quando toma decisões”³⁴. Assim, o segundo axioma³⁵ impõe que se duas proposições forem desse tipo darão o mesmo resultado, seja qual for a situação em que se apliquem. Dos axiomas restantes, os principais são o terceiro e o quarto, que tratam da transitividade entre os resultados.

Pela aplicação desses axiomas, o comportamento de um agente será considerado racional se ele for capaz de ordenar as suas preferências, e com isso estabelecer uma função-utilidade para as possíveis ações.

Tendo um modo de medir utilidades, falta ainda uma etapa para chegar à medição da crença em geral. Neste último passo, que vai da medida da crença à medida de probabilidade, **Ramsey** utiliza o conceito de *probabilidade condicional*, que trata do “grau de crença em p dado q ”³⁶, para demonstrar que o grau de crença em uma proposição deve satisfazer aos axiomas da probabilidade, e que são as “leis fundamentais da crença provável”:

³² “Uma proposição atômica p é chamada de eticamente neutra se dois mundos possíveis que só diferirem com respeito à verdade de p forem sempre de mesmo valor; e uma proposição não-atômica p é chamada de eticamente neutra se todos os seus valores-verdade atômicos forem eticamente neutros” (Ramsey, p. 73). Ou, “definindo grosseiramente, grau $\frac{1}{2}$ de crença é aquela em que é indiferente apostar em um modo ou no modo contrário, supondo a mesma quantia apostada” (idem, p. 74).

³³ Idem, pp. 74-5.

³⁴ Sahlin, pp. 26-7.

³⁵ Idem, pp. 28-30. Seguimos Sahlin neste e nos outros comentários sobre os axiomas.

³⁶ Ramsey, p. 76, dá a seguinte explicação para ‘o grau de crença em p dado q ’: “Isso não significa o grau de crença em “se p , então q ”, ou em “ p implica q ”, ou [o grau de crença] que o sujeito teria em p se conhecesse q , ou o grau de crença que deveria ter. [O ‘grau de crença em p dado q ’] expressa aproximadamente as chances sob as quais o sujeito apostaria agora em p , sendo a aposta válida apenas caso q seja verdadeiro”.

Sahlin, p. 36, exemplifica: “queremos estabelecer a crença parcial em p , a situação econômica vai melhorar, dado q , o câmbio será desvalorizado em pelo menos 20%”.

- (1) Grau de crença em p + grau de crença em $\sim p = 1$.
- (2) Grau de crença em p dado q + grau de crença em $\sim p$ dado $q = 1$.
- (3) Grau de crença em $(p \text{ e } q) = \text{grau de crença em } p \times \text{grau de crença em } q \text{ dado } p$.
- (4) Grau de crença em $(p \text{ e } q) + \text{grau de crença em } (p \text{ e } \sim q) =$
 $= \text{grau de crença em } p$.³⁷

Em resumo, o caminho completo do método começa nas ações, que indicam preferências, que fornecem medidas de utilidade e que, por sua vez, se relacionam com os graus de crença em uma proposição. Se esses graus de crença forem consistentes, isto é, se satisfizerem aos axiomas da teoria da probabilidade, serão reconhecidos como medidas de probabilidade subjetiva³⁸ dessa proposição, conforme esquematizado na Figura 1.3:

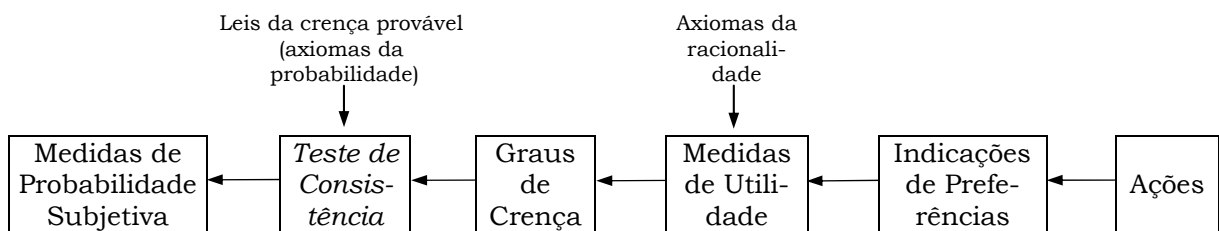


Figura 1.3 – Os graus de crença permitem relacionar medidas de utilidade com medidas de probabilidade

Nas etapas já descritas, são fundamentais as hipóteses relativas ao comportamento do agente racional: sua capacidade de ter dúvidas e a obediência aos axiomas de racionalidade. Ao indicar uma ordem de preferências das suas possíveis escolhas, e assim propiciar uma avaliação de suas respectivas utilidades, o agente demonstra certa consistência³⁹, mas ainda no âmbito da lógica formal. Aqui estamos de volta ao ponto alto da argumentação, já referido no início deste item, em que **Ramsey** afirma:

³⁷ Ramsey, p. 77.

³⁸ Ver Apêndice 1, “A regra de Bayes e as teorias da decisão”.

³⁹ Como se verá logo em seguida, no texto que faz referência à Nota 43.

Uma consideração precisa da natureza da crença parcial revela que as leis da probabilidade são leis de consistência, uma extensão à lógica formal, a lógica da consistência, para que esta possa abrigar as crenças parciais. As leis da probabilidade não dependem para o seu significado de qualquer grau de crença em uma proposição que seja especificamente apontada como racional; elas meramente distinguem os conjuntos de crenças que as obedecem como consistentes. ⁴⁰

Apesar de ter abandonado a idéia de aposta no desenvolvimento de seu método, **Ramsey** continua produzindo analogias entre aposta e ação em vários momentos de *Truth and Probability* ⁴¹. **Sahlin** defende que, mais do que uma substituição de métodos, houve uma melhoria, pois “o método de apostas e as definições de **Ramsey** não são essencialmente diferentes” ⁴². Na conclusão de seu argumento **Ramsey** confirma essa relação:

Ter um grau qualquer de crença implica certa medida de consistência, isto é, uma disposição para apostar em uma dada proposição com os mesmos riscos para qualquer aposta, as apostas sendo medidas em termos de valores últimos. Ter graus de crença que obedecem às leis da probabilidade implica uma medida adicional de consistência, qual seja, uma consistência entre os riscos aceitáveis em proposições diferentes, tal que previna um *livro* feito contra você.

(...) Isto é baseado fundamentalmente em apostas, mas não parecerá irrazoável se pensarmos que, de certo modo, em toda a nossa vida estamos apostando. Sempre que vamos para a estação estamos apostando que um trem realmente vai passar, e, se não tivéssemos um grau suficiente de crença nesse fato, deveríamos recusar a aposta e ficar em casa. ⁴³

O *livro* a que **Ramsey** se refere, hoje chamado de *livro holandês* (*Dutch Book*) ⁴⁴, é a possibilidade de alguém, intencionalmente, percebendo que o conjunto de crenças de um agente é inconsistente, tirar proveito disso, levando-o a fazer apostas em que este perderia em qualquer circunstância. A

⁴⁰ Ramsey, p. 78. Ver texto que faz referência à Nota 4.

⁴¹ Ver referências a *bet*, *betting* e *better* (= *bettor*) em Ramsey, pp. 67-8, 72, 74, 76, 78 e 79.

⁴² Sahlin, p. 39.

⁴³ Ramsey, pp. 78-9. Grifo nosso.

⁴⁴ Hacking, p. 169, prefere chamar o *Dutch Book* de “contrato de perda certa” (“*sure-loss contract*”).

A respeito do termo *Dutch Book*, Wakker apresenta algumas especulações históricas: (a) ele teria sido inventado no século XIX, quando companhias de seguro holandesas teriam feito seguros de navios de modo a ganhar dinheiro em qualquer contingência. Mas não há evidências desse fato. (b) O idioma inglês tem várias expressões negativas que usam o termo *Dutch*, tais como *Dutch uncle*, *Dutch treat*, etc. Tal terminologia parece datar do século XVII, quando a Holanda e a Inglaterra tinham frotas poderosas e rivais.

demonstração do livro holandês feita por **De Finetti**⁴⁵ implica em uma exigência de *coerência*, termo que parece ser mais apropriado, conforme chama a atenção o tradutor desse autor para o inglês:

O termo ‘consistência’, usado por alguns autores ingleses e americanos, é perfeitamente aceitável a De Finetti, mas é ambíguo (do ponto de vista lógico) porque, aplicado a crenças, tem outro significado muito preciso e explícito na lógica formal. Da forma como as palavras são usadas nesta tradução, dizer que um corpo de crenças é ‘consistente’ é o mesmo que dizer (como em lógica) que ele não contém duas crenças contraditórias. Dizer em adição que o corpo de crenças é ‘coerente’ é dizer que os *graus* de crença satisfazem certas condições além dessa⁴⁶.

Assim, se as crenças de um agente (ou suas probabilidades subjetivas) são coerentes, elas obedecem às leis do cálculo de probabilidade. Para **Resnik**, “se aceitarmos que a coerência seja uma condição de racionalidade, a conclusão de **De Finetti** pode ser parafraseada como *as leis da probabilidade são condições necessárias para a racionalidade*”⁴⁷, e que é possível demonstrar que ter probabilidades subjetivas que obedecem ao cálculo de probabilidades é suficiente para a coerência. Devido a essa conversão, o agente fica liberado para modificar seus graus de crença à luz de novos dados: ao seguir as regras de condicionalidade, suas novas probabilidades subjetivas obedecerão às leis de probabilidade, e então continuarão coerentes⁴⁸.

Podemos comparar a definição de coerência de **Ramsey/De Finetti** com aquela que é adotada pelas teorias coerentistas do conhecimento. Estas, em geral, defendem que uma crença é justificada por outras crenças, formando um sistema de crenças consistentes entre si. Entretanto, apenas a consistência não é suficiente para a coerência, e cada variante dessa teoria

⁴⁵ Resnik, p. 69.

⁴⁶ De Finetti, p. 55. Texto introdutório de Henry E. Kyburg, Jr.

⁴⁷ Resnik, p. 75.

⁴⁸ Ver texto que faz referência à Nota 46.

acrescenta algum requisito, tal como a força (ou o grau) das relações inferenciais que unem as crenças⁴⁹.

Uma dessas variantes atribui ao *equilíbrio reflexivo* do nosso sistema de pensamento a capacidade de transformar meras opiniões em entendimento. O equilíbrio reflexivo é o ponto de convergência de um processo de ajustamentos em que, de um lado, estão as inferências que ligam nossas crenças – o que **Elgin** chama de “compromissos epistêmicos anteriores” – e, de outro, as investigações que fornecem as evidências empíricas necessárias à revisão dessas inferências e teorias. Dessa forma, o equilíbrio reflexivo acrescenta à coerência, que é puramente formal, um contexto, que inclui a força das inferências anteriores bem como as novas experiências⁵⁰.

Ao fazer um balanço das suposições que **Ramsey** adota, vemos que, em geral, a ação do sujeito está voltada para a satisfação de seus desejos, e que para isso escolhe os meios necessários para obter os melhores resultados. Tal agente vê a vida como uma disputa, em que está disposto a apostar para ganhar; fixado o seu interesse e avaliadas as crenças parciais que tem a respeito do mundo, ele pode então organizar suas preferências para a ação. O agente é considerado racional quando for capaz de ordenar tais preferências de modo consistente; **Ramsey**, entretanto, acrescenta uma nova condi-

⁴⁹ Ver Craig, verbete “*Knowledge and justification, coherence theory of*”, assinado por L. Bonjour, e verbete “*Justification, epistemic*”, assinado por R. Foley.

⁵⁰ Elgin, pp. 49-51.

O equilíbrio reflexivo também pode ser pensado: (1) em termos da relação geral x particular; (2) para a introdução de temas substantivos, como os éticos, a respeito da natureza humana e (3) para a consideração da sensibilidade da escolha em relação ao contexto – mudou o contexto, muda a escolha. Os temas 2 e 3 são preocupações de Amartya Sen, como se verá no item 1.2.

A relação entre crenças e equilíbrio pode também ser vista no texto que faz referência à Nota 119, Capítulo 2, e no Apêndice 1, item 2.

ção: que os graus de crença obedeçam às leis da probabilidade, tornando o agente e suas apostas imunes ao *Dutch Book*.

Esta é a proposta de **Ramsey** para uma teoria da racionalidade, compreendendo o que é ser racional, e o que é escolher e agir racionalmente. Como vimos, **Ramsey** tem algumas suposições básicas a respeito do homem racional, quais são seus objetivos e sua visão do mundo. A escolha racional é definida em termos de uma capacidade especial de implementar tais objetivos, dadas as suas crenças. A consistência, ou coerência, é a principal característica de um conjunto de escolhas racionais. E a ação de um indivíduo vai interagir com as ações semelhantes de outros indivíduos também racionais, como em um jogo.

1.2 Sen e a crítica da maximização do auto-interesse

Vamos examinar neste item a concepção de racionalidade de Amartya **Sen**. Enquanto em **Ramsey** vimos o conceito de racionalidade ser desenvolvido em um âmbito formal, toda a reflexão que vamos examinar agora é feita no campo da economia, por isso é importante acompanhar um certo desenvolvimento da teoria econômica, em que o conceito de racionalidade, pelo menos na crítica de **Sen**, ocupa um papel “intermediário” entre o comportamento real do ser humano e um modelo explicativo desse comportamento.

Sen aponta duas origens históricas distintas para a economia, uma relacionada à “ética” e outra à “engenharia”⁵¹. Na primeira, a origem está em **Aristóteles** que, na *Ética a Nicômacos*, associa a economia à busca dos fins; nessa abordagem, há duas questões cruciais: a motivação própria do homem, representada pela questão socrática: “como devemos viver?” e, em seguida, o problema da realização do homem, cuja finalidade é alcançar o bem, não isolada, mas socialmente, na polis. Na outra origem, a da engenharia, a economia está empenhada em resolver questões logísticas, isto é, quer encontrar os meios para atingir os fins, pois estes já seriam dados diretamente.

O desenvolvimento da abordagem “engenheira” da economia é mais recente e ganhou amplo predomínio em relação às questões éticas. Para **Sen**, a “natureza da economia moderna foi substancialmente empobrecida pelo distanciamento crescente entre economia e ética”⁵², em especial pelo abandono das considerações relacionadas à motivação humana e à realização no âmbito social. Na economia moderna o pressuposto de um “comportamento ra-

⁵¹ Sen, 1987, p. 20.

⁵² Idem, p. 23.

cional” do homem tem papel fundamental: descreveria um comportamento real, servindo de modelo para previsões.

Sen argumenta que a caracterização do comportamento racional, da forma como é feita pelo modelo dominante, pode não estar correta; mesmo se estivesse, as pessoas não agiriam necessariamente assim, por causa de erros, confusões e experimentações⁵³. Por isso, **Sen** defende uma concepção de racionalidade que admita padrões de comportamento alternativos.

Na teoria econômica tradicional há dois métodos predominantes para definir a ação racional: “um deles consiste em conceber a racionalidade como uma *consistência interna* da escolha; o outro, em identificar racionalidade com *maximização do auto-interesse*”⁵⁴. No primeiro caso, **Sen** enfatiza que a consistência em si mesma não é um critério suficiente, pois se uma pessoa agisse contra seus próprios interesses de maneira consistente, não poderia ser considerada racional; além disso, a consistência de um conjunto de escolhas “deve depender da *interpretação* dessas escolhas e de algumas características *externas* à escolha propriamente dita (por exemplo, a natureza das nossas preferências, objetivos, valores, motivações)”⁵⁵.

Sen consolida essas idéias em *Internal Consistency of Choice*, onde “argumenta contra a imposição de exigências *a priori* de ‘condições internas’ de escolha”⁵⁶, negando que seja possível haver uma tal consistência independentemente do contexto. Isso fica claro especialmente no campo das ciências sociais, em que a consistência interna, ou “racionalidade social”, pre-

⁵³ Idem, p. 27.

⁵⁴ Idem, p. 28. A concepção de Ramsey, como vista no item 1.1, integra os dois métodos como complementares. Ver Nota 24.

⁵⁵ Idem, p. 30.

⁵⁶ Sen, 1993, p. 495.

cisaria ser justificada em termos de noções não imediatas, como “preferência da sociedade” e “utilidade social”. Essa crítica, para **Sen**, apenas aparentemente contesta a teoria da preferência revelada ⁵⁷, de **Samuelson**, pois na realidade ela seria contra uma certa interpretação dessa teoria ⁵⁸.

Os problemas com a consistência interna de escolha estariam em dois níveis: no fundacional e no prático. No nível fundacional, a dificuldade básica é com a suposição implícita de que “atos de escolha são, por si mesmos, como *proposições* que podem ou se contradizer, ou ser consistentes umas com as outras” ⁵⁹. Ilustra a questão:

As proposições A e $\sim A$ são contraditórias de um jeito que escolher x de $\{x, y\}$ e y de $\{x, y, z\}$ podem não ser. Se o primeiro par de opções devesse acarretar respectivamente as proposições

- (1) x é uma alternativa melhor que y , e
- (2) y é uma alternativa melhor que x ,

então haveria de fato uma contradição (...). Mas essas opções não acarretam, *por si mesmas*, quaisquer proposições desse tipo. Se fossem *dadas* algumas idéias do que a pessoa está tentando fazer (e essa é uma correspondência externa), poderíamos ser capazes de ‘interpretar’ essas ações como insinuado pelas proposições. Mas não podemos concluir isso sem invocar uma referência externa. Não há tal coisa de consistência de escolha *puramente* interna. ⁶⁰

⁵⁷ A teoria da preferência revelada diz que “se uma coleção de bens y pudesse ter sido comprada por um certo indivíduo dentro de seu orçamento, quando de fato ele fosse observado a comprar outra coleção x , será presumido que ele revelou uma preferência de x sobre y . (...) Do ponto de vista introspectivo da pessoa em questão, o processo acontece da sua preferência à sua escolha, mas do ponto de vista de um observador científico, a seta vai na direção oposta: as escolhas são observadas primeiro e então as preferências são presumidas a partir dessas observações. A condição de consistência sobre a qual Samuelson baseou sua teoria, que se tornou conhecida como o ‘Axioma Fraco da Preferência Revelada’, diz que se uma pessoa revelar uma preferência (...) de x sobre y , então ela também deve não revelar uma preferência de y sobre x .” (Sen, 1973, pp. 60-1).

Apesar de parecer inócua, Samuelson conseguiu, na opinião de Sen, ter bons resultados com a aplicação dessa teoria ao estudo do comportamento dos consumidores.

⁵⁸ Sen, 1993, p. 496. Citação de P. A. Samuelson, “A Note on the Pure Theory of Consumers’ Behaviour”, *Economica*, 5, 1938, pp. 61-71.

Segundo Sen, Samuelson lhe comunicou pessoalmente que “não via os axiomas da teoria da preferência revelada como condições de consistência interna da escolha ou do comportamento (...) [pois] os axiomas podem ser interpretados como *conseqüências* da maximização da utilidade” (Sen, 1993, p. 496, n. 5).

⁵⁹ Sen, 1993, pp. 498-9. A expressão “*statement*”, utilizada por Sen nesta e em outras passagens da referência citada, foi tomada como sinônima de “*proposition*”, e assim traduzida por “*proposição*”, para indicar que se trata do conteúdo de uma sentença.

⁶⁰ Idem, p. 499.

Sen vai além: “mesmo as ações aparentemente contraditórias de ‘dizer A ’ e ‘dizer $\sim A$ ’ podem não ser rigorosamente contraditórias, da forma como são as duas proposições”⁶¹, pois, ao executá-las, a pessoa pode querer se mostrar mentalmente perturbada ou simplesmente querer confundir⁶². O problema com a consistência interna é que há muitos tipos de escolha, de interesse de economistas e outros cientistas sociais, que não funcionam na prática. Por exemplo, a “propriedade α ”⁶³, uma das condições básicas de consistência, pode ser violada da maneira como se mostra a seguir:

Suponha que uma pessoa, em uma mesa de jantar, tenha de escolher entre pegar a última maçã de uma cesta (y) ou não pegar nada (x). Ela decide se comportar decentemente e não pegar nada (...). Se, em vez disso, a cesta tivesse duas maçãs, e a escolha agora fosse entre não pegar nada (x), pegar uma maçã (y) ou pegar duas maçãs (z), ela poderia muito razoavelmente escolher pegar uma maçã (y), sem violar qualquer regra de bom comportamento. A presença de outra maçã (z) torna a escolha de uma das duas maçãs uma escolha decente, mas essa combinação de escolhas violaria as condições-padrão de consistência, incluindo a propriedade α , mesmo que não haja nada de particularmente ‘inconsistente’ neste par de opções.⁶⁴

⁶¹ Ibidem. **Sen** se apega ao conceito tradicional da contradição. Entretanto, como o próprio princípio de não-contradição requer um mesmo tempo e uma mesma referência para afastar qualquer contradição (Aristotle, *Metaphysics*, IV.3.1005^b19), fica aberta a possibilidade de dúvidas de caráter pragmático, pelo menos no nível prático em que **Sen** está desenvolvendo a sua crítica à consistência interna de escolha. Assim, nem mesmo a conjunção de uma proposição com a sua negação constituiria sozinha um par contraditório, pois a proposição, ao fornecer o conteúdo, é apenas parte de um ato de fala – a outra parte é a força ilocucionária, que é a interação com o ambiente. Poderíamos, então, fixados um agente e um contexto, perceber a contradição nos atos que concorrem para fins opostos.

⁶² A importância do contexto nos atos de fala é enfatizada por **Recanati**, para quem “duas sentenças expressam a mesma proposição se e somente se elas forem verdadeiras nas mesmas circunstâncias” (Recanati, p. 4). Outra forma de relacionar uma proposição com um contexto é indicar que o conteúdo de uma sentença é indexical, pois depende de quem a proferiu, onde e quando. Como consequência, “um sujeito racional pode acreditar e não acreditar na mesma proposição, desde que acredite se ela estiver sob certo disfarce, e não acredite se ela estiver sob outro disfarce” (ibidem), o que é uma maneira de descrever a relação de crença entre uma evidência e uma proposição, como fez **Ramsey** (ver Figura 1.2).

⁶³ Sen, 1993, p. 500. A propriedade α estabelece que se um elemento de um conjunto S é escolhido e pertence a um subconjunto T de S , então ele deve ser escolhido também em T .

⁶⁴ Idem, p. 501. Uma descrição semelhante é dada pela “*lógica não-monotônica*”. Uma relação de inferência é chamada de não-monotônica se a adição de uma nova premissa altera a conclusão anteriormente obtida. Enquanto as inferências dedutivas são monotônicas, as inferências do dia-a-dia tem um caráter não-monotônico, pois as premissas práticas são dedutivamente insuficientes. Nesse caso, as inferências envolvem risco e qualquer nova informação tende a provocar a revisão das conclusões anteriores. (Ver **Craig**, verbete “*Non-Monotonic Logic*”, assinado por A. **Fuhrman**).

Na opinião de **Sen**, a pessoa que escolhe uma maçã quando poderia escolher duas continua sendo um maximizador porque, apesar da ordem de preferência das alternativas variarem com o menu, para cada um desses há uma opção que maximiza a decisão. Assim, as condições de consistência interna⁶⁵ podem deixar de ser necessárias em uma interpretação mais ampla de maximização.

A violação das condições de consistência interna acontece de várias maneiras, categorizadas por **Sen** conforme os contextos que representam situações de escolha bem definidas em termos de tempo e espaço⁶⁶:

- (i) *Escolha posicional*: a pessoa não quer pegar o maior pedaço de bolo, nem ser o primeiro a terminar um trabalho ou furar uma greve (...).
- (ii) *Valor epistêmico do menu*: o que é oferecido para escolha pode dar informações sobre a situação subjacente, que então influencia a preferência entre as alternativas (...).
- (iii) *Liberdade para rejeitar*: algumas escolhas são engendradas para rejeitar, de um modo livre, certas ações ou efeitos em favor de uma alternativa desejada. Por exemplo, jejuar não é o mesmo que estar faminto, mas estar faminto deliberadamente quando há liberdade para comer. (...) ⁶⁷

Ocorrendo condições externas relevantes, há uma pluralidade de correspondências internas que torna difícil estabelecer padrões que sejam úteis como justificativas de consistência – esta seria apenas aparência, pois, dependendo das circunstâncias, seria facilmente violada na prática. Assim, fora de um contexto, não seria possível estabelecer uma ordem de preferências das escolhas, que é fundamental no método proposto por **Ramsey**⁶⁸.

* * *

⁶⁵ Ver Nota 58, sobre a comunicação entre Samuelson e Sen.

⁶⁶ Haveria ainda um segundo tipo de contexto, definido pela cultura, que impõe valores a serem obedecidos nas decisões, tornando necessárias certas ações. Esse segundo tipo vai ser examinado junto à questão ética, alguns parágrafos à frente. Na prática, os dois tipos de contexto, o da ocasião e o da cultura, coexistem com uma certa gradação. Ver Nota 75.

⁶⁷ Sen, 1993, p. 502.

⁶⁸ Ao longo da crítica à consistência interna de escolha, Sen reconhece a existência de padrões mais complexos de comportamento, como os que vão ser explorados no Capítulo 3.

Voltemos ao ponto em que **Sen** apresenta, em *Sobre Ética e Economia*, os dois métodos predominantes na teoria tradicional para definir a ação racional. Já examinamos o primeiro deles, o que concebe a racionalidade como uma consistência interna da escolha; vamos agora examinar aquele que identifica a racionalidade com maximização do auto-interesse. Nesse segundo método, é pressuposta uma correspondência entre as escolhas que uma pessoa faz e seu auto-interesse. A origem dessa hipótese é atribuída a Adam **Smith**, mas este, segundo **Sen**, teria na verdade examinado

as diferentes disciplinas de ‘amor-próprio’, ‘prudência’, ‘simpatia’, ‘generosidade’ e ‘espírito público’, entre outras, e discutiu não somente sua importância intrínseca, mas também seu papel instrumental no sucesso da sociedade e sua influência prática no comportamento. ⁶⁹

Na avaliação de **Sen**, **Smith** considerava a prudência, que seria a união do amor-próprio e do autodomínio, “como sendo ‘de todas as virtudes a que mais auxilia o indivíduo’, ao passo que ‘humanidade, justiça, generosidade e espírito público são as qualidades mais úteis aos outros’” ⁷⁰. O que prevaleceu na história, contudo, foi a visão da prudência como maximização do auto-interesse, e do auto-interesse como qualidade única da razão⁷¹. **Sen** quer demonstrar que essa interpretação não tem validade universal:

por que deveria ser *unicamente* racional empenhar-se pelo auto-interesse, excluindo todo o resto? Evidentemente, pode não ser de todo absurdo afirmar que a maximização do auto-interesse não é irracional, pelo menos não necessariamente, mas asseverar que tudo o que não for maximização do auto-interesse tem que ser irracional parece absolutamente insólito. ⁷²

⁶⁹ Sen, 1995, p. 15.

⁷⁰ Sen, 1987, p. 39. Citação de Smith, 1790 (Oxford, 1975, p. 189).

⁷¹ Para Sen, Smith foi vítima de uma ironia, pois, enquanto criticava os filósofos pela tendência a adotar um princípio único para explicar todos os fenômenos, “essa ‘particular predileção’ viria a ser atribuída ao próprio Smith por seus exaltados admiradores ao fazê-lo ‘guru’ do auto-interesse (contrariando o que ele realmente afirmou)” (idem, p. 40).

⁷² Idem, p. 31.

A estratégia desse modelo é “usar o conceito de racionalidade como um ‘intermediário’ (...) para conduzir à proposição de que o comportamento *real* tem de ser maximizador do auto-interesse”⁷³. A crítica de **Sen** é que, além de ser inapropriada, a concepção da racionalidade centrada no auto-interesse rejeita a motivação ética, ou seja, não considera como racional qualquer comportamento que tenha objetivos desvinculados do interesse próprio. Enquanto é aceitável dizer que as pessoas buscam a maximização de seus interesses, é absurdo sustentar que a maximização do auto-interesse, isto é, o egoísmo universal, seja requisito de racionalidade⁷⁴.

Embora haja muitos argumentos a favor da teoria do auto-interesse, alguns baseados na expectativa de eficiência, os resultados empíricos mostram a ocorrência de uma pluralidade de motivos para a ação, além do auto-interesse, tais como o “dever, a lealdade e a boa-vontade”⁷⁵.

⁷³ Idem, pp. 31-2.

⁷⁴ Idem, p. 32.

⁷⁵ Idem, p. 34. Ver Nota 66, que distingue os contextos de ocasião e de cultura.

Ao comentar o *Ensaio sobre a dádiva*, de Marcel Mauss (1923-24), Laburthe-Tolra e Warnier (p. 341) dizem que a dádiva é uma motivação paradoxal: “ela é por definição voluntária, e, por isso, gratuita. Entretanto, todas as sociedades obrigam os indivíduos a praticar a troca. A dádiva é ao mesmo tempo voluntária e obrigatória. Ela possui uma das características do fato social tal como o definia Durkheim: impõe-se ao indivíduo. É coercitivo. (...) Da mesma forma que a obrigação de dar pesa sobre o doador, o beneficiário da dádiva encontra-se na obrigação de aceitá-la, e de retribuí-la”. O exemplo da dádiva mostraria também que “em todas as sociedades, os indivíduos fazem escolhas que (...) [são] dependentes dos quadros de percepção e ação fornecidos pela civilização” (idem, p. 347) e que, por isso, qualquer modelo de racionalidade deve levar em conta o “predomínio das civilizações sobre os comportamentos individuais” (ibidem).

Entretanto, a pesquisa de Mauss não fala em “predomínio”, mas em “influência” do contexto na ação humana, o que parece dar fundamento à tese de Sen. Para Mauss (p. 188), “entre a economia relativamente amorfa e desinteressada [em sociedades primitivas] e a economia individual e de puro lucro (...), sobrepôs-se toda uma mesma série de instituições e de acontecimentos, a qual não é governada pelo racionalismo econômico”. O motivo que anima as dádivas e os gastos naquelas sociedades “não é a fria razão do mercador, do banqueiro e do capitalista”, mas uma ação que é feita “para ‘obrigar’” ou “para humilhar o primeiro doador” (ibidem). Na nossa própria sociedade ocidental, nem todos somos “animais econômicos”, pois “o gasto puro e irracional é de prática corrente” (idem, p. 189). Por outro lado, a tese de Sen vai além do auto-interesse e do contexto social como motivadores da ação. Ele acrescenta, como se verá adiante, a liberdade da condição de agente do homem.

Sen acrescenta a essa lista a identidade social, que é tida em muitas teorias como

o principal determinante da compreensão das pessoas sobre o mundo, seus modos de raciocinar e suas concepções de racionalidade, suas normas de comportamento e práticas, sua moralidade pessoal e compromissos políticos.⁷⁶

A conclusão de **Sen** é que este “homem econômico”, que busca seus próprios interesses, é um modelo de racionalidade limitado e equivocado. Teria essa crítica um valor especial por vir de alguém com experiência de vida em outra cultura, e que por isso é sensível aos defeitos do modelo dominante na sociedade ocidental? Ou, ao contrário, esse fato invalidaria a crítica?⁷⁷

Em primeiro lugar, **Sen** descreve mais semelhanças do que diferenças entre o Ocidente e o Oriente. As alegações de que uma abordagem racional é tipicamente ocidental, ou de que as sociedades não-ocidentais dão pouca ênfase a valores tais como liberdade, tolerância, respeito mútuo e simpatia, não resistem à constatação de que “muitas idéias consideradas quintessencialmente ocidentais também floresceram em outras civilizações”⁷⁸.

Em segundo lugar, **Sen** reconhece que há uma “dominação da cultura ocidental contemporânea sobre as nossas percepções”⁷⁹, o que dificulta a verificação de fronteiras culturais e de até que ponto algumas crenças que realçam qualidades ocidentais estão implícitas em muitas discussões.

Em terceiro lugar⁸⁰, uma pessoa pode ter simultaneamente múltiplas identidades, como, por exemplo, nascer na Itália, ser mulher, feminista, ve-

⁷⁶ Sen, 1999, p. 6.

⁷⁷ O *princípio de caridade* nos faz supor que Sen “pensa como os ingleses”, parodiando a ironia de Blackburn a respeito do princípio formulado por Davidson.

⁷⁸ Sen, 2000, p. 14.

⁷⁹ Idem, p. 10.

⁸⁰ Sen, 1999, pp. 14-17.

getariana, novelista, etc., e que a relevância de cada uma depende do contexto. Apesar da forte influência da cultura original e de outras restrições práticas, a identidade é construída por um processo de escolhas que se repete continuamente, em que a razão, por meio do conhecimento e da atitude crítica, tem um papel fundamental.

* * *

A moderna economia do bem-estar combina (1) o comportamento auto-interessado como a única base de escolha econômica e (2) o julgamento da realização no âmbito social baseado no critério de utilidade. Com o desenvolvimento de uma tendência antiética a partir de 1930, na chamada “fase pós-utilitarista da economia do bem-estar”⁸¹, o critério de utilidade predominante, até então o utilitarista simples⁸², passou a ser o da “otimalidade de **Pareto**”⁸³, ou da “eficiência econômica”, que se refere “exclusivamente à eficiência no espaço das utilidades, deixando de lado as considerações distributivas relativas à utilidade”⁸⁴. Com isso, “pode-se dizer que o critério utilitarista e também o da eficiência de **Pareto** foram atrativos por não exigirem demais da imaginação ética do economista tradicional”⁸⁵.

Sen faz duas críticas à prática da economia do bem-estar de considerar a utilidade como única fonte de valor. A primeira é que “o êxito de uma

⁸¹ Sen, 1987, p. 65.

⁸² O *critério utilitarista simples* julga o êxito “segundo a magnitude da *soma total* de utilidade criada” (Sen, 1987, p. 46).

⁸³ “Considera-se que um determinado estado social atingiu um ótimo de Pareto se e somente se for impossível aumentar a utilidade de uma pessoa sem reduzir a utilidade de outra” (Sen, 1987, p. 47). Esse critério aceita, por exemplo, que em um estado ótimo haja “algumas pessoas na miséria extrema e outras nadando em luxo, desde que os miseráveis não possam melhorar suas condições sem reduzir o luxo dos ricos” (Sen, 1987, p. 48).

⁸⁴ Sen, 1987, p. 49.

⁸⁵ Idem, p. 66.

pessoa não pode ser julgado exclusivamente em termos de seu bem-estar”⁸⁶, pois ela pode atribuir valor a determinados objetivos não relacionados à melhora de seu próprio bem-estar. Haveria “uma ‘dualidade’ essencial e irreduzível” que, removida a camisa-de-força da motivação baseada apenas no auto-interesse, mostraria a pessoa

em termos de sua condição de agente [*agency*], reconhecendo e respeitando sua capacidade para estabelecer objetivos, comprometer-se [*commitments*], valores etc., e também (...) [mostraria] essa pessoa em termos de bem-estar [*well-being*], o que igualmente requer atenção.⁸⁷

Uma dificuldade dessa proposta, e que apareceria ao se dar importância à condição de agente de cada pessoa e ao que ela valoriza, é que ela teria uma concepção subjetivista da ética. A defesa de **Sen** é que a objetividade não estaria excluída: a importância dos valores individuais não seria atribuída incondicionalmente, mas avaliada segundo um procedimento objetivo; em especial, uma visão objetivista da ética pode incluir “entre os objetos valiosos a capacidade da pessoa para obter o que de fato ela valoriza”⁸⁸.

Se a utilidade fosse interpretada como “escolha”, sua relação com a condição de agente poderia ser estabelecida mais imediatamente, em detrimento da vinculação da utilidade com o bem-estar; mas, para dificultar essa interpretação, “a escolha é considerada importante na perspectiva utilitarista típica devido precisamente a sua alegada congruência com o bem-estar”⁸⁹.

A segunda crítica à economia de bem-estar questiona a utilidade como a condição que melhor representa o bem-estar pessoal, pois, julgada “exclu-

⁸⁶ Idem, p. 56.

⁸⁷ Idem, p. 57. As citações em inglês são originais da edição brasileira.

⁸⁸ Idem, p. 58.

⁸⁹ Idem, p. 60 n. 15.

sivamente pela métrica da felicidade ou satisfação de desejos”⁹⁰, sofreria a influência de circunstâncias contingentes, como as distorções que ocorrem nas avaliações feitas por pessoas que têm uma vida de infortúnios.

As críticas de **Sen** à atual concepção da economia do bem-estar se resumem em que “(1) o bem-estar não é a única coisa valiosa; (2) a utilidade não representa adequadamente o bem-estar”⁹¹. Mas há ainda outros pontos. Por exemplo, se o comportamento é afetado por considerações éticas⁹², então estas também são relevantes para as previsões sobre esse comportamento. Para **Sen**, a “sensação de invulnerabilidade à ética que parece permear a economia preditiva”⁹³ se deve às suposições estruturais sobre o comportamento, que somente são válidas se a economia ficar confinada ao “reduzido compartimento que proclama a adequação da otimalidade de **Pareto**”⁹⁴.

A suposição do comportamento auto-interessado também tem problemas em relação à liberdade de uma pessoa, que pode ser julgada importante em adição ao que essa pessoa realiza, pois “[a] liberdade pode ser valorizada não meramente porque auxilia a realização, mas também em razão de sua própria importância, extrapolando o valor do resultado alcançado”⁹⁵. Ao aplicar a perspectiva da liberdade ao bem-estar e à condição de agente, **Sen** distingue quatro categorias de informações sobre uma pessoa, que são:

⁹⁰ Idem, p. 61. Algumas páginas adiante, Sen cita John Rawls (1971): “Embora sentir-se feliz seja uma realização de grande importância, não é a única realização que importa para o bem-estar de uma pessoa” (Idem, p. 76).

⁹¹ Idem, p. 63.

⁹² Idem, pp. 67-8.

⁹³ Idem, p. 68.

⁹⁴ Idem, p. 69.

⁹⁵ Idem, p. 76.

‘realização de bem-estar’ (*well-being achievement*), ‘liberdade de bem-estar’ (*well-being freedom*), ‘realização da condição de agente’ (*agency achievement*) e ‘liberdade da condição de agente’ (*agency freedom*),⁹⁶

que foram reduzidas, na economia do bem-estar, a uma única categoria, a da “realização do bem-estar”. O que **Sen** critica nessa abordagem é que a liberdade seja considerada apenas pelo seu valor instrumental, e que a condição de agente, por se orientar exclusivamente ao bem-estar individual, seja depreciada do valor da condição de agente em si mesma.

Exatamente para cobrir essas categorias, **Sen** defende que se incluam nas avaliações éticas uma multiplicidade de condições valiosas, em oposição à “metodologia ‘monista’, que insiste na homogeneidade descritiva do que deve ser valorizado”⁹⁷. A consideração dessa diversidade tem a desvantagem de, às vezes, dificultar a “ordenação completa ponderada”⁹⁸ de todas as alternativas, nos casos em que houver um conflito insolúvel de princípios⁹⁹.

Uma *decisão difícil*, ilustrada pelo dilema de Agamenon¹⁰⁰, não é resolvida pelo ajustamento forçado das preferências antes da decisão. Mas, apesar de não ser bem vista pelos economistas nem pela maioria dos filósofos, a falta de uma ordenação completa não pode paralisar a tomada de decisão.

⁹⁶ Idem, p. 77. As citações em inglês são originais da tradução adotada.

⁹⁷ Idem, p. 78.

⁹⁸ Essa abordagem “requer que os conflitos sejam resolvidos antes de as decisões serem tomadas” (idem, p. 81), deixando em aberto exatamente a questão do que fazer se os conflitos não forem resolvidos.

⁹⁹ Barcan Marcus (p. 127) argumenta que escolher uma das opções de um dilema moral não implica em “apagar a obrigação original em relação à outra opção”. A solução de um dilema sempre deixa um resíduo que continua a ter uma “força dinâmica” no controle de nossas vidas e instituições, o que pode se manifestar como um “sentimento de culpa” (idem, p.136) e influenciar o comportamento futuro. Os dilemas vão ser tratados no item 2.1 deste trabalho.

¹⁰⁰ Sen, 1987, p. 82. O dilema de Agamenon, descrito por Ésquilo, é a decisão que o comandante grego teve que tomar para provocar uma mudança no vento que levasse os milhares de navios na direção de Tróia, ao preço do sacrifício da vida de sua própria filha, Ifigênia.

No âmbito das políticas públicas, por exemplo, há um forte apelo a favor da ordenação completa, pela necessidade de se ter avaliações inequívocas. Mas há também a exigência de que alguma decisão deva ser tomada, mesmo sem “uma *razão adequada* para escolher um caminho e não outro”¹⁰¹. Isso implica que, às vezes, as decisões sejam tomadas com base em justificação parcial, sem significar um afastamento da racionalidade¹⁰².

No âmbito das decisões pessoais, pesquisas empíricas sobre o comportamento revelam a ocorrência de “inconsistências sistemáticas na apreciação do risco e na avaliação comparativa de decisões alternativas”¹⁰³. Mesmo que prevaleça uma interpretação desses resultados como “simplesmente ‘erros’ de percepção ou raciocínio”, **Sen** considera “ser acertado dar lugar para afastamentos dos requisitos usuais de ‘racionalidade’ na compreensão do comportamento real”¹⁰⁴. Esses “erros” são também indícios de problemas na maneira como a decisão é tratada na concepção tradicional de racionalidade.

Existem outros comportamentos importantes que também podem ser considerados desvios em relação às características-padrão do agente auto-interessado¹⁰⁵. É de conhecimento comum que

o uso da escolha orientada para o próprio objetivo por parte de uma comunidade de pessoas com objetivos díspares pode levar a uma menor satisfação dos objetivos de cada uma delas do que se elas seguissem uma regra de comportamento diferente.¹⁰⁶

¹⁰¹ Idem, p. 83.

¹⁰² Sen (ibidem) cita o *asno de Buridan* (apud Jean Buridan, c.1295/1300-c.1360), que poderia, para não morrer de fome, ter comido qualquer um dos dois montes de feno à sua frente, mesmo se não tivesse uma razão para preferir um ao outro, ou seja, fazendo uma escolha que teria sido apenas parcialmente justificada.

¹⁰³ Sen, 1987, p. 85. Sen cita em especial os estudos de Kenney & Raiffa (1976) e Kahnemann, Slovik & Tversky (1982).

¹⁰⁴ Idem, p. 85.

¹⁰⁵ Idem, p. 96.

¹⁰⁶ Idem, p. 97.

O referido “comportamento diferente” – e mais cooperativo – levaria à maior satisfação dos objetivos de cada um, contrariando frontalmente a suposição básica da racionalidade que diz que a “escolha orientada para o próprio objetivo” maximiza o auto-interesse¹⁰⁷.

Sen já havia apresentado em *Rational Fools* ¹⁰⁸ uma razão por que a concepção de homem adotada em modelos econômicos tende a ser a de um egoísta interesseiro: “não importa o que uma pessoa fizer, ela pode ser vista como buscando o seu próprio interesse em todo ato de escolha” ¹⁰⁹. É difícil escapar desse reducionismo, pois se alguém fizer uma escolha, revela uma preferência, ou uma utilidade, maior; se, em um segundo momento, essa pessoa agir de forma oposta, um observador poderia dizer que ela estaria sendo inconsistente ou que suas preferências teriam mudado¹¹⁰.

Por trabalhar no campo da lógica, a concepção de racionalidade de **Ramsey** é neutra em relação à ética, aos fins e ao contexto – exatamente os pontos que **Sen** critica nos dois modelos de racionalidade dominantes na teoria econômica, quais sejam, o da maximização do auto-interesse e o da consistência interna como condição suficiente de racionalidade. Os dois casos são considerados restritivos por não contemplarem a multiplicidade de valores envolvidas na ação humana. Tais concepções deveriam, em resumo:

1. Admitir padrões de conduta alternativos que contemplem desvios da maximização do auto-interesse, ora vistos como irracionais. Implicaria re-

¹⁰⁷ Esse é um caso de *dilema do prisioneiro*, que vai ser caracterizado no Capítulo 2.

¹⁰⁸ Para Sen, a expressão ‘*rational fool*’ não se refere a uma pessoa autocentrada, mas a alguém que não tem a liberdade para distinguir entre questões flagrantemente diferentes, embora interligadas, e insiste em uma única resposta a essas questões (Sen, 2002, p. 7 n. 4).

¹⁰⁹ Sen, 1977, p. 88.

¹¹⁰ Idem, p. 89.

conhecer e valorizar a condição de agente das pessoas, incluindo sua liberdade (importante em si mesma, e não apenas no sentido instrumental), sua capacidade de estabelecer objetivos, bem como sua realização como pessoa.

2. Abrir espaço nas decisões para as considerações éticas relativas ao contexto, que afetam o comportamento das pessoas e que ora são vistos como meramente inconsistentes.

1.3 Blackburn e o domínio das paixões

O ponto de partida de **Blackburn** é exatamente onde paramos no item anterior: o da crítica ao conceito do *homo economicus*. Esse “ator racional” da teoria econômica e da ciência social e política seria¹¹¹ o agente básico da sociedade, que persegue objetivos que refletem seu auto-interesse. Sua ação resulta de um processo que envolve preferências consistentes e estáveis e, dentre as opções disponíveis e com conseqüências conhecidas, escolherá aquela de maior utilidade esperada. Tal paradigma, de que as pessoas racionais sempre agem no sentido de maximizar a utilidade esperada, seria indicador de uma característica humana universal ou representaria apenas uma imagem cultural, relativa ao individualismo do capitalismo ocidental?

Vamos acompanhar ao longo deste item a análise de **Blackburn** sobre três interpretações do princípio da maximização da utilidade esperada¹¹²:

1. O princípio estabelece uma verdade empírica, que produz previsões.
2. O princípio é uma verdade normativa, que produz recomendações.
3. O princípio é um modelo destinado a tratar matematicamente o processo de deliberação, impondo um modo de ver e interpretar as coisas.

Blackburn denuncia que existe uma tentação de se misturar as três interpretações, preservando-se algo empírico ou normativo, enquanto se defende o princípio como tautológico – é a chamada “falácia do economista”¹¹³, na qual os filósofos não estão inocentes. **Mill**, por exemplo, parte dos desejos

¹¹¹ Blackburn, 1998, p. 134. Citação de Monroe & Maher, “Psychology and Rational Actor Theory”, *Political Psychology*, 16 (1995), 1-21.

¹¹² Idem, p. 135.

¹¹³ Idem, p. 136. Falácias são erros de raciocínio que levam a uma conclusão falsa. O caso de John Stuart Mill citado por Blackburn parece que se enquadra como uma “falácia da afirmação do conseqüente”, em que dados $p \rightarrow q$ e q , conclui-se invalidamente que p . Assim, o raciocínio de Mill poderia ser colocado na seguinte forma equivalente: se um objeto é desejável, então é prazeroso; tal objeto é prazeroso; então é desejável.

empíricos das pessoas, passa pelo pensamento de que um objeto desejável é a mesma coisa que um objeto prazeroso, e conclui que é “impossível desejar alguma coisa a não ser na mesma proporção da idéia de seu prazer” ¹¹⁴.

Blackburn mostra que há um sentido em que isso é verdade:

Se eu desejar a morte dos infiéis, posso pensar que é agradável que os infiéis morram. Se eu desejar a sobrevivência das baleias, penso que é agradável que as baleias sobrevivam. Mas isso não é relevante para a questão empírica com que **Mill** começa, isto é, se as pessoas desejam apenas sua própria felicidade ou prazer. O fato de pensar na sobrevivência das baleias como agradável nem mesmo implica que penso que isso será agradável *para mim*. Poderia esperar estar morto antes que a sobrevivência das baleias ou a morte dos infiéis estivessem asseguradas. ¹¹⁵

Essa crítica se opõe à idéia de **Mill** de que “todo o mundo em geral está preocupado com o prazer de todo o mundo, ou com a felicidade geral” ¹¹⁶, porque todo o mundo, na verdade, está preocupado com o seu próprio prazer e assim *ninguém* está preocupado com o prazer de todo o mundo.

A primeira das três interpretações do princípio de maximização, a da verdade empírica que produz previsões, parece ser a mais difícil de contestar, pois pressupõe um “egoísmo psicológico” que seria uma dessas ervas daninhas que nunca morrem, embora os filósofos acreditem que tenham arrancado completamente suas raízes ¹¹⁷. **Blackburn** chama a atenção de que

¹¹⁴ Idem, p. 137. Citação de John Stuart Mill, *Utilitarianism*, cap. IV.

¹¹⁵ Idem, p. 137.

¹¹⁶ Blackburn, 2001, p. 88.

¹¹⁷ Uma tentativa importante, para Blackburn, foi a do bispo J. Butler (1692-1752), para quem “mesmo que exista um princípio de auto-interesse, tem que ser verdade que (...) visamos coisas externas particulares”. Contra a opinião dos egoístas psicológicos, que dizem que as coisas externas que desejamos são apenas meios para nossa felicidade, Butler mostra que as *coisas externas em si mesmas*, visadas por nossas paixões particulares, são distintas do *prazer que surge delas* – este seria um prazer meu, pois foi o meu desejo que foi satisfeito; mas não posso concluir que o princípio de minha ação seja sempre o amor-próprio. Outro ponto é que, quando um desejo é satisfeito, há uma mudança subjetiva de um estado de desejo para o de realização, e este seria o verdadeiro objeto de desejo. Ora, ‘desejar ter um desejo satisfeito’ é um desejo de segunda ordem, pois pressupõe um desejo antecedente e assim não pode ser meu principal objeto de interesse. Blackburn, 1998, p. 138. Citação de J. Butler, *Fifteen Sermons Preached at the Royal Chapel* (1726).

não há nem mesmo uma oposição direta entre o altruísmo e um princípio de amor-próprio, pois pensar assim seria o mesmo que

se pensar como em uma transação de caixa: se o meu próximo ganhar o dinheiro, eu não ganho; se eu ganhar, ele não ganha, e eu posso ter de escolher um resultado ou o outro. (...) Mas nem todos os relacionamentos são do tipo 'soma zero' (se eu ganhar, você perde, e vice-versa). Uma pessoa pode compartilhar bens dentro de um grupo mesmo quando os bens envolvidos forem escassos. ¹¹⁸

É uma contingência que as pessoas sejam ou não dessa maneira, pois umas podem ser mais benevolentes ou mais fechadas para relacionamentos que outras. Mesmo assim, está disseminada uma “permanente interpretação pessimista” da condição humana, que se desdobra em duas crenças:

Primeiro, por trás dos desejos que uma pessoa aparenta ter escondem-se as forças sombrias do inconsciente. Segundo, quando descrevemos o que uma pessoa realmente quer, acredita ou intenciona, sentimos que não temos ancoragens: quase tudo serve. ¹¹⁹

Na opinião de **Blackburn**, há argumentos que refutam essa visão pessimista¹²⁰, cuja inspiração é atribuída a **Freud** ¹²¹.

Mas o retorno mais recente ao *egoísmo psicológico*, segundo **Blackburn** é produzido pela síntese da teoria da evolução com a sociobiologia. De um lado, há uma forte tendência de se interpretar a teoria da evolução em termos de uma batalha cruel e egoísta pela sobrevivência; de outro, uma inter-

¹¹⁸ Blackburn, 1998, pp. 142-3.

¹¹⁹ Idem, p. 143.

¹²⁰ A melhor refutação ao pessimista seria, para Blackburn, o argumento que Butler usou contra a sugestão de Hobbes de que ter pena de alguém é o mesmo que imaginar o infortúnio acontecendo conosco no futuro. Primeiro, se fosse assim, compaixão e medo, frutos da mesma percepção, seriam idênticos. Segundo, a compaixão é considerada virtude, mas como seria se um homem compassivo, merecedor da nossa bondade, de repente se tornasse aflito? Terceiro, um amigo debaixo de um ônibus não é melhor sinal de perigo do que se fosse um desconhecido. Quarto, se uma pessoa estiver sob ameaça, qualquer ruído insignificante renovará esse medo, mas será ridículo falar desse som como um objeto de compaixão. (Idem, p. 144. Citação de Butler, *Fifteen Sermons*, Sermon V, 'Upon Compassion', 85).

¹²¹ Apesar de o alvo ser claramente a visão pessimista do homem, a referência a Freud pode ser questionada, pois nem toda motivação escondida seria atribuível ao inconsciente freudiano. Além disso, o inconsciente poderia ser, pelo menos como representação, tornado consciente pela terapia psicanalítica, por meio da linguagem (Roudinesco, p. 377).

pretação da teoria do gene egoísta, de **Dawkins**, de que “se nossos genes são ‘implacavelmente egoístas’, assim também devem ser os seres humanos”¹²². Daí a conclusão, criticada por **Blackburn**, de que, se quisermos desafiar os genes e nos livrar dessa tendência da natureza, devemos ensinar generosidade e altruísmo às crianças.

Blackburn descarta a motivação biológica para realizar tal esforço, pois, segundo o próprio **Dawkins**, a influência dos genes no comportamento humano é irrelevante; além do mais, “os genes não são literalmente egoístas implacáveis, não têm cérebro nem têm como representar escolhas a si mesmos. (...) Tudo o que podem fazer é se reproduzir, e algumas vezes essa reprodução sobreviverá, outras não”¹²³.

Esse esclarecimento seria suficiente para derrubar a “falácia do biólogo”, ou seja, a inferência que vai da premissa (a) à conclusão (b) abaixo:

- (a) um gene que dá aos membros de uma espécie uma característica *Y*, que é boa (...) para se reproduzir em um determinado ambiente, explica razoavelmente a evolução da espécie em que os membros têm tal característica; (...)
- (b) os seres humanos devem ser vistos como seres que perseguem, consciente ou inconscientemente, seu sucesso genético (...) se exibirem a característica *Y*.¹²⁴

¹²² Blackburn, 1998, p. 145. Referência a Dawkins, *The Selfish Gene*, 1989, pp. 200-201. Segundo Blackburn, o próprio autor não concorda com essa interpretação, e por isso “trava uma batalha contínua contra a implicação que o título do livro sugere” (ibidem).

¹²³ Idem, p. 146.

Na opinião de Hornsby, esse é um exemplo típico da mistura indevida de dois níveis de explicação: o pessoal, que faz referência a pessoas como um todo, e o subpessoal, que faz referência a partes de pessoas, como o cérebro ou os genes. Para ela, a filosofia contemporânea freqüentemente “se esquece de que as pessoas e seus atos são os sujeitos” (Hornsby, p. 157) e suprime a menção a pessoas em frases como “[as ações] são ‘causadas por crenças e desejos’” (ibidem), ou “‘duas crenças produzem uma terceira’, para dizer que uma pessoa chegou a uma conclusão” (ibidem). Assim, o problema estaria não na literalidade da frase, mas na passagem indevida de um nível de explicação a outro. Enquanto na “falácia do biólogo”, apresentada em seguida no texto, a passagem é do nível subpessoal para o pessoal, no título do livro de Dawkins o sentido da passagem é o inverso, pois atribui ao gene uma característica tipicamente humana.

¹²⁴ Blackburn, 1998, p. 148.

Entretanto, a linguagem comum reforça a falsa inferência pela repetição das duas proposições¹²⁵. **Blackburn** cita o que talvez seja a instância mais clara da falácia, a crença de que “a teoria genética ‘mostra’ que todos nós temos um desejo ardente e implacável de perpetuar nossos genes”¹²⁶, o que não é de todo verdadeiro em vista do celibato voluntário e dos casais que não querem ter filhos¹²⁷.

Um exemplo do egoísmo psicológico em uma forma renovada, segundo **Blackburn**, é a tese que relaciona desejos que podem ser considerados egoístas, tais como o desejo da própria felicidade ou da auto-afirmação, com uma extensão para outros desejos que, apesar de não serem voltados primariamente para nós mesmos, têm alguma relação conosco e por isso são “estimulantes de motivos egoístas”¹²⁸. Entretanto, todos os demais desejos voltados para os outros podem ser referidos a nós mesmos de uma maneira fraca, pois vemos ou ouvimos alguma coisa que se torna parte da nossa decisão, mas daí não se pode traçar um vínculo forte entre esses desejos e o

¹²⁵ A consideração dos dois níveis de explicação, conforme Hornsby (ver Nota 123), por si só eliminaria a falácia do biólogo, pois a primeira tese não pode ser suporte para a segunda. Entretanto, Blackburn apela para o reforço que acontece na linguagem comum, que parece ser da mesma natureza da “distração” apontada por Hornsby (p. 157) na filosofia contemporânea, quando esta se esquece de mencionar as pessoas e cria uma linguagem impessoal que esconde a distinção entre as descrições feitas em níveis diferentes.

¹²⁶ Blackburn, 1998, p. 150.

¹²⁷ Hornsby aponta ainda para outro problema além da mera falácia e da confusão dos níveis de explicação (ver Notas 123 e 125). Em uma concepção *impessoal* como essa, “uma ação seria um elo em uma cadeia causal que poderia ser vista sem se prestar atenção às pessoas, os elos sendo referidos ao funcionamento causal do mundo” (Hornsby, p. 129), o que subverteria a concepção que temos de nós mesmos como agentes livres (isto é, que têm poder causal), pois, “ao tratar ações como acontecimentos, e tratar qualquer acontecimento como ocorrência inevitável, expõe a liberdade como ilusão” (Hornsby, p. 241).

Na linguagem do monismo-anômalo de Davidson (Hornsby, p. 136), seria um ‘erro categorial’ atribuir à ação, que envolve o mental, uma descrição puramente física, que é do tipo nomológica, isto é, sujeita às leis da natureza.

¹²⁸ Blackburn, 1998, pp. 153-4. Referência a Broad, “Egoism as a Theory of Human Motives”, in *Broad’s Critical Essays in Moral Philosophy*, ed. E. D. Cheney, 248.

nosso interesse particular nem interpretá-los como se não fossem “realmente’ voltados” ¹²⁹ para os outros.

Dois exemplos do que foi descrito seriam a preocupação com o mundo que virá depois da nossa morte e a decisão de nos tornarmos mártires, em que agimos em função do que *nós* podemos causar no *nosso* mundo, sem que isso seja a revelação de um motivo egoísta que estaria escondido.

Mas ainda são possíveis outras versões do egoísmo psicológico. Uma delas é o caso da vingança: mesmo que não tenhamos qualquer ganho pessoal com o prejuízo daquele de quem queremos nos vingar, pode-se reconhecer algo egoísta nessa motivação – seria uma força estimulante do ego. Outra versão se mostra em um caso de amor entre duas pessoas ou de envolvimento de alguém em um movimento ecológico, em que é difícil calcular o peso do egoísmo na motivação quando, a partir de uma decepção ou perda, “ocorrerem mudanças na disposição” ¹³⁰ do agente supostamente desinteressado.

Blackburn acredita que, quando assistimos a uma peça, nosso relacionamento emocional com os personagens é devido a uma “humanidade comum” ¹³¹ que compartilhamos com eles. Essa capacidade em simpatizar nos foi ensinada quando crianças, por meio de treinamento de recompensa e punição, aproveitando nosso interesse egoísta de ganhar um e evitar o outro.

Até aqui acompanhamos a crítica de **Blackburn** à primeira versão do princípio de maximização da utilidade esperada, por meio do exame de diversos casos empíricos da tese do egoísmo psicológico. Retomando o que foi

¹²⁹ Idem, p. 155.

¹³⁰ Idem, p. 156.

¹³¹ Idem, p. 157.

mencionado no início deste item 1.3 e lembrado na Figura 1.4, as três versões sob análise são:

N°	Versão do princípio de maximização	Resultado	Pressuposto
(1)	Verdade empírica	Produz previsões	Egoísmo psicológico
(2)	Verdade normativa	Produz recomendações	Leva à felicidade
(3)	Modelo matemático para deliberação	Impõe um modo de ver	Tautologia

Figura 1.4 – Três interpretações do princípio de maximização da utilidade esperada ¹³²

A segunda versão do princípio é uma recomendação que leva à felicidade. O comentário de **Blackburn** é que o excesso de amor-próprio é mau conselheiro de nosso interesse próprio, pois nos conduz ao “paradoxo do hedonismo” ¹³³, pelo qual só seremos felizes se deixarmos de ficar interessados exclusivamente em nossa própria felicidade. Uma saída seria o meio-termo: há momentos em que se deve pensar mais nos outros e menos em nós mesmos, e há momentos que se deve fazer o contrário. E aqui a razão nada tem a dizer em um caso ou no outro: “a pessoa que trabalha pelo alívio dos famintos ou pela abolição da escravidão é, no mínimo, tão razoável quanto a pessoa que não trabalha por isso” ¹³⁴.

A terceira interpretação do princípio de maximização da utilidade esperada é um modelo destinado a tratar matematicamente o processo de decisão. O modelo não produz previsões relativas ao comportamento humano nem dá recomendações para a ação, apenas impõe um modo de ver – a teoria da escolha racional – pelo qual *invertemos* a leitura: “em vez de vermos um princípio específico, com a ‘utilidade’ sendo um tipo particular de alvo,

¹³² Figura baseada em Blackburn, 1998, p. 135.

¹³³ Blackburn, 1998, p. 158. Citação de Butler, *Fifteen Sermons*, Sermon IX, p. 171.

¹³⁴ Idem, p. 160.

vemos a utilidade ela mesma como um construto (...) resultante dos modos como [o agente] manipula seus interesses”¹³⁵.

Esta terceira interpretação é exatamente a mesma proposta por **Ramsey**, já descrita no item 1.1, em que um agente atribui valores, ou utilidades, às opções de um conjunto ordenado conforme suas preferências. Sabemos o que um agente prefere observando o que ele escolhe, mas, para **Blackburn**, temos ainda que saber interpretar as preferências desse agente, o que significa conhecer a situação em que ele está e as crenças que ele tem a respeito do que está fazendo. “Em particular, precisamos distinguir entre o que um agente *escolhe* e o que ele *prefere* quando as preferências se relacionam a aspectos de uma situação que estão fora de seu controle”¹³⁶. Juntando as duas pedras fundamentais da teoria da escolha racional, temos:

(*Util*) Uma função utilidade (...) pode ser definida sobre um conjunto de opções se as preferências satisfizerem duas condições de consistência: (...) a totalidade, e (...) a transitividade.¹³⁷

(*Revpref*) A escolha é o comportamento primitivo (...). Fazer escolhas é como estar equipado com uma relação de preferência em que uma escolha é preferível a outras, à luz daquilo que [se] acredita sobre a situação. (...)

A primeira parte dessa abordagem produz utilidades como ‘construções lógicas’ das preferências; a segunda produz preferências como construções lógicas das escolhas feitas, dadas as crenças.¹³⁸

Na opinião de **Blackburn**, as pessoas com preferências consistentes em um conjunto de opções deveriam ser chamadas de “qualificadas” e não de racionais, “como freqüentemente é feito, porque isso perpetua a ilusão de que estamos falando de um tipo especial de pessoa, ou dando *recomenda-*

¹³⁵ Idem, p. 162.

¹³⁶ Idem, p. 163.

¹³⁷ Em outro texto, Blackburn comenta que, nesse primeiro caso, “as preferências são estados psicológicos que antecedem nossas decisões” (Blackburn, 1995, p. 699).

¹³⁸ Blackburn, 1998, p. 164.

ções em nome da razão”¹³⁹. Mas, para **Blackburn**, é possível melhorar a apresentação da situação, com a substituição de *Revpref* por *Revconc*:

(*Revconc*) A escolha é o comportamento primitivo (...).Fazer escolhas é como estar equipado com uma relação de interesses (*concerns*), em que um interesse (*concern*) é mais bem atendido por uma escolha que por outras, à luz do que [se] acredita sobre a situação.¹⁴⁰

Revconc pretende ser um conceito mais amplo do que *Revpref* por incluir, como mencionado no início deste item 1.3, outras preocupações além da própria felicidade. Entretanto, um conceito suficientemente amplo de preferência faria com que *Revpref* e *Revconc* se tornassem iguais.

A partir desses conceitos, **Blackburn** propõe que o *homo economicus* não seja mais visto como um tipo especial de agente, “mas um agente qualquer, descrito em termos de interesses e utilidades”¹⁴¹. Tal agente se encaixaria perfeitamente no modelo de **Ramsey**, em que não há previsões empíricas, porque nada que o agente faça é inconsistente com ele, nem há recomendações, porque nada do que ele faça seria considerado como contrário à maximização da utilidade esperada. Dessa forma, um agente que ignore seu próprio bem-estar estaria igualmente maximizando a utilidade tanto quanto aquele que satisfaz sua vingança, ou mesmo aquele que perdoa.

Embora um sujeito obediente à lógica se mostre participante de um sistema normativo no qual seus hábitos inferenciais são formados, **Blackburn** defende que tal sujeito deveria conformar suas deliberações apenas “até certo ponto”¹⁴² na obediência às normas da teoria da utilidade esperada,

¹³⁹ Ibidem.

¹⁴⁰ Idem, p. 167.

¹⁴¹ Ibidem.

¹⁴² Idem, p. 168.

pois esta nada poderia dizer se uma pessoa fizesse algo diferente – seria irracionalidade ou mero desconhecimento de algum procedimento operacional?

Para esclarecer essa questão, **Blackburn** faz uma distinção entre uma situação empírica (ou um jogo empírico) e uma situação teórica (ou um jogo interpretado) ¹⁴³. No primeiro caso a situação é descrita em termos empíricos, mas não significa que essa descrição esteja necessariamente nos mesmos termos que um agente percebe a situação em que faz sua própria escolha – ele pode se preocupar com coisas diferentes daquelas que lhe poderiam ser atribuídas, sem qualquer restrição de número ou qualidade. Esse número indefinidamente grande de possibilidades tem a ver com a variedade de contextos em que as escolhas são feitas, dando margem ao aparecimento de intransitividade nas preferências, sem que isso revele irracionalidade.

No segundo caso, o da situação teórica, a interpretação é crítica, pois a intransitividade em um jogo é condenada como irracional. Para ilustrar a distinção entre as situações empírica e teórica, **Blackburn** exemplifica:

Adão e Eva (...) enfrentam uma escolha simples, X ou Y. Se eles escolherem X então eles ganham \$1 e o outro ganha \$0; e vice-versa se escolherem Y. Isso descreve a situação empírica, que é idêntica para cada um deles. Agora suponha que Adão escolha X. Essa escolha mostra o que ele prefere, e mostra qual é a sua utilidade esperada: ela é maior se ele ganhar \$1 do que se Eva ganhar. Suponha que Eva escolha Y. Isso mostra o que ela prefere, e mostra que a sua utilidade esperada é maior se Adão ganhar um dólar do que se ela ganhar. Pode ser que ela seja altruísta, ou espere algum benefício de Adão em troca, ou odeie dinheiro, ou queira sabotar o experimntador, ou o que quer que seja. ¹⁴⁴

O significado disso é que os jogadores estão em jogos teóricos diferentes, e o modelo de ação de um não serve como recomendação para o outro.

Um exemplo um pouco mais complexo é o caso da chantagem:

¹⁴³ Idem, pp. 168-9.

¹⁴⁴ Idem, p. 170.

Antes de começar o jogo, Eva cometeu um deslize. Se Adão não fizer nada, ele mantém \$1 e Eva \$2. Se ele a chantagear e ela aceitar, ele toma \$1 de Eva. Mas, se ela não se submeter e der com a língua nos dentes (*blows the gaff on him*), revela-o como chantagista, mas também revela seu próprio deslize, ao preço de tudo o que eles têm, deixando ambos com \$0.¹⁴⁵

Nessa situação empírica não há uma resposta única, pois depende do jogo teórico que cada um percebe e de como está disposto a jogar¹⁴⁶:

Suponha, primeiro, que, para cada um deles, a utilidade esperada (...) simplesmente coincida com os seus próprios dólares. Isso significa que eles não se preocupam com nada, exceto com a posse de seus próprios dólares. Suponha, também, que cada um saiba isso a respeito do outro. Então Adão sabe o que acontecerá na vez de Eva: ela jogará “pomba”, para manter \$1, em vez de jogar “falcão” e ficar com \$0. E, sabendo disso, Adão jogará “falcão”, porque, dado que sua preocupação maior é com os dólares, seria uma contradição (...) se ele ficasse com menos utilidades, isto é, com menos dólares, quando por uma simples escolha ele pudesse ficar com mais. Assim, ele jogará “falcão”, Eva se submeterá, e Adão tomará um dos dólares de Eva. Pobre Eva.

O mesmo caso empírico, mas com interpretação teórica alterada devindo a um novo contexto – a educação –, tem desta vez um desfecho diferente:

Suponha agora que Eva frequentou uma boa escola onde lhe ensinaram a ser orgulhosa e vingativa. E suponha que Adão saiba disso. Então Adão sabe que a situação *teórica* de Eva não é representada pelos dólares envolvidos. Submeter-se e deixar Adão se sair bem pode ser impensável para Eva. Sua preferência, e conseqüentemente sua utilidade, seria dar com a língua nos dentes (*to blow the gaff*), mesmo que isso significasse provocar sua própria ruína financeira.¹⁴⁷

Blackburn considera que na vida real podemos conhecer um jogo empírico, mas é improvável que conheçamos o jogo teórico, pois este só se revelará quando as ações e as preferências se mostrarem. Assim, um especialista em teoria dos jogos pode somente sugerir que se adote um modelo de comportamento em um dado jogo, mas nunca recomendar que, em um jogo empírico, se deva jogar desta ou daquela determinada maneira.

* * *

¹⁴⁵ Idem, p. 171.

¹⁴⁶ Conforme a menor ou maior agressividade da ação escolhida, as opções desse jogo são chamadas de “pomba” e “falcão”.

¹⁴⁷ Idem, pp. 171-2.

Vimos ao longo desta análise que **Blackburn** tem uma visão reduzida do poder da razão na ação humana, ora mostrando que a razão não tem nada a dizer no domínio da preferência e da escolha, ora negando à razão autoridade para fazer recomendações. Além disso, segue **Hume** ao dizer que

a esfera própria da razão está confinada à matemática e à lógica, enquanto o conhecimento a respeito de como as coisas são se deve exclusivamente à experiência dos sentidos. [A Razão] não nos dá qualquer princípio substantivo de conduta.¹⁴⁸

Essa “Razão” (com letra maiúscula) seria algo que “teria uma influência necessária, ou o que os filósofos algumas vezes chamam de força ‘apodíctica’”¹⁴⁹, para comandar os agentes racionais. Ao negá-la, **Blackburn** sustenta que no domínio da preferência e da escolha, ou seja, da ética, “o coração, ou o que **Hume** chamou de paixão ou sentimento, governa tudo”¹⁵⁰. O que as paixões pedem nós fazemos, com a ajuda da razão e da experiência; talvez haja limites formais em nossas preferências, como considerar “irracionais” certas escolhas inconsistentes,

mas não há restrições substantivas em nossas paixões que sejam impostas exclusivamente pela razão. Isso pode ser posto em termos de um contraste entre *descrição* e *prescrição*. A razão está envolvida em obter descrições corretas do mundo. O que então prescrevemos está além da sua jurisdição.¹⁵¹

Em *Practical Tortoise Raising*, **Blackburn** reforça seu ponto de vista: uma decisão não depende apenas de uma combinação de fatos e da razão, mas de algo mais que está sempre presente – as paixões. Mas reconhece que

a conhecida afirmação de Hume, no *Tratado*, de que ‘a razão é a escrava das paixões e não pode aspirar a nenhuma outra função exceto servi-las e obedecê-las’ é o reverso deliberado da imagem platônica da razão (o cocheiro) que domina as muito desreguladas paixões (os cavalos).¹⁵²

¹⁴⁸ Blackburn, 2001, pp. 110-1.

¹⁴⁹ Idem, p. 109.

¹⁵⁰ Idem, p. 111.

¹⁵¹ Idem, pp. 111-2.

¹⁵² Blackburn, 1997, pp. 332-3. Verbete “*racionalidade*”.

1.4 **É possível conciliar a discussão?**

Nos três autores escolhidos para ilustrar a discussão contemporânea do conceito de racionalidade, percorremos um caminho que vai de uma visão tradicional, passamos pela contestação de algum aspecto do conceito, e chegamos à algumas proposições que abrem espaços novos, antes interditados.

Ao admitirmos, com **Ramsey**, a possibilidade de crenças parciais, a consistência interna de um sistema de crenças, definida como inexistência de contradição entre seus elementos, ganha uma condição a mais: a de que elas devem obedecer às leis da probabilidade. Essa *coerência*, como mais recentemente passou a ser chamada, possibilita a convivência de crenças que de outra forma seriam tomadas como contraditórias, pois estariam a salvo graças aos graus parciais de crença.

A valorização da condição de agente presente no homem, defendida por **Sen**, e que o vê como capaz de estabelecer objetivos e valores, expande o modelo de maximização do auto-interesse, adotado pela moderna teoria econômica. No novo modelo, a liberdade não se reduziria a uma função instrumental, mas seria integrada à escolha dos fins. Assim, as decisões públicas poderiam aceitar justificações parciais, sem que isso negasse a racionalidade da escolha; no campo pessoal, um comportamento cooperativo daria maior satisfação às pessoas, contrariando a suposição tradicional.

A concepção de **Blackburn**, estruturalmente idêntica à de **Ramsey**, aceita que a escolha seja o comportamento primitivo a ser observado, mas nega autoridade à razão, ao retirar o caráter normativo do *homo economicus*, pois não seria mais possível fazer previsões de seu comportamento ou fazer

recomendações em nome da razão. Por um caminho diferente, essa concepção também se aproxima da proposta de **Sen**, ao defender um maior cuidado na avaliação do contexto em que as preferências e as escolhas acontecem.

O alargamento do conceito de racionalidade, propiciado tanto pela coerência definida em termos de probabilidades subjetivas, quanto pela valorização da condição de agente do homem e por um modelo sensível às suas emoções, abre a possibilidade do erro e da criatividade, e nos fornece modos de tentar justificar e compreender o conflito e o dilema.

Essa abertura para uma explicação mais ampla, ao mesmo tempo em que torna possível uma conciliação entre pressupostos e princípios diferentes, exige uma certa tolerância com a inconsistência e nos dá motivos para descrever a razão como algo paradoxal, ou que a razão, em si mesma, abriga paradoxos.

No próximo capítulo, tentaremos avançar a discussão na direção de um produto típico do século XX: a teoria dos jogos.

CAPÍTULO 2 – DOS PARADOXOS À TEORIA DOS JOGOS

- 2.1 Paradoxos e dilemas*
- 2.2 Uma introdução à história da teoria dos jogos*
- 2.3 Classificação e representação dos jogos*
- 2.4 Um recorte na teoria para examinar o dilema do prisioneiro*
- 2.5 Caracterização do dilema do prisioneiro*
- 2.6 O dilema do prisioneiro e o conceito de racionalidade*

A discussão do conceito de racionalidade, iniciada no capítulo anterior pela apresentação de três diferentes concepções, vai agora ser levada ao contexto dos paradoxos e da teoria dos jogos. Veremos que um tipo especial de paradoxo, o dilema do prisioneiro, examinado à luz da teoria dos jogos, é capaz de levantar questões filosóficas no campo em que racionalidade e valores se tangenciam. O dilema do prisioneiro representa um tipo de situação em que há interesses em parte comuns e em parte conflitantes – os primeiros levam a cooperar, enquanto os segundos levam a não cooperar. A teoria dos jogos, ao buscar soluções, supõe que um agente racional, dotado de um conjunto coerente de preferências, escolhe a ação que lhe dará a máxima utilidade. Mas o seu oponente, igualmente racional, faz com que as escolhas sejam estratégicas, isto é, que dependam da expectativa do que o outro vai fazer. Assim, aquela ação que é a melhor individualmente, produz um resultado que é pior para todos – esse é um dilema que envolve cotidianamente não só pessoas, mas entidades, como empresas e nações.

2.1 Paradoxos e dilemas

Examinamos no capítulo anterior alguns aspectos do conceito de racionalidade, e em quase todos tivemos a presença de um conceito mais básico, o da consistência. Inicialmente normativa e axiomática, chegamos ao final da discussão a certa flexibilização da consistência, um alargamento¹ que permite à racionalidade abrigar conflitos e perplexidades.

Uma classe especial desses “desvios” é a dos paradoxos², em que, pela definição de **Sainsbury**, “uma conclusão aparentemente inaceitável [é] obtida de premissas aparentemente aceitáveis, por meio de um raciocínio aparentemente aceitável”³. Haveria então nesses casos uma “camuflagem” encobrindo a realidade, e esta se revelaria apenas em aparências – tanto das evidências que conduzem às proposições, quanto dos raciocínios lógicos que vão relacionar essas proposições. Esse jogo enganador nos desafia a fazer uma escolha: “ou a conclusão não é totalmente inaceitável, ou há algum erro que não é óbvio no ponto de partida ou no raciocínio”⁴.

¹ A linha de pesquisa adotada neste trabalho é a de buscar um conceito mais amplo da racionalidade, ao invés de destacar um papel importante para a irracionalidade na ação humana. Essa opção permite, por exemplo, que certas ações, vistas como irracionais por um observador, possam ser consideradas como fruto de uma deliberação racional, visando determinado fim, como mencionado no Capítulo 1.

² O termo *paradoxo* vem do grego *para* (além) e *doxa* (opinião), e tem como sinônimos os “argumentos” (*logoi*), as falácias (*paralogismoî*), os sofismas e as aporias, todos eles termos usados por Platão e Aristóteles, bem como os paralogismos e as antinomias, termos usados por Kant (Rescher, p. 3).

Para Poundstone, a falácia é o tipo mais simples de paradoxo, pois se trata de “uma contradição que vem de um erro trivial, mas bem camuflado, de raciocínio. (...) Em uma falácia, o paradoxo é uma ilusão” (Poundstone, 1988, p. 16). Ver Nota 113, Capítulo 1.

³ Sainsbury, p. 1.

⁴ Ibidem.

Em outra definição proposta por **Rescher**, um paradoxo seria constituído por proposições que são individualmente plausíveis⁵, mas que formam um conjunto logicamente inconsistente. Nesse caso, “premissas plausíveis acarretam uma conclusão cuja negação é também plausível”⁶.

Enquanto a definição de **Sainsbury** deixa a solução dos paradoxos para cada caso, **Rescher** propõe uma teoria, a *aporética*, na tentativa de estabelecer um método geral de solução. Ele se baseia na suposição de que nos paradoxos “consideramos como plausível mais do que o reino dos fatos e da realidade é capaz de acomodar”⁷, resultando uma situação inconsistente. Resolver um paradoxo seria, em geral, restaurar tal consistência, exigindo que algumas teses fossem abandonadas, de preferência aquelas que mostrassem uma plausibilidade mais fraca. Essa busca seria facilitada pela existência de um tropismo a favor da alternativa mais plausível, que é “um instrumento de prudência epistêmica”⁸. A aplicação desse critério não significa, entretanto, que as teses abandonadas devam ser rejeitadas como falsas, pois, em virtude de sua plausibilidade, elas poderão ainda ter utilidade cognitiva em algum outro contexto⁹. Além disso, o reconhecimento da falsidade de uma proposição pode exigir um conhecimento técnico acima do comum¹⁰.

⁵ Uma proposição é considerada plausível “se tiver alguma chance de ser verdadeira” (Rescher, p. 16). Por isso, ela será aceita provisoriamente, enquanto as atuais circunstâncias perdurarem. A plausibilidade de uma proposição não é uma questão de sim ou não, mas de mais ou menos, pois depende da “força epistêmica” a que está submetida.

⁶ Idem, p. 6.

⁷ Idem, p. 9.

⁸ Idem, p. 19. Essa abordagem é conhecida na teoria da decisão como *estratégia minimax da perda potencial*.

⁹ Idem, p. 37.

¹⁰ Idem, pp. 57 ss. Ver o texto que faz referência à Nota 8, Capítulo 1, que trata da diferença entre crença e aceitação.

O tipo especial de paradoxo que nos interessa neste trabalho é o *dilema*, em que, sabendo-se que P e Q violam algum princípio,

- dadas as premissas (se C , então P ; e se não- C , então Q)
- e uma disjunção inevitável (ou C ou não- C),
- chega-se a um resultado inaceitável (ou P ou Q).¹¹

Um caso particular é o dilema moral. Neste, qualquer que seja a escolha ou a ação inevitável que a pessoa fizer (ou C ou não- C), resultará a violação de um certo princípio moral que também é obrigatório (ou P ou Q)¹².

Contrária à visão do dilema como inconsistência, conforme se vê nas definições de **Sainsbury** e **Rescher** acima, **Barcan Marcus** defende a existência de dilemas mesmo quando não houver inconsistência entre os princípios que estiverem em jogo¹³. Para **Barcan Marcus**, a consistência de princípios morais não acarreta que, ao se resolver um dilema pela escolha de um desses princípios, as obrigações morais com respeito ao outro se apaguem – trata-se do *resíduo moral*¹⁴. Ao contrário, o reconhecimento da realidade dos dilemas “tem uma força dinâmica (...) [que] nos motiva a arranjar nossas vidas e instituições de modo a evitar tais conflitos”¹⁵.

As teorias que negam a realidade dos dilemas e os tomam apenas como conflitos aparentes se caracterizam por um formalismo ético¹⁶ ou por um intuicionismo moral¹⁷. A primeira característica considera que, se as regras de um código moral forem corretamente ampliadas ou corrigidas, a inconsis-

¹¹ Idem, p. 34, e Honderich, verbete ‘*dilemma*’, assinado por Ruth Barcan Marcus.

¹² Blackburn, 1994, p. 102.

¹³ Barcan Marcus, p. 127.

¹⁴ O *resíduo moral* foi citado anteriormente na Nota 99, Capítulo 1.

¹⁵ Barcan Marcus, p. 127.

¹⁶ Idem, p. 129.

¹⁷ Ibidem.

tência em um dilema poderá ser eliminada em qualquer circunstância¹⁸. A segunda característica pressupõe que “há sempre uma escolha moralmente correta entre as opções em conflito”, e que essa escolha “é uma questão de intuição”¹⁹. A principal crítica de **Barcan Marcus** é que essas teorias buscam a solução do conflito em um princípio único, tal como o imperativo categórico de **Kant** ou o princípio de utilidade, o que é um erro, pois de um mesmo princípio sempre podem derivar opções moralmente equivalentes e produzir “um análogo ao asno de **Buridan**”²⁰.

A consistência é “somente uma condição necessária, mas não suficiente, para um conjunto de regras morais”²¹. Se quisermos agir em todas as circunstâncias de acordo com cada uma das nossas regras, nossa alternativa como agentes morais não será distribuir as cartas ao acaso, como mandam as regras dos jogos de baralho, mas “tentar trapacear no jogo (*to stack the deck*) tal que os dilemas não apareçam”²². Fazer a melhor escolha possível em uma situação difícil não significa apagar retroativamente uma das obrigações conflitantes, nem dizer que é um erro sentir culpa ou remorso por não ter agido de acordo com essa obrigação. Para **Barcan Marcus**, todos os “os sentimentos de culpa a respeito da alternativa rejeitada devem ser levados em consideração na conduta futura das nossas vidas. Se quisermos evitar dilemas, devemos estar motivados a agir assim”²³, isto é,

¹⁸ Essa correção poderia ser feita pela adoção de um princípio mais geral, pela aplicação de regras de prioridade ou pelo uso de pesos para ordenar os princípios (ibidem).

¹⁹ Idem, p. 130.

²⁰ Idem, p. 131.

²¹ Idem, p. 135.

²² Ibidem.

²³ Ibidem.

devemos agir de modo que, se se deve fazer x e se deve fazer y , então podemos fazer ambos x e y . (...) Não há razão para supor (...) que podemos (...) ter sucesso em prevenir e evitar tal conflito. Não é uma mera falha da vontade, ou falha da razão, que impede as máximas morais de se tornarem leis universais. São as contingências do mundo. ²⁴

Quando um agente decide entre alternativas conflitantes, não significa ter ele descoberto um princípio de prioridade para ordenar suas ações, pois “não há razões morais para generalizar sua própria escolha” ²⁵, mas sim que escolheu “o tipo de pessoa que gostaria de ser e o tipo de vida que gostaria de levar” ²⁶. **Barcan Marcus** conclui que “nem todas as questões de valor são questões morais, e pode ser que nem todos os dilemas morais sejam resolvíveis por princípios para os quais uma justificativa *moral* possa ser dada” ²⁷.

O tema dos valores está na origem dos conflitos entre as diferentes razões que são apresentadas para a tomada de decisão entre cursos alternativos de ação. A questão central desses paradoxos da racionalidade é determinar qual é a melhor ação que deve ser feita, dadas as circunstâncias.

Anteriormente, neste trabalho, a discussão entre **Ramsey**, **Sen** e **Blackburn** já nos trouxe até este ponto em que estão em jogo questões de valor e sua associação com a racionalidade. Vamos examinar, neste e nos próximos capítulos, dois dos paradoxos que tratam desses temas: o *dilema do prisioneiro* e o *problema de Newcomb*.

²⁴ Idem, p. 140.

²⁵ Idem, p. 141.

²⁶ Ibidem.

²⁷ Ibidem.

2.2 Uma introdução à história da teoria dos jogos

Antes de ir para os EUA em 1938, fugindo da perseguição nazista, o matemático húngaro John **von Neumann**²⁸ (1903-1957) tinha desenvolvido o teorema *minimax*, no qual demonstrava que em jogos de dois jogadores há sempre um curso racional de ação (um jeito ótimo de jogar), desde que os interesses dos jogadores sejam completamente opostos. Em 1944, juntamente com o economista Oskar **Morgenstern**, publicou *Theory of Games and Economic Behavior*, obra clássica da fundamentação matemática da economia, que lançou a teoria dos jogos ou da decisão estratégica²⁹. O termo “jogo” indicava “uma situação de conflito onde alguém tem que fazer uma escolha, sabendo que outros estão fazendo escolhas também, e o resultado do conflito será determinado de alguma maneira por todas as escolhas feitas”³⁰.

A corrida armamentista entre duas potências atômicas criou, na década de 1950, a possibilidade de uma nação destruir a outra, e cada uma delas poderia ser a vítima desse ataque de surpresa. A tentação de uma guerra preventiva justificava, na política americana, a defesa de uma “agressão para a paz”, cujo objetivo seria retomar o monopólio perdido em 1949, quando a União Soviética explodiu sua primeira bomba atômica. Logo depois da segunda grande guerra, o esforço estratégico americano já contava com a *Rand*

²⁸ Seguimos o roteiro histórico apresentado por Poundstone, 1992, cap. 1, esp. pp. 1-9. Na Nota 3, Capítulo 1 deste trabalho, foi mencionado que antes de von Neumann e Morgenstern houve um desenvolvimento paralelo dessa teoria por Ramsey (1926) e por De Finetti (1934), sem que qualquer deles conhecesse o trabalho do outro.

²⁹ A teoria dos jogos é a parte da teoria da escolha racional que trata das decisões estratégicas, nas quais as decisões de todos os envolvidos são interdependentes. No Capítulo 3 vamos examinar as decisões em situações não-estratégicas.

³⁰ Poundstone, 1992, p. 6.

Corporation, formada por mais de 500 cientistas³¹, entre eles **von Neumann**, para desenvolver aplicações militares da teoria dos jogos.

Em 1950, Merrill **Flood** e Melvin **Dresher**, cientistas da *Rand*, “inventaram um ‘jogo’ simples, mas enganoso, que desafiava parte das bases teóricas da teoria dos jogos”³². Logo depois, Albert W. **Tucker**, precisando exemplificar uma situação de conflito, deu a esse jogo o nome de *dilema do prisioneiro* que, a partir daí e ao longo dos últimos 50 anos, se constituiu em uma referência para estudos em vários campos da ciência e da filosofia.

Uma versão do dilema do prisioneiro tem o seguinte enunciado:

Um delegado oferece a dois prisioneiros que aguardam julgamento as seguintes opções. (A situação é simétrica para os prisioneiros; eles não podem se comunicar para coordenar as ações em resposta à proposta do delegado ou, se puderem, eles não têm nenhum meio para forçar qualquer acordo que possam desejar). Se um prisioneiro confessar e o outro não, o primeiro é liberado e o segundo recebe uma pena de 12 anos de prisão; se ambos confessarem, cada um recebe uma pena de 10 anos de prisão; se nenhum confessar, cada um recebe uma sentença de 2 anos.³³

O fato paradoxal nesse dilema é que a decisão racional que cada prisioneiro tende a tomar – a de confessar, que é a melhor para cada um individualmente, independentemente da ação presumida do outro – elimina a possibilidade de obterem uma sentença mais favorável para ambos, se escolhessem a opção de não confessar. Esse fato não contradiz qualquer pressuposto da ação racional, mas explicita que nem sempre um cálculo individualmente racional produz um resultado coletivamente desejado³⁴.

³¹ Idem, p. 92.

³² Idem, p. 8.

³³ Nozick, 1993, p. 50.

³⁴ O aspecto teórico crucial do dilema do prisioneiro, como vai ser visto no item 2.4 (primeiro modelo), é que, apesar de haver um equilíbrio de Nash no ponto em que os dois prisioneiros confessam, essa solução não é um ótimo de Pareto.

O interesse no dilema do prisioneiro é que ele representa situações que envolvem, ao mesmo tempo, o conflito de interesses e a possibilidade de cooperação³⁵ entre duas ou mais partes, que podem ser pessoas, empresas ou nações. São exemplos o conflito de poder entre um homem e uma mulher em um casamento, a disputa de mercado entre empresas, e a corrida armamentista entre nações. Além desses casos que representam a *ação individual* de poucos elementos, sejam pessoas ou instituições, há ainda os casos de *ação coletiva* que interessam de perto às ciências sociais, dentre eles as interações no âmbito de um grupo numeroso de agentes, como em um racionamento de água, no pagamento de impostos, no compartilhamento de recursos públicos e na utilização do meio ambiente.

Em situações como essas há um campo de opções à disposição dos agentes, caracterizado pela sobreposição de interesses convergentes e divergentes, que faz com que a decisão de cada um seja contingente às decisões que os outros tomarem³⁶. A escolha de cada parte não depende apenas dos próprios desejos e crenças nem traz conseqüências tão-só para si própria, mas deve ser considerada em conjunto com os desejos e crenças das outras partes, que também vivem a mesma situação e que são, da mesma forma, agentes intencionais e sabem que não têm um controle absoluto das conse-

³⁵ No exemplo do dilema do prisioneiro, a cooperação se dá entre os prisioneiros, e não entre estes e as autoridades. Cooperar, nesse caso, significa *não confessar*.

³⁶ Apesar de haver uma interdependência estrutural na escolha do que cada um vai fazer, algumas dessas opções são claramente dominantes, como se verá no item 2.4, por darem sempre resultados melhores que outras, independentemente da escolha que o outro fizer. É o caso já citado de escolher “confessar” em um dilema do prisioneiro.

qüências de sua decisão. Essa interdependência³⁷ das escolhas, que não é exclusiva do dilema do prisioneiro, forma um jogo estratégico.

Apesar de a história da teoria dos jogos ter começado recentemente, a questão do interesse pessoal em conflito com o bem comum já havia aparecido em **Platão**, *A República*, em que Gláucon defende a tese de que “ninguém é justo por sua vontade, mas constrangido”³⁸. Mas as referências mais diretas ao dilema do prisioneiro são encontradas em **Hobbes** e **Hume**.

Na teoria política de **Hobbes**, o Estado é a força que leva os homens a cooperar na realização do bem comum, em detrimento do interesse particular. Sem essa coação, os homens viveriam no estado da natureza, o de guerra de todos contra todos em que “cada homem não busca apenas maximizar o seu resultado, mas também a sua eminência, [que é] a diferença entre o seu resultado e o das outras pessoas”³⁹.

A teoria de **Hume**⁴⁰ não exige a presença em todos os casos de um agente coator para a possibilidade da cooperação, pois se houvesse necessidade de uma força que obrigasse sempre a ação cooperativa, não haveria o dilema. **Hume** distingue dois casos que têm características diferentes:

³⁷ A dependência da ação humana ao contexto foi analisada por Sen, como exposto no Capítulo 1. Mesmo não existindo uma definição prévia do que seja o contexto, é importante lembrar algumas características do ambiente em que acontecem as ações humanas: (a) elas envolvem emoções, expectativas e simbologias; (b) apesar de conter elementos naturais, como os mecanismos subpessoais e os fatos biológicos, o contexto da ação é parcialmente criado pelo homem, pois inclui a cultura e as possibilidades de manipulação (c) a definição de um contexto é fruto de interpretações dos agentes, dos interessados e dos observadores; (d) o contexto não é inteiramente dependente do acaso.

A relação ação-contexto é reforçada por Morton, para quem “em uma situação estratégica as ações dos agentes que interagem são suscetíveis a alguns fatores do ambiente, os quais são eles mesmos afetados pelas ações dos agentes” (Morton, p 4). Essa caracterização está presente mais à frente no item 2.3, no texto que faz referência à Nota 61 e seguintes.

³⁸ Platão, *A República*, Livro II, 359d-360c, história do anel de Gíges.

³⁹ Taylor, pp. 1 e 32-3.

⁴⁰ Conforme analisada por Taylor, p. 15.

Dois vizinhos podem concordar em drenar um pântano que possuem em comum; porque é fácil para eles saber o que cada um tem em mente; e cada um tem de perceber que a consequência imediata de falhar em sua parte é o abandono do projeto inteiro. Mas é muito difícil, e realmente impossível, que mil pessoas possam concordar em alguma ação desse tipo; sendo difícil para elas coordenar tão complicado acerto, e ainda mais difícil executar isso; enquanto cada uma busca um pretexto para se livrar da dificuldade e despesa, e pôr o fardo inteiro nas outras.⁴¹

No primeiro caso, nenhum dos dois vizinhos é capaz de realizar todo o trabalho sozinho, configurando uma situação que é diferente daquela do dilema do prisioneiro⁴². Nesse caso, em que estão envolvidos apenas dois vizinhos, a cooperação é voluntária, pois o benefício da cooperação é maior do que o da não-cooperação e assim não há um conflito entre a racionalidade e a cooperação.

No segundo caso (mil vizinhos), a omissão de alguns pode não ser suficiente para impedir a realização do trabalho pelos outros vizinhos, o que faz surgir a tentação racional a cruzar os braços, que é o conflito entre a racionalidade e a cooperação, característica do dilema do prisioneiro. Para reprimir a possibilidade do *carona*, há necessidade, nesse caso, de coação, porque “os homens [nas grandes sociedades] não cooperarão voluntariamente (...), eles são jogadores de um superjogo do dilema do prisioneiro e suas taxas de desconto são bastante altas”⁴³. Nessa situação, há uma preferência pelo ganho individual sobre o coletivo e pelas vantagens imediatas sobre as de longo prazo.

⁴¹ Hume, 3.2.7.8.

⁴² No item 2.5, no texto que faz referência às Notas 136 e 137, o caso da drenagem do pântano por apenas dois vizinhos será apresentado como um exemplo do *jogo da coordenação*.

⁴³ Taylor, p. 159. Ver Notas 100-103 sobre o conceito de “desconto”.

Um superjogo é um jogo repetido, ao qual se aplica o “*folk theorem*”, que será caracterizado no item 2.4, no texto que faz referência à Nota 106.

Skyrms (2004, p. 5) mostra que um dilema do prisioneiro repetido se transforma, a partir do segundo estágio, em um jogo da coordenação. Ver item 2.5, referência à Nota 141.

A mesma preferência provoca a “tragédia das terras comuns”⁴⁴, que descreve o compartilhamento de um pasto pelos rebanhos dos membros de uma comunidade. Se todos agirem para maximizar os seus próprios ganhos, mais cabeças serão adicionadas até que a capacidade do pasto se esgote. A partir daí, cada nova cabeça dará benefícios apenas ao seu proprietário, enquanto o custo será pago por todos da comunidade. Esse mesmo processo de degradação também ocorre, em larga escala, nos oceanos, lagos e rios.

O ingrediente principal do dilema do prisioneiro é a tentação que cada jogador tem de satisfazer o interesse próprio, o que levaria a um resultado ruinoso se *todos* decidissem fazer o mesmo. O exemplo mais comum de dilema do prisioneiro em nossa vida diária é o do *carona*, que se aproveita do esforço de muitos – tal pessoa não paga uma taxa coletiva, ou não economiza durante um racionamento.

⁴⁴ Taylor, p. 21, citação de texto de Garrett Hardin. Este, por sua vez, cita Whitehead (1948, p. 17), que destaca que a essência da tragédia não é a infelicidade, mas “a solenidade do funcionamento sem remorso das coisas”, ou a inevitabilidade do destino, como tematizado nas tragédias gregas. Taylor acrescenta que “a tragédia reside no fato que a ação ‘racional’ de cada indivíduo produz um estado de coisas que ninguém quer” (Taylor, p. 21).

2.3 Classificação e representação dos jogos

Antes de fazer uma análise das formas de representação dos jogos, vamos examinar como os jogos são classificados.

Os jogos se dividem em jogos de soma zero e em jogos de soma não-zero. Nos de soma zero, o que um jogador ganha é exatamente o mesmo que o outro perde, e nesse caso se pode dizer que os jogadores têm interesses completamente opostos. Nos jogos de soma não-zero, ou de soma variável, há ganhos e perdas diferenciados, pois os interesses dos jogadores não são completamente opostos – existe a possibilidade de cooperação.

Dependendo de como a cooperação é considerada, distinguem-se os jogos cooperativos e os jogos não-cooperativos⁴⁵. No primeiro caso, há uma *teoria dos jogos cooperativos*, cujo modelo de análise visa a melhor solução coletiva e, por isso, considera a cooperação como “axiomática, apelando frequentemente ao ótimo de **Pareto**⁴⁶, à justiça e à equidade”⁴⁷. O exemplo mais comum de um jogo cooperativo é o *jogo da barganha*⁴⁸. A suposição básica nos jogos cooperativos é que, havendo comunicação entre os jogadores e

⁴⁵ A distinção usual entre esses dois tipos, segundo Rasmusen, é que “um jogo cooperativo é um jogo no qual os jogadores podem fazer compromissos obrigatórios (*binding commitments*), ao contrário de um jogo não-cooperativo, no qual eles não podem fazer tal acordo” (Rasmusen, p. 21). Entretanto, insiste, “a diferença real reside (...) nos tipos de conceitos de solução empregados” (ibidem), e que são descritos em seguida no texto.

“Um *conceito de equilíbrio* ou um *conceito de solução* (...) é uma regra que define um equilíbrio baseado nas possíveis combinações de estratégias e funções de payoffs” (idem, p. 18). Morton acrescenta: “um conceito de solução é um critério para uma escolha satisfatória, tal que se a escolha de uma pessoa satisfizer a esse critério, e da mesma forma as de outras pessoas que interagem, então o resultado será razoavelmente bom para aquele agente” (Morton, p. 5). São exemplos o critério minimax e o equilíbrio de Nash.

⁴⁶ Ver Nota 83, Capítulo 1, sobre o conceito de ótimo de Pareto.

⁴⁷ Rasmusen, p. 21.

⁴⁸ Em um *jogo da barganha* entre poucos agentes, cada jogador procura obter para si o melhor acordo possível em uma partilha ou na determinação de um preço, ao mesmo tempo em que tenta evitar um impasse, no qual ninguém ganharia nada. Um caso particular de barganha é o da divisão de um determinado bem (uma quantia em dinheiro, um bolo, etc.). A teoria dos jogos cooperativos, não tematizada neste trabalho, foi desenvolvida inicialmente por Nash (Rasmusen, pp. 296-ss., e Skyrms, 2004, pp. 17-ss).

em função da força das promessas que fizerem entre si, é possível obter um compromisso tão firme que leve ao compartilhamento efetivo dos ganhos⁴⁹.

O segundo caso, que é o que nos interessa neste trabalho, é tratado na *teoria dos jogos não-cooperativos*, que difere da anterior por ter um “sabor econômico, com conceitos de solução baseados em jogadores que se empenham em maximizar suas próprias funções de utilidade, sujeitas às restrições dadas”⁵⁰. Nessa teoria, há o pressuposto de que todos os jogadores agem racionalmente, isto é, eles escolhem sempre as opções que levam aos melhores resultados para si próprios⁵¹. Ou seja, é adotada uma concepção instrumental de racionalidade: a ação visa um objetivo de maneira consistente, e são escolhidos os melhores meios para tal fim.

Neste capítulo, em especial no próximo item, a teoria dos jogos não-cooperativos, ou simplesmente teoria dos jogos, vai ser apresentada com ênfase em sua aplicação ao dilema do prisioneiro⁵², que é, ao mesmo tempo, um jogo de soma não-zero e um jogo não-cooperativo. Nele, por definição, não há comunicação entre as partes e, mesmo se houvesse, seria uma questão de confiança acreditar que um acordo de cooperação seja mantido até o último instante da tomada de decisão, desprezando um prêmio que favorece a não-cooperação e incentiva o rompimento de qualquer acordo.

⁴⁹ Rasmusen, p. 21, lembra que na barganha pode haver pagamentos laterais, ou seja, transferências entre os jogadores, que mudam *por fora* os resultados prescritos no jogo. Também podem ocorrer *perdas* (idem, p. 25) como, por exemplo, o pagamento devido aos advogados na partilha de bens no divórcio de um casal.

⁵⁰ Idem, p. 21. Ver definição do conceito de utilidade na Nota 23, Capítulo 1.

⁵¹ O interesse próprio pode incluir o bem-estar de outros, e assim “os valores numéricos das utilidades supostamente refletem tais considerações éticas” (Luce e Raiffa, p. 96). Sen observa, no entanto, que pode haver uma motivação baseada na “responsabilidade social” que é distinta de qualquer consideração de bem-estar (ver texto referente à Nota 74).

Adicionalmente, não haveria nenhuma expectativa de “explorar os desatinos dos oponentes” (Davis, p. 9).

⁵² Conforme descrito no item 2.2 deste Capítulo.

A modelagem do dilema do prisioneiro apresenta dois jogadores, duas opções à disposição de cada jogador e quatro resultados possíveis, que são pares formados pela combinação dessas opções. Essa variabilidade pode ser representada por uma árvore de resultados – é a *forma extensiva* – ou por uma matriz de resultados – é a *forma normal* ou *estratégica*.

A forma extensiva reúne em um diagrama de árvore todos os passos em que um jogo se desenrola, consistindo em “ramos”, ou segmentos de reta, que ligam dois pontos, chamados de “nós”. Em geral, cada nó é um ponto de decisão em que um jogador escolhe o próximo passo que irá tomar. Assim, um conjunto de ramos e nós descreve, dinamicamente, um caminho que pode vir a ser seguido por um jogador.

Como exemplificado na Figura 2.1, a forma extensiva apresenta:

(1) os jogadores, (2a) a vez de cada jogador, (2b) o que cada jogador pode fazer na sua vez de jogar, (2c) o que cada jogador sabe nesse momento, e (3) o prêmio de cada jogador, conforme a combinação dos movimentos que forem escolhidos por eles.⁵³

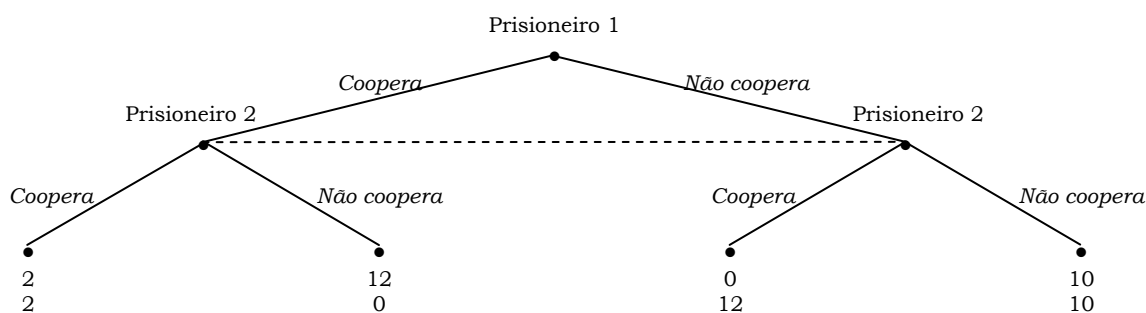


Figura 2.1 – Um dilema do prisioneiro representado na forma extensiva⁵⁴

Obs.: (1) A seqüência do jogo é de cima para baixo. (2) A linha tracejada liga duas representações do mesmo nó, desdobrado para mostrar as opções disponíveis no nó antecedente. (3) Os primeiros números dos pares apresentados no final da árvore são do Prisioneiro 1.

⁵³ Conforme Gibbons, 1992, pp. 115-6.

⁵⁴ Idem, p. 116. Os valores originais da figura foram substituídos pelos dados da versão apresentada por Nozick (1993, p. 50). Ver texto que faz referência à Nota 33.

A segunda maneira de representar um jogo é a forma normal, ou estratégica, que é “estática, ignora todas as questões de sincronia (*timing*) e trata os jogadores como se fizessem suas escolhas simultaneamente”⁵⁵. Entretanto, isso não significa que “os jogadores necessariamente *ajam* simultaneamente: é suficiente que cada um escolha a sua ação sem o conhecimento da escolha dos outros”⁵⁶. Nessa forma normal, adotada neste trabalho, todas as informações estão reunidas em uma matriz simétrica, na qual os jogadores compartilham a mesma visão e os mesmos elementos de análise.

No dilema do prisioneiro exemplificado nas Figuras 2.2a e 2.2b, cada jogador escolhe uma das duas opções: na primeira, ele coopera com o outro prisioneiro, mantendo-se calado e sem denunciar o companheiro; na segunda, ele não coopera com o companheiro, pois resolve confessar e denunciar.

		Prisioneiro 2	
		<i>Coopera</i>	<i>Não coopera</i>
Prisioneiro 1	<i>Coopera</i>	2, 2	12, 0
	<i>Não coopera</i>	0, 12	10, 10

Figura 2.2a – Um dilema do prisioneiro representado em número de anos de prisão⁵⁷
 (Em cada quadrícula, o primeiro número de cada par corresponde ao Prisioneiro 1).

		Prisioneiro 2	
		<i>Coopera</i>	<i>Não coopera</i>
Prisioneiro 1	<i>Coopera</i>	2, 2	0, 3
	<i>Não coopera</i>	3, 0	1, 1

Figura 2.2b – Um dilema do prisioneiro representado na forma normal⁵⁸
 (Os números agora representam as utilidades das sentenças para os prisioneiros.)

⁵⁵ Myerson, p. 47.

⁵⁶ Gibbons, 1992, p. 4.

⁵⁷ Nozick, 1993, p. 50. Ver Nota 33 e Nota 54. O número de anos de prisão deve ser visto como *desutilidade*, isto é, quanto maior, *menos* desejado.

⁵⁸ Poundstone, 1992, p. 120.

Da Figura 2.2a para a Figura 2.2b, o tempo de prisão foi transformado em utilidades, para indicar o valor que cada prisioneiro atribui à sentença que receberia, medido em seu próprio sistema de valores.

Nessa forma normal de representação estão presentes os *jogadores*, que escolhem as *estratégias* disponíveis nas diversas circunstâncias, e a *matriz de resultados* que indica os ganhos, ou *utilidades*, que os jogadores terão como prêmio pelas estratégias escolhidas por eles⁵⁹. A notação utilizada para representar a forma normal é dada pela seguinte definição:

A forma normal de representação de um jogo de n jogadores especifica os *espaços de estratégias* S_1, \dots, S_n dos jogadores e suas funções de resultado (*payoff functions*) u_1, \dots, u_n . Indicamos esse jogo por $G = \{S_1, \dots, S_n; u_1, \dots, u_n\}$.⁶⁰

O exame das Figuras 2.2a e 2.2b revela que um jogador não tem o poder de sozinho, por meio de sua escolha, definir o resultado do jogo, pois este depende também da escolha que o outro fizer. Um jogador pode apenas, na sua vez, apontar para uma das duas linhas, se ele for o nº 1, ou para uma das duas colunas, se ele for o nº 2, e aguardar que o outro aponte a sua. O resultado do jogo será a quadrícula em que se cruzarem a linha e a coluna que forem escolhidas pelos jogadores.

As opções à disposição dos jogadores são chamadas de *estratégias*⁶¹, porque são interdependentes: elas refletem as expectativas de cada um a respeito do outro, como também indicam as diferentes respostas às possíveis ações que os jogadores vierem a escolher. Por esse motivo, a teoria dos jogos é associada ao ramo das ciências sociais que estuda a tomada de decisão em

⁵⁹ Gibbons, 1992, p. 3, e Davis, p. 6.

⁶⁰ Gibbons, 1992, p. 4.

⁶¹ A questão do *contexto estratégico* já foi mencionada na Nota 37.

situações em que as pessoas têm interesses tanto comuns quanto conflitantes, e por isso estão envolvidas em ações interativas e interdependentes⁶².

A estratégia de um jogador é “uma regra que lhe diz qual ação escolher em cada instante do jogo, dado o seu conjunto de informações”⁶³, e nesse conjunto se incluem os desejos e propósitos do jogador, bem como as suas crenças em relação ao ambiente e às ações dos outros jogadores. A estratégia de um jogador deve lhe dizer “quais ações realizar em cada situação concebível, mesmo que ele não espere encontrar tal situação”⁶⁴.

Há então um elemento contrafactual⁶⁵ na análise estratégica, pois ela procura relacionar todas as opções de escolha que sejam viáveis para os participantes, e combina essas opções para simular as conseqüências potenciais que serão lançadas em termos de utilidades na matriz de resultados. Como a realidade (ou a possibilidade) em geral é mais rica do que a sua representação – o que ocorre, por exemplo, quando as conseqüências se mostram diferentes daquelas inicialmente presumidas –, o modelo deve dar margem a revisões, pois “a observação de um fato discrepante da teoria não invalida a teoria”⁶⁶. Ao contrário, os indivíduos aprendem com a experiência, e uma nova informação pode fazer com que o modelo evolua para representar essa nova situação, de um modo heurístico⁶⁷.

Essa observação é útil devido à dúvida se de fato é possível, em um contexto estratégico, conhecer as preferências e as utilidades a partir das

⁶² Dixit e Nalebuff (pp. 2, 13 e 33).

⁶³ Rasmusen, p. 16.

⁶⁴ Idem, p. 17.

⁶⁵ Uma proposição contrafactual é do tipo “Se eu fizesse *a*, então aconteceria *c*” (Gibbard e Harper, p. 133).

⁶⁶ Kreps, p. 113.

⁶⁷ Idem, p. 155.

ações. O pressuposto desse modelo de inferência ⁶⁸ é que há um comportamento racional que “não é simplesmente um comportamento inteligente, mas um comportamento motivado por um cálculo consciente de vantagens, (...) baseado em um *sistema de valores explícito e internamente consistente*”

⁶⁹. No entanto, esse comportamento racional é paradoxal, pois

não é uma vantagem universal (...) [ser] manifestadamente racional na decisão e na motivação. (...) Pode ser perfeitamente racional (...) suspender ou destruir a própria “racionalidade”, pelo menos até certo ponto. (...) Uma teoria que faz da racionalidade um postulado explícito está em condições não somente de modificar tal postulado e examinar o seu significado, como também desvelar parte do mistério que ela contém. (...) E os resultados alcançados por uma análise teórica do comportamento estratégico *são* freqüentemente um pouco paradoxais; eles contradizem freqüentemente o senso comum e as regras geralmente aceitas. ⁷⁰

É bem possível que um observador, por desconfiar das observações e por desconhecer *a priori* os sistemas de valores dos envolvidos, seja obrigado a adotar, como representação da racionalidade, um conjunto-padrão de preferências, que lhe permitirá presumir, pelo menos inicialmente, as utilidades a serem lançadas em uma matriz de resultados.

Essa dificuldade é vista por **Sen**⁷¹ de um modo diferente daquele provocado pelos eventuais “erros” cometidos com o propósito de enganar. Ele aponta que, mais do que as ações estratégicas não revelarem as preferências, é importante reconhecer que as escolhas nem sempre estão “rigidamente limitadas às próprias preferências” ⁷².

Sen descreve duas situações em que o comportamento pode ser diferente do esperado pela racionalidade-padrão e que é exemplificado no caso

⁶⁸ Esse modelo é o ponto de partida tanto da teoria da ação de Ramsey quanto da teoria da preferência revelada. Ver textos que fazem referência às Notas 24 e 57, Capítulo 1.

⁶⁹ Schelling, p. 4. Grifos nossos.

⁷⁰ Idem, p. 18.

⁷¹ Ver item 1.2, em especial o texto que faz referência à Nota 53, Capítulo 1.

⁷² Sen, 1973, p. 72.

do dilema do prisioneiro: na primeira, quando for relativamente fácil garantir o cumprimento de um contrato de não-confissão; na segunda, quando não houver essa “solução contratual” e, mesmo assim, os prisioneiros não confessarem⁷³. Nessa segunda situação – que é a que nos interessa – há ainda uma importante distinção a registrar: a cooperação pode ser motivada por uma “responsabilidade social”, e não devida meramente a uma preferência pelo próprio bem-estar, ainda que este inclua o bem-estar do outro⁷⁴.

Enfim, a teoria não pode subestimar o “fato de que o homem é um animal social” que, diferentemente dos outros animais, não revela tão claramente suas preferências⁷⁵. Na opinião de **Sen**,

um ato de escolha para esse animal social é, em um sentido fundamental, sempre um ato social. Ele pode estar apenas vagamente atento aos imensos problemas de interdependência que caracterizam a sociedade (...) mas seu comportamento é *algo mais* que uma mera tradução de suas preferências individuais.⁷⁶

Ao criticar o sentido restrito da expressão “preferências individuais”⁷⁷, **Sen** quer realçar a necessidade de se adotar um conceito de racionalidade que extrapole os limites da consistência interna, pois esta só adquire significado no próprio contexto das ações interdependentes⁷⁸. Aquele “algo mais” apontado por **Sen**, e que não é descrito pelas preferências individuais, seria a motivação que está além do bem-estar e que mostraria a pessoa “em termos de sua condição de agente”, capaz de estabelecer objetivos, compromissos e valores⁷⁹. Mas a descrição de uma pessoa não está confinada às

⁷³ Ibidem. Um caso extremo seria cada prisioneiro tentar maximizar o bem-estar do outro.

⁷⁴ Idem, p. 76. Ver comentário na Nota 51.

⁷⁵ Idem, p. 73.

⁷⁶ Ibidem. Grifo nosso.

⁷⁷ Ver item 1.2, em especial texto relacionado à Nota 86, Capítulo 1.

⁷⁸ Ver item 1.2, texto que faz referência à Nota 55 e seguintes, Capítulo 1.

⁷⁹ Sen, 1987, p. 57. Citação já referida na Nota 87 do Capítulo 1.

suas realizações: ela incluiria a própria liberdade do agente, cujo valor intrínseco mantém aberta a possibilidade de “algo” por fazer e ainda não catalogado nas observações empíricas da relação entre ações e preferências⁸⁰.

Uma comparação entre as versões da relação ações/preferências, conforme apresentadas até aqui, é resumida na Figura 2.3.

Autor	Relação entre Ações e Preferências	Conceito de Racionalidade	Sistema de Valores / Utilidades
Sen	Ações <i>estratégicas</i> revelam <i>mais que</i> preferências <i>individuais</i> .	Significado da ação depende do contexto.	Responsabilidade social. Valor intrínseco da liberdade.
Schelling	Ações <i>estratégicas</i> não revelam preferências.	Inclui aparente inconsistência.	“Racionalidade” para enganar.
Samuelson (Ramsey)	Ações revelam preferências (crenças).	Consistência interna.	Bem-estar próprio. Inclui bem-estar de outros.

Figura 2.3 – A relação entre ações e preferências, e o conceito de racionalidade (A tabela deve ser lida de baixo para cima, a partir da tese básica).

Essa análise mostra o cuidado que a representação de uma situação social na teoria dos jogos exige. O modelo não pode simplesmente enquadrar os agentes em um sistema de valores que não é o deles, mas derivado de uma racionalidade padronizada. Se os valores reais forem diferentes daqueles que o observador supõe, ou “se o jogo empírico for diferente do jogo teórico” ⁸¹, as escolhas que o agente fizer parecerão irracionais e as conclusões serão enganosas. O caráter heurístico do modelo não pode ser esquecido, pois permite que a sua normatividade, ou sua capacidade de nos ensinar a agir, seja sempre provisória – já que ele trabalha com preferências estimadas e contém elementos contingentes, que são descritivos e sujeitos a mudança.

⁸⁰ Ver texto que faz referência à Nota 95 e seguintes, Capítulo 1.

⁸¹ Expressão de Blackburn. Ver texto que faz referência às Notas 143-ss, Capítulo 1.

2.4 Um recorte na teoria para examinar o dilema do prisioneiro

A teoria dos jogos trabalha com quatro modelos de solução dos jogos, conforme eles sejam estáticos ou dinâmicos, e as informações estejam ou não completamente disponíveis aos jogadores. Esses quatro modelos, que serão resumidos neste item, são os seguintes:

1. Jogos estáticos de informação completa.
2. Jogos dinâmicos de informação completa.
3. Jogos estáticos de informação incompleta.
4. Jogos dinâmicos de informação incompleta.

O primeiro modelo se aplica aos *jogos estáticos de informação completa*⁸², em que todos os resultados possíveis são de conhecimento comum e a eles está associada a noção de que, se houver uma única solução para o jogo, tal solução deve estar em *equilíbrio de Nash*, isto é:

Dado um jogo de n jogadores, de forma normal $G = \{S_1, \dots, S_n; u_1, \dots, u_n\}$, as estratégias (s^*_1, \dots, s^*_n) estão em *equilíbrio de Nash* se, para cada jogador i , s^*_i for a melhor resposta de i às estratégias especificadas para os outros $n-1$ jogadores, $(s^*_1, \dots, s^*_{i-1}, s^*_{i+1}, \dots, s^*_n)$.⁸³

Esse equilíbrio de **Nash** significa que nenhum dos jogadores tem razões para escolher uma opção que não seja a melhor para ele nesse jogo. Uma estratégia é considerada *estável* para um jogador se qualquer mudança unilateral para outra estratégia implicar um resultado pior para esse jogador. Se houver um ponto em que coincidam estratégias estáveis de todos os jogadores, tal resultado é um *ponto de equilíbrio*. Essa noção de equilíbrio permite eliminar um regresso infinito que surgiria no processo de escolha de

⁸² Rasmusen, p. xii.

⁸³ Idem, p. 8.

cada jogador que, ao antecipar as ações dos outros, teria que fazer também uma estimativa do que os outros esperariam que ele próprio fizesse⁸⁴.

Para que a definição do ponto de equilíbrio esteja correta, “é necessário que cada jogador esteja disposto a escolher a estratégia predita pela teoria” ⁸⁵, ou seja, que cada jogador aja racionalmente e não se decida por nenhuma estratégia estritamente dominada, conforme a seguinte definição:

Em um jogo de forma normal $G = \{S_1, \dots, S_n; u_1, \dots, u_n\}$, sejam s'_i e s''_i estratégias viáveis para o jogador i (isto é, s'_i e s''_i são membros de S_i). A estratégia s'_i é estritamente dominada pela estratégia s''_i se para cada combinação de estratégias viáveis dos outros jogadores, o resultado de i ao jogar s'_i for estritamente menor que o resultado de i ao jogar s''_i .⁸⁶

No exemplo do dilema do prisioneiro apresentado na Figura 2.2b (repetida abaixo), a estratégia de cooperação que está à disposição do Prisioneiro 1 é *dominada* pela estratégia de não-cooperação. Isso acontece porque, qualquer que seja a escolha que o Prisioneiro 2 fizer, o resultado para o Prisioneiro 1 será pior se ele escolher a estratégia de cooperação (no caso, menos utilidade significa mais anos de prisão). A mesma análise, com os resultados simetricamente alterados, pode ser feita do ponto de vista do Prisioneiro 2.

		Prisioneiro 2	
		<i>Coopera</i>	<i>Não coopera</i>
Prisioneiro 1	<i>Coopera</i>	2, 2	0, 3
	<i>Não coopera</i>	3, 0	1, 1

Figura 2.2b (repetição) – Um dilema do prisioneiro representado na forma normal⁸⁷
 (Os números são utilidades. O primeiro número de cada par corresponde ao Prisioneiro 1.)

⁸⁴ Elster, p. 7.

⁸⁵ Rasmusen, p. 8. Esse é o aspecto formalmente normativo a que se faz referência no final do item 2.3.

⁸⁶ Idem, p. 5.

⁸⁷ Poundstone, 1992, p. 120. Repetição da Figura 2.2b.

A solução do jogo é, então, o ponto de equilíbrio de **Nash**, ou seja, a quadrícula (1, 1), em que os dois jogadores não cooperam. O que é surpreendente, daí ser um dilema, é que o equilíbrio de **Nash**, no caso do dilema do prisioneiro, não é um ótimo de **Pareto**⁸⁸, pois haveria um resultado melhor para ambos os jogadores, que seria a quadrícula (2, 2), de cooperação mútua. Ou seja, a melhor decisão do ponto de vista da racionalidade individual viola um certo critério de bem-estar coletivo, pois leva a um resultado que nem sempre é o melhor para todos⁸⁹.

* * *

Um segundo modelo é aplicável aos *jogos dinâmicos de informação completa*⁹⁰, no qual um jogo se desenrola em vários estágios e tem como resultado final a soma dos resultados que efetivamente ocorrem em cada estágio. No recorte que estamos fazendo da teoria dos jogos para examinar o dilema do prisioneiro, o foco está nos jogos dinâmicos que, em cada estágio, têm movimentos simultâneos⁹¹, para os quais é válida a seguinte proposição:

Se um jogo simples G tem um único equilíbrio de Nash, então, para qualquer número finito de repetições T , o jogo repetido $G(T)$ tem um único resultado do subjogo perfeito: o equilíbrio de Nash de G é jogado em cada estágio. ⁹²

⁸⁸ Como citado no Capítulo 1, Nota 83, um ponto ótimo de Pareto é uma posição em que nenhum jogador pode melhorar o seu resultado sem piorar o de outro. Qualquer posição em que não houver essa unanimidade será um subótimo. No caso do dilema do prisioneiro, todos os resultados são ótimos de Pareto, exceto o ponto de equilíbrio de Nash.

⁸⁹ Segundo Skyrms, uma “*justiça utilitária* requer que o estado do grupo seja ótimo de Pareto. (...) Em jogos de soma não-zero [diferentemente de jogos de soma zero] os equilíbrios podem não ser ótimos de Pareto e os pontos ótimos de Pareto podem não ser equilíbrios” (Skyrms, 1990, p. 137), mas essa diferença não impede que a escolha de jogadores racionais continue sendo o equilíbrio.

⁹⁰ Rasmusen, p. 55.

⁹¹ Quando em um jogo os jogadores fizerem um movimento de cada vez, pode-se aplicar o método da *indução retroativa*, partindo-se do fim do jogo para o início, ponto a ponto, o que não é possível fazer se o jogo tiver movimentos simultâneos ou horizonte infinito.

⁹² Gibbons, 1992, p. 84.

A partir do segundo estágio de um jogo dinâmico, para cada um dos resultados possíveis do estágio t corrente, há todos os resultados que já ocorreram anteriormente até o estágio $t-1$, configurando uma história do jogo até aquele instante $t-1$. O restante do jogo, a partir daí, é um leque de possibilidades. Um *subjogo* é um pedaço de um jogo dinâmico, definido como “a parte que resta a ser jogada, começando em qualquer ponto no qual a história completa do jogo até então for de conhecimento comum entre os jogadores”⁹³. Em um dilema do prisioneiro de dois estágios, por exemplo, há quatro subjogos, correspondentes à segunda parte dos jogos que se seguem aos quatro possíveis resultados do primeiro estágio.

Por definição, “um equilíbrio de **Nash** é um subjogo perfeito se as estratégias dos jogadores constituírem um equilíbrio de **Nash** em cada subjogo”⁹⁴. Um jogo dinâmico de informação completa pode ter vários equilíbrios de **Nash**, mas apenas os que tiverem ameaças e promessas críveis são “equilíbrios de subjogo perfeito de **Nash**”⁹⁵.

A aplicação dessa teoria ao dilema do prisioneiro com um número finito de estágios mostra que o equilíbrio continua sendo o mesmo do jogo estático, ou seja, as estratégias não-cooperativas dominam, pois “a cooperação (...) não pode ser alcançada em nenhum estágio”⁹⁶ do jogo.

Entretanto, em um *jogo repetido*, que tem horizonte de tempo infinito ou número indefinido de repetições, “as pessoas interagem ao longo do tempo, [e dessa forma] as ameaças e promessas a respeito do comportamento

⁹³ Idem, p. 94. Aqui se distingue a noção de *informação perfeita*, pela qual em cada estágio os jogadores conhecem toda a história do jogo até aquele momento.

⁹⁴ Idem, 1992, p. 95. Citação de Selten (1985).

⁹⁵ Idem, 1992, p. 57.

⁹⁶ Idem, 1992, p. 84.

futuro podem influenciar o comportamento presente”⁹⁷. Além disso, “em um jogo repetido [indefinidamente] (...) nenhuma jogada é necessariamente a última, [porque] um jogador pode sempre considerar o efeito que sua jogada terá nas jogadas e na informação de outros jogadores no futuro”⁹⁸, criando as condições para que o jogo se torne mais cooperativo ou mais beligerante.

Como se verá adiante, em um dilema do prisioneiro repetido – como em outros jogos repetidos –, é possível sustentar em equilíbrio um comportamento cooperativo e “a chave para esse equilíbrio generoso é que, sempre que os jogadores se encontrem, eles acreditem haver uma probabilidade muito alta de voltarem a jogar”⁹⁹. Tal crença não seria efetiva se eles soubessem, com antecedência, o momento em que o jogo iria acabar.

Ao lidar com resultados distribuídos no tempo, deve-se introduzir nos cálculos um *fator de desconto*¹⁰⁰, que vai representar em termos presentes a diminuição de valor que o jogador atribui às utilidades que são esperadas no futuro¹⁰¹. Dessa forma, o resultado de um jogo dinâmico não vai ser, como descrito acima, a mera soma dos resultados dos vários estágios, mas o *valor presente* da seqüência de resultados, dado pelo somatório dos resultados de cada estágio multiplicados pelo fator de desconto do respectivo estágio.

⁹⁷ Gibbons, 1997, p. 136, e também Gibbons, 1992, p. 86.

⁹⁸ Myerson, p. 308.

⁹⁹ Idem, p. 309.

Como já citado na Nota 43, Skyrms (2004, p. 5) chama a atenção de que um dilema do prisioneiro repetido se transforma, a partir do segundo estágio, em um *jogo da coordenação*, no qual o prêmio para cooperar é maior que para não cooperar. Ver item 2.5, Nota 141.

¹⁰⁰ “O fator de desconto $\delta = 1/(1+r)$ é o valor hoje de um \$1 que será recebido um estágio depois, onde r é a taxa de juros [ou taxa de desconto] por estágio” (Gibbons, 1992, p. 89).

¹⁰¹ Esse fenômeno é descrito por Hume como “miopia” (expressão de Taylor, pp. 153-54), pois tudo “o que estiver mais perto de nós em termos de tempo ou espaço afeta nossa imaginação com maior força do que o que estiver mais afastado, o efeito do tempo sendo maior que aquele do espaço” (Hume, 2.3.7.1-4).

No caso geral, dado o fator de desconto δ , o valor presente de uma sequência infinita de resultados $\pi_1, \pi_2, \pi_3, \dots$, é igual à seguinte soma:

$$\pi_1 + \delta\pi_2 + \delta^2\pi_3 + \dots = \sum_{t=1}^{\infty} \delta^{t-1} \pi_t \quad 102,$$

onde t é o número de estágios que separa o instante de realização da utilidade esperada do instante de referência (o *presente*).

A introdução do valor presente pode alterar o resultado do jogo dinâmico, dependendo de qual valor se adote para a taxa de desconto.

De um lado, se a taxa de desconto for muito alta,

o jogo praticamente se transforma em um jogo de um estágio apenas. Quando a taxa de desconto real for de 1.000%, um pagamento no próximo ano é um pouco melhor que daqui a cem anos, e assim o próximo ano se torna praticamente irrelevante. ¹⁰³

Dessa forma, uma taxa de desconto muito alta reduz significativamente o valor presente dos futuros ganhos da cooperação, fazendo com que “o único resultado de equilíbrio seja confessar eternamente” ¹⁰⁴.

Por outro lado, se a taxa de desconto for baixa,

o jogador que tiver um potencial para não-cooperar deve avaliar se o valor presente da cooperação continuada não compensaria o ganho de curto-prazo, fruto da não-cooperação, mas que é depreciado pelas perdas de longo prazo [que ele sofrerá por não cooperar, e que será a concretização de promessas e ameaças dos demais]. Se os jogadores forem suficientemente pacientes (isto é, se a taxa de desconto for suficientemente baixa), a cooperação pode ocorrer no equilíbrio de um jogo repetido, mesmo que não ocorra no jogo estático. ¹⁰⁵

¹⁰² Gibbons, 1992, p. 90.

Para taxas de desconto positivas, isto é, $r \geq 0$, a expressão $\delta = 1/(1+r)$ fornece valores de δ entre zero e um. Então, o fator de desconto de cada estágio t , dado por δ^{t-1} , diminui quando t cresce – o que justifica a “miopia” mencionada na Nota 101.

¹⁰³ Rasmusen, p. 113.

Para Gibbons, a taxa de desconto pode refletir “não só a taxa de preferência no tempo, como também a probabilidade de que o estágio corrente seja o último, [levando um] jogo ‘infinitamente repetido’ a terminar em uma data aleatória” (Gibbons, 1997, p. 136). Assim, quando a taxa de desconto r for alta, o fator de desconto de um estágio t tenderá *rapidamente* para zero à medida que t cresça, tornando desprezível o valor presente das parcelas a partir de determinado estágio, como se o jogo terminasse naquele ponto.

¹⁰⁴ *Ibidem*.

¹⁰⁵ Gibbons, 1997, p. 137.

A análise dos jogos infinitamente repetidos, inclusive para o dilema do prisioneiro repetido, é resumida então pelo *Folk theorem*, que estabelece as condições em que é possível o surgimento espontâneo da cooperação mútua:

Seja G um jogo estático finito de informação completa. Sejam (e_1, \dots, e_n) os resultados de um equilíbrio de Nash de G , e sejam (x_1, \dots, x_n) quaisquer outros resultados viáveis de G . Se $x_i > e_i$ para todo jogador i e se δ for suficientemente próximo de um, então existe um equilíbrio de Nash no subjogo perfeito do jogo infinitamente repetido $G(\infty, \delta)$ que tende ao resultado médio (x_1, \dots, x_n) .¹⁰⁶

* * *

Um terceiro modelo, que se aplica aos *jogos estáticos de informação incompleta*¹⁰⁷, tem a seguinte característica: “pelo menos um dos jogadores não está certo sobre a função de resultados dos outros jogadores”¹⁰⁸. Um exemplo clássico é o de um leilão onde “cada licitante conhece a sua própria avaliação do bem que está sendo vendido, mas não conhece a avaliação de nenhum outro licitante”¹⁰⁹. Para esse caso, a teoria apresentada até aqui é convenientemente alterada para abrigar um *equilíbrio bayesiano de Nash*, com a introdução das crenças dos jogadores, representadas por probabilidades subjetivas¹¹⁰, para suprir a falta do conjunto completo de informações.

Em uma situação estratégica complexa e na qual seja necessária uma decisão imediata, um agente pode ainda preferir uma solução “boa o suficiente (...) na qual ele não fique mal, e que seja provável que os outros também

¹⁰⁶ Gibbons, 1992, p. 97. Gibbons atribui esse enunciado a Friedman (1971).

Rasmusen acrescenta a condição de haver uma “probabilidade zero, ou positiva, mas suficientemente pequena, de que o jogo termine em qualquer repetição” (Rasmusen, p. 112).

¹⁰⁷ “Jogos de informação incompleta”, ou “jogos bayesianos” (idem, p. 138).

¹⁰⁸ Gibbons, 1997, p. 138.

¹⁰⁹ Ibidem.

¹¹⁰ Ver Anexo 1 a respeito da regra de Bayes e probabilidade subjetiva.

a prefiram para, da mesma forma, não ficarem mal”¹¹¹. Esse distanciamento dos pressupostos da teoria tradicional requer uma regra prática¹¹², tal como a que propõe adotar um resultado ótimo de **Pareto**, pois enquanto é possível desenvolver um método simples para se chegar ao “equilíbrio de **Pareto**”, não há um método semelhante para se chegar ao equilíbrio de **Nash**¹¹³.

* * *

O quarto modelo, e o mais geral deles, é aplicável aos *jogos dinâmicos de informação incompleta*, nos quais a solução é obtida pela combinação do conceito de equilíbrio de subjogo perfeito de **Nash** com a regra de **Bayes** – é um *equilíbrio bayesiano perfeito*. Um exemplo da aplicação dessa teoria é um *jogo da reputação*, “no qual pelo menos um dos jogadores tem algo a aprender sobre o outro”¹¹⁴, como no caso de um jogador desconhecer o *tipo*¹¹⁵ de seu oponente. O jogador que detém a informação privada tenta comunicar (ou enganar), enquanto o outro tenta aprender e responder.

Uma conclusão importante é que “um dilema do prisioneiro repetido e finito, que começa com uma pitada de informação privada, mas do tipo adequado, pode ter cooperação em equilíbrio em todos os períodos, exceto naqueles próximos do final”¹¹⁶. Basta, para isso, que um dos jogadores credi-

¹¹¹ Morton, p. 143.

¹¹² Um exemplo de regra prática aplicada à decisão é o conceito de “*satisficing*”, termo cunhado por Herbert Simon (1957) pela combinação de “*satisfying*” com “*sufficing*”, que sugere adotar “a primeira escolha satisfatória encontrada dentre aquelas (...) disponíveis a curto prazo, em oposição a visualizar antecipadamente a escolha ótima e persegui-la até que seja encontrada. [Por exemplo:] é mais provável que um jovem propenso ao casamento faça uma proposta à mulher mais atraente dentre as disponíveis no seu círculo de relacionamento, do que procure longamente por uma companheira ideal preconcebida” (Wilson, pp. 224-25).

¹¹³ Morton, p. 143.

¹¹⁴ Gibbons, 1992, p. 225, n75.

¹¹⁵ “O *tipo* de um jogador é o conjunto de estratégias, informação e função de resultados que (...) [ele tem] no começo de um jogo de informação incompleta” (Rasmusen, p. 53).

¹¹⁶ Gibbons, 1997, p. 142. Citação de Kreps, Milgrom, Roberts e Wilson (1982).

te que o outro vá jogar somente a estratégia *tit-for-tat* ¹¹⁷, mesmo que isso aconteça com baixa probabilidade. É surpreendente que “uma pequena quantidade de informação incompleta possa fazer uma grande diferença nos resultados” ¹¹⁸. A novidade desse *equilíbrio bayesiano perfeito* é que

as crenças são elevadas ao nível de importância das estratégias na definição do equilíbrio. Isto é, a definição de equilíbrio não mais consiste apenas de uma estratégia para cada jogador, mas inclui agora também uma crença para cada jogador, sempre que o jogador tiver a vez e for incerto sobre a história do jogo até então. A vantagem de fazer das crenças dos jogadores uma parte explícita do equilíbrio é que, exatamente como antes insistimos que os jogadores escolhessem estratégias críveis (isto é, subjogos perfeitos), podemos agora também insistir que tenham crenças razoáveis.

¹¹⁹

¹¹⁷ Na estratégia *tit-for-tat* (*olho por olho, dente por dente*) o jogador coopera na jogada inicial, e daí em diante faz o que o oponente fizer na jogada anterior (Gibbons, 1992, p. 225).

¹¹⁸ Rasmusen, p. 149.

¹¹⁹ Gibbons, 1997, p. 142.

2.5 Caracterização do dilema do prisioneiro

O dilema do prisioneiro tem a aparência da Figura 2.4, em que os dois jogadores *I* e *II* podem ou não cooperar um com o outro:

		<i>Jogador II</i>	
		<i>Coopera</i>	<i>Não coopera</i>
<i>Jogador I</i>	<i>Coopera</i>	2, 2	0, 3
	<i>Não coopera</i>	3, 0	1, 1

Figura 2.4 – O dilema do prisioneiro¹²⁰
 (Os valores representam *utilidades*. Os primeiros números de cada par são do jogador *I*.)

O dilema do prisioneiro é caracterizado por uma *tentação* a não cooperar (valor 3) maior que a *recompensa* pela cooperação mútua (valor 2), ambos maiores que a *punição* pela não-cooperação (valor 1), e o *prejuízo do “otário”* (valor 0) que coopera sozinho e propicia o ganho do outro¹²¹.

Não basta envolver dois agentes, haver duas opções e ter soma não-nula para ser considerado um dilema do prisioneiro. Apesar de simples, essa estrutura 2x2 produz setenta e oito jogos distintos¹²² que podem ser formados pela combinação dos valores relativos da matriz de resultados. Dentre esses, estamos interessados no subconjunto dos vinte e quatro jogos simétricos, em que os resultados são idênticos quando os jogadores trocam de posição. Para facilitar a análise desse subconjunto, são identificados a seguir todos os quatro resultados possíveis:

- CC significa que os dois jogadores cooperam.
- DD que os dois não cooperam.

¹²⁰ Poundstone, *PD*, p. 120. Esta figura é semelhante à Figura 2.2b.

¹²¹ Idem, pp. 118-27. Poundstone acrescenta ainda a condição de que “a recompensa deve ser maior que a média da tentação e do prejuízo do otário” (idem, p. 127).

Blackburn sugere os seguintes nomes para as opções: *Vitória* (valor 3), *Cooperação* (valor 2), *Guerra* (valor 1) e *Ruína* (valor 0) (Blackburn, 1998, p. 177).

¹²² Poundstone, 1992, p. 215. Referência a Guyer e Rapoport, “A Taxonomy of 2x2 Games”, in *General Systems* (1966), 11, 203-14.

CD significa que o jogador *I* coopera e o *II* não.
 DC significa que o jogador *I* não coopera e o *II* coopera. ¹²³

Toda a variedade de jogos pode, então, ser descrita em termos dos valores relativos dos quatro resultados, conforme resumido na Figura 2.5:

		<i>Jogador II</i>	
		<i>Coopera</i>	<i>Não coopera</i>
<i>Jogador I</i>	<i>Coopera</i>	CC	CD
	<i>Não coopera</i>	DC	DD

Figura 2.5 – Um jogo 2x2 simétrico: duas pessoas, duas alternativas, soma não-nula.

A maioria dos vinte e quatro jogos simétricos pode ser descartada da análise, pois em alguns o prêmio por cooperar é maior que por não cooperar, e em outros não há qualquer ganho por não cooperar. O dilema aparece quando há um incentivo à não-cooperação e, nesse caso, como mencionado, o “bem comum [dos dois jogadores] é subvertido pela racionalidade individual: cada jogador deseja a cooperação do outro, enquanto ele mesmo é tentado a não cooperar” ¹²⁴. Há quatro jogos simétricos que merecem destaque. São chamados de jogos sociais pela sua aplicação nas ciências sociais, e têm entre si semelhanças e distinções importantes, resumidas na Figura 2.6:

Ordem de preferência para o jogador I ¹²⁵	Nome do jogo	Relação com o dilema do prisioneiro
DC>DD>CC>CD	Jogo do Impasse (<i>Deadlock</i>)	Recompensa e punição trocadas
DC>CC>DD>CD	Dilema do prisioneiro	–
DC>CC>CD>DD	Jogo do Covarde (<i>Chicken</i>)	Inversão dos prêmios do “otário”
CC>DC>DD>CD	Jogo da Coordenação (<i>Stag Hunt</i>)	Recompensa e tentação trocadas

Figura 2.6 – Os quatro jogos sociais¹²⁶

¹²³ As letras *C* e *D* vêm do inglês *Cooperate* (cooperar) e *Defect* (falhar, ou não cooperar).

¹²⁴ Poundstone, 1992, p. 216.

¹²⁵ Devido ao jogo ser simétrico, essa é a ordem dos resultados também para o *jogador II*.

¹²⁶ Figura baseada em Poundstone, 1992, p. 217.

Exemplos numéricos dos quatro jogos são apresentados na Figura 2.7:

<p>Jogo do Impasse</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td></td> <td style="text-align: center;">C</td> <td style="text-align: center;">D</td> </tr> <tr> <td style="text-align: center;">C</td> <td style="text-align: center;">1, 1</td> <td style="text-align: center;">0, 3</td> </tr> <tr> <td style="text-align: center;">D</td> <td style="text-align: center;">3, 0</td> <td style="text-align: center;">2, 2</td> </tr> </table>		C	D	C	1, 1	0, 3	D	3, 0	2, 2	<p>Dilema do Prisioneiro</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td></td> <td style="text-align: center;">C</td> <td style="text-align: center;">D</td> </tr> <tr> <td style="text-align: center;">C</td> <td style="text-align: center;">2, 2</td> <td style="text-align: center;">0, 3</td> </tr> <tr> <td style="text-align: center;">D</td> <td style="text-align: center;">3, 0</td> <td style="text-align: center;">1, 1</td> </tr> </table>		C	D	C	2, 2	0, 3	D	3, 0	1, 1
	C	D																	
C	1, 1	0, 3																	
D	3, 0	2, 2																	
	C	D																	
C	2, 2	0, 3																	
D	3, 0	1, 1																	
<p>Jogo do Covarde</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td></td> <td style="text-align: center;">C</td> <td style="text-align: center;">D</td> </tr> <tr> <td style="text-align: center;">C</td> <td style="text-align: center;">2, 2</td> <td style="text-align: center;">1, 3</td> </tr> <tr> <td style="text-align: center;">D</td> <td style="text-align: center;">3, 1</td> <td style="text-align: center;">0, 0</td> </tr> </table>		C	D	C	2, 2	1, 3	D	3, 1	0, 0	<p>Jogo da Coordenação</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td></td> <td style="text-align: center;">C</td> <td style="text-align: center;">D</td> </tr> <tr> <td style="text-align: center;">C</td> <td style="text-align: center;">3, 3</td> <td style="text-align: center;">0, 2</td> </tr> <tr> <td style="text-align: center;">D</td> <td style="text-align: center;">2, 0</td> <td style="text-align: center;">1, 1</td> </tr> </table>		C	D	C	3, 3	0, 2	D	2, 0	1, 1
	C	D																	
C	2, 2	1, 3																	
D	3, 1	0, 0																	
	C	D																	
C	3, 3	0, 2																	
D	2, 0	1, 1																	

Figura 2.7 – Exemplos numéricos dos jogos sociais ¹²⁷
 (Os equilíbrios de Nash estão em negrito. Os ótimos de Pareto estão sombreados)

No dilema do prisioneiro a estratégia não-cooperativa de cada jogador domina a estratégia cooperativa, o que faz da não-cooperação mútua um equilíbrio de **Nash**. Mas esse equilíbrio único é também o único resultado que *não* é ótimo de **Pareto** – ou seja, é um equilíbrio ineficiente, evidenciando o citado conflito entre o interesse individual e o bem coletivo.

O *jogo do covarde (chicken)*¹²⁸ pode ser descrito como uma disputa entre adolescentes que dirigem carros em alta velocidade, um de encontro ao outro, até que um deles, o covarde, desvie. Nesse jogo há dois equilíbrios de

¹²⁷ Figura baseada em Poundstone, 1992, pp. 120, 199, 218-20.

Além da possibilidade de haver jogos *híbridos* dos quatro tipos citados (Taylor, p. 39), há ainda outros jogos sociais que merecem referência, como os abaixo. O primeiro é um *jogo de coordenação pura*, como ocorre quando duas pessoas tiram o telefone do gancho para ligar uma para a outra no mesmo instante; o segundo é a *Guerra dos Sexos*, em que duas pessoas querem sair juntas, mas discordam sobre aonde ir (Hausman e McPherson, p. 184-5).

<p>Jogo de Coordenação Pura</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td></td> <td style="text-align: center;">A</td> <td style="text-align: center;">B</td> </tr> <tr> <td style="text-align: center;">A</td> <td style="text-align: center;">1, 1</td> <td style="text-align: center;">0, 0</td> </tr> <tr> <td style="text-align: center;">B</td> <td style="text-align: center;">0, 0</td> <td style="text-align: center;">1, 1</td> </tr> </table>		A	B	A	1, 1	0, 0	B	0, 0	1, 1	<p>Guerra dos Sexos</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td></td> <td style="text-align: center;">A</td> <td style="text-align: center;">B</td> </tr> <tr> <td style="text-align: center;">A</td> <td style="text-align: center;">2, 1</td> <td style="text-align: center;">0, 0</td> </tr> <tr> <td style="text-align: center;">B</td> <td style="text-align: center;">0, 0</td> <td style="text-align: center;">1, 2</td> </tr> </table>		A	B	A	2, 1	0, 0	B	0, 0	1, 2
	A	B																	
A	1, 1	0, 0																	
B	0, 0	1, 1																	
	A	B																	
A	2, 1	0, 0																	
B	0, 0	1, 2																	

¹²⁸ Poundstone, 1992, pp. 187-99.

O jogo deve seu nome à metáfora cunhada por B. Russell, que associou o *beco sem saída* da corrida armamentista à disputa de delinquentes juvenis. Esse também foi o tema do filme *Rebel Without a Cause (Juventude Transviada)*, dirigido por Nicholas Ray, 1955.

Nash, que ocorrem quando um jogador coopera e o outro não¹²⁹ – são as melhores respostas que um pode dar às estratégias do outro. A grande diferença em relação ao dilema do prisioneiro é que o pior resultado é a não-cooperação, que levaria, no caso, à destruição ou à morte. Por isso, cada jogador prefere fazer exatamente o oposto ao que o outro fizer: cooperar é melhor do que enfrentar um oponente até o fim decidido a não cooperar.

Um *jogo do covarde* pode acontecer na produção de um bem público, como, por exemplo, na manutenção de um dique para irrigação, em que “cada um dos envolvidos é capaz de fazer todo o trabalho sozinho, mas prefere que o outro o faça; contudo, se o outro não o fizer, ele mesmo faz, para evitar o desastre”¹³⁰. Nesse jogo cada jogador prefere “forçar” a cooperação do outro, esperando caronear a ação cooperativa¹³¹.

O *jogo do impasse (deadlock)*¹³² representa uma situação em que os dois jogadores não querem mesmo cooperar, por isso não chega a ser um verdadeiro dilema. Diferentemente do dilema do prisioneiro, o equilíbrio (o ponto de não-cooperação mútua) é ótimo de **Pareto**. Um exemplo é a decisão a ser tomada por uma empresa sobre se deve se associar ou não a um concorrente para desenvolver novos produtos. O melhor para a sociedade seria que os dois tivessem êxito ao desenvolver processos concorrentes¹³³.

¹²⁹ No jogo do covarde os dois equilíbrios coincidem com dois pontos ótimos de Pareto.

¹³⁰ Taylor, p. 36. Itálico do autor, sublinhando uma condição presente apenas no jogo do covarde e que o faz diferente de um dilema do prisioneiro, pois neste cada participante, por definição, não seria capaz de realizar todo o trabalho sozinho. Essa capacidade é indicada pela ordem de preferências de um jogador (ver Figura 2.6), que prefere a cooperação unilateral à não-cooperação mútua, que seria desastrosa.

¹³¹ Idem, p. 39.

¹³² Poundstone, 1992, p. 218.

¹³³ Bêrni, p. 84.

O jogo da coordenação (*stag hunt*)¹³⁴, como o do covarde, pode ser descrito como um acordo entre dois adolescentes que combinam ir ao último dia de aula com um corte de cabelo ridículo, e ficam indecisos se cumprem o combinado ou não. Há dois equilíbrios, o de cooperação mútua e o de não-cooperação mútua, e ambos os jogadores preferem o primeiro, que também é o único ótimo de **Pareto** do jogo. O jogo da coordenação, na opinião de **Skyrms**, representa melhor o contrato social que o dilema do prisioneiro¹³⁵, como revelado em dois clássicos exemplos de **Hume**: o de dois homens que remam um barco¹³⁶; e o da drenagem de um pântano entre dois vizinhos¹³⁷.

No jogo da coordenação, a cooperação e a racionalidade caminham no mesmo sentido, o que não ocorre no dilema do prisioneiro, onde os dois princípios conflitam. Por outro lado, o jogo da coordenação se aproxima do dilema do prisioneiro e difere do jogo do covarde por ter um incentivo a não cooperar, embora menor que para cooperar. Mesmo assim, “o que é racional para um jogador escolher depende de suas crenças a respeito do que o outro vai escolher”, o que significa que “é melhor caçar o veado se o outro caçar o veado e é melhor caçar a lebre se o outro caçar a lebre”¹³⁸.

¹³⁴ Poundstone, 1992, pp. 218-20. O nome *stag hunt* é derivado da metáfora sobre a importância da cooperação, utilizada por Rousseau no *Discurso sobre a desigualdade* (1755). Alguns autores (como Taylor, p. 18) preferem o nome *jogo da confiança*, ou *da garantia* (*assurance game*). Preferimos a expressão *jogo da coordenação* porque identifica mais claramente um jogo que freqüentemente é confundido com o dilema do prisioneiro.

¹³⁵ Skyrms, 2004, p. xii. A presença de dois equilíbrios no jogo da coordenação leva Skyrms (idem, p. 9) a descrever a adoção do contrato social como um esforço para sair de um equilíbrio, o do estado da natureza, e chegar a outro equilíbrio, o do contrato. Da mesma forma, a modificação do contrato requer um esforço para passar para um novo equilíbrio.

¹³⁶ Hume, 3.2.2.10. O exemplo dos dois remadores é um caso de cooperação em que não há promessa, apenas reconhecimento do interesse mútuo “que os induz a regular a conduta por meio de certas regras, (...) e que dá [a ambos] uma confiança da regularidade futura dessa conduta” (ibidem).

¹³⁷ Idem, 3.2.7.8. Ver Notas 41 e 42.

¹³⁸ Skyrms, 2004, p. 3.

Skyrms propõe uma variante, dada na Figura 2.8, que dramatiza as duas forças a que o jogador racional está submetido, o benefício e o risco.

Jogo da Coordenação – original		Jogo da Coordenação – variante	
	<i>C</i>	<i>D</i>	
<i>C</i>	3, 3	0, 2	<i>C</i>
<i>D</i>	2, 0	1, 1	<i>D</i>

Figura 2.8 – O jogo da coordenação original e uma variante¹³⁹

Nessa variante, os *payoffs* da não-cooperação estão equalizados tal que, se um jogador resolver caçar lebre, é completamente indiferente para ele se o outro vai caçar lebre ou veado. Além disso, apesar de os *payoffs* dos dois equilíbrios de **Nash** estarem próximos, mesmo assim a ação cooperativa exige uma dose de confiança muito grande no outro jogador. No equilíbrio cooperativo, o benefício domina, “mas o jogador corre o risco de que o outro não coopere”; no equilíbrio não-cooperativo, o risco domina, e o jogador apenas “abre mão do prêmio potencial da caça ao veado bem sucedida”¹⁴⁰.

Um dos argumentos a favor da maior atenção ao jogo da coordenação é que um dilema do prisioneiro de dois jogadores, se repetido indefinidamente com uma determinada probabilidade, pode se transformar, a partir do segundo estágio, em um jogo da coordenação¹⁴¹. Basta, para isso, que um dos jogadores use a estratégia de nunca cooperar e que o outro jogador escolha a estratégia de inicialmente cooperar e em seguida retaliar. A proposta de **Skyrms** aponta na direção de um possível desenvolvimento da teoria dos jogos, o que talvez explique algumas das dúvidas que serão levantadas a seguir.

¹³⁹ Ibidem.

¹⁴⁰ Ibidem.

¹⁴¹ Idem, p. 5.

2.6 O dilema do prisioneiro e o conceito de racionalidade

Vamos complementar neste item as críticas de Amartya **Sen** e de Simon **Blackburn** ao conceito de racionalidade, já apresentadas no Capítulo 1 deste trabalho, agora sob o aspecto da possibilidade de cooperação em situações representadas pelo dilema do prisioneiro.

Para **Sen**, a abordagem comum relaciona o surgimento da cooperação em um dilema do prisioneiro finito ao desconhecimento que os jogadores teriam quanto ao número restante de rodadas, e assim seria impossível aplicar um raciocínio de indução retroativa que manteria, em todos os estágios do jogo, a ação dominante não-cooperativa. Contra essa concepção de “defeito”¹⁴², **Sen** considera que as preferências das pessoas poderiam perfeitamente ser diferentes das utilidades apresentadas na matriz de resultados – o que faria com que o problema *real* deixasse de ser um dilema do prisioneiro. Mas é também possível que a motivação para a cooperação surja do reconhecimento de uma “interdependência mútua” dos objetivos ou de um “senso de identidade” social¹⁴³, que teria uma “enorme importância *instrumental* na promoção dos respectivos objetivos dos membros desse grupo”¹⁴⁴.

Entretanto, **Sen** reconhece que identificar o que “uma pessoa está maximizando depende de uma certa interpretação do que ela considera serem as variáveis de controle apropriadas”¹⁴⁵, e assim a decisão por uma ação cooperativa poderia ser acusada de instrumental, isto é, ao promover o bem coletivo estaria visando, no fundo, o próprio interesse. Trata-se, nesse caso,

¹⁴² Situação já mencionada no item 2.3, no texto que faz referência à Nota 71.

¹⁴³ Sen, 1987, p. 101.

¹⁴⁴ *Ibidem*.

¹⁴⁵ *Idem*, p. 102.

segundo **Sen**, de “uma genuína ambigüidade quanto ao que a razão verdadeiramente determina”¹⁴⁶, ou seja, a razão cria um dilema que ela mesma não consegue resolver satisfatoriamente, mas revela que há outras motivações para a ação humana além da maximização do interesse próprio¹⁴⁷.

A crítica de **Blackburn** é mais contundente: ele simplesmente nega¹⁴⁸ que a teoria dos jogos seja capaz de prever o comportamento ou dar qualquer recomendação para a ação. Se, por exemplo, no dilema do prisioneiro, não seguissemos o conselho para jogar a estratégia dominante (*não cooperar*), estaríamos revelando uma preferência diferente, mostrando que os números da matriz estariam errados – enfim, negando o dilema do prisioneiro.

Para **Blackburn** há uma diferença entre um jogo teórico e uma situação empírica¹⁴⁹. Na teoria, é feita uma modelagem de regras e comportamentos que podem não refletir toda a diversidade de opções que os agentes vão encontrar na prática, e cujas preferências só vão ser reveladas ao agirem. Assim, se a matriz teórica de resultados for conhecida com antecedência, é tautológico recomendar que se jogue a estratégia dominante (*não cooperar*), pois os agentes representados pelo modelo não têm realmente que fazer uma escolha: eles já sabem qual é o melhor resultado para eles mesmos. Mas, “em um dilema empírico a tautologia *nunca* se aplica”¹⁵⁰, pois é perfeitemen-

¹⁴⁶ Idem, p. 103.

¹⁴⁷ O tema da existência de vários objetos de valor foi tratado no item 1.2, no texto que faz referência à Nota 100 e seguintes, Capítulo 1.

¹⁴⁸ Como visto no Capítulo 1, item 1.3, para Blackburn há três interpretações para o princípio da maximização da utilidade esperada, que é uma das premissas básicas da teoria dos jogos. Ele seria: (1) uma verdade empírica que produz previsões; ou (2) uma verdade normativa, que produz recomendações; ou (3) um modelo matemático, que impõe um modo de ver. A terceira interpretação é a preferida de Blackburn.

¹⁴⁹ Blackburn, 1998, pp. 168-72. O mesmo argumento foi mencionado no item 1.3, no texto que faz referência à Nota 143 e seguintes, Capítulo 1.

¹⁵⁰ Idem, p. 183.

te racional que as pessoas prefiram as estratégias dominadas, a despeito do sacrifício que essa escolha possa significar.

Blackburn quer evitar o debate entre os que defendem escolhas fortemente dominadas como racionais e os que negam seja possível sustentar no longo prazo escolhas fora do equilíbrio¹⁵¹. Para ele, a melhor saída seria, à maneira de **Ramsey**¹⁵², “mudar a natureza da decisão”, não por tomar como racional algo que de outra forma seria considerado irracional, mas “mudando o modo pelo qual a utilidade será realizada na situação em questão”¹⁵³. É o que ele procura mostrar por meio do *paradoxo da toxina*:

Um experimentador, que conhece muito bem as minhas intenções, (...) me promete uma grande quantia, que receberei na segunda-feira somente se eu, na segunda-feira, tiver sinceramente a intenção de beber na quarta-feira uma toxina que me provocará náuseas e me adoecerá. Eu sei (e o experimentador sabe que eu sei, e assim por diante) que na quarta-feira não haverá absolutamente nenhuma razão para eu fazer isso, e haverá algumas razões para eu não fazer, pois a toxina é desagradável. (...) O experimentador me deu um motivo para eu formar a intenção de fazer algo para o qual haverá razões *apenas* para não fazer. Formarei a intenção? E nesse caso, serei irracional?¹⁵⁴

Este é um teste para verificar se uma pessoa é do tipo que forma uma intenção para agir e sustenta essa intenção mesmo se, na hora da ação, houver apenas razões contrárias à ação. Uma pessoa que tomasse a atitude de não reabrir a questão e fixasse o curso de ação até o fim, não seria irracional, mas bem sucedida, pois teria agido conforme a decisão que tomou ao receber *com antecedência* o prêmio. Se ela adotasse um plano supostamente racional de formar a intenção, receber o prêmio e na hora não agir conforme

¹⁵¹ Idem, p. 184. Blackburn inclui Gauthier (1986) e McLennen (1990) entre os que defendem escolhas fortemente dominadas no longo prazo – chamados de *teóricos da decisão normativa* (idem, p. 185), e Binmore (1994) entre aqueles que negam essa possibilidade.

¹⁵² Como visto no Capítulo 1, Ramsey propõe um método para se atribuir utilidades às ações, a partir da ordenação das preferências.

¹⁵³ Blackburn, 1998, p. 185.

¹⁵⁴ Idem, pp. 186-7. Citação de Kavka, ‘The Toxin Puzzle’, *Analysis*, 43 (1983), 33-6.

Ver estrutura semelhante à desse paradoxo no *problema de Newcomb*, no Capítulo 3.

a intenção, dificilmente poderia ser chamada de bem sucedida. Ela de fato não teria formado a intenção, se pensasse em revogá-la em seguida – e, nesse caso, o experimentador, sabendo disso, não lhe daria o prêmio.

O mesmo raciocínio desenvolvido para o paradoxo da toxina poderia ser aplicado ao dilema do prisioneiro. Uma pessoa pode vir a cooperar “não exatamente por causa de um princípio, ou por altruísmo, mas porque se sentiria desconfortável se, por exemplo, ela escolhesse não-cooperar enquanto outra pessoa cooperasse”¹⁵⁵. Tudo se passa como se, por educação ou por hábito, houvesse uma intenção para agir formada *antecipadamente*, uma utilidade que se realizaria¹⁵⁶ quando determinada situação surgisse.

A teoria de **Ramsey** “não nos diz se as pessoas têm certos valores [tais como cooperação, promessas, bem-estar, reputação, ou conforto], e menos ainda se é racional tê-los”¹⁵⁷. **Blackburn** propõe que, tal como **Ramsey** identifica os estados mentais, como crença e preferência, simplesmente pela sua função, deveríamos também aceitar que uma pessoa está *maximizando* quando tem interesses por uma variedade de coisas: “ela pode escolher de maneira diferente [de qualquer outra pessoa], mas ainda assim estará maximizando” algo¹⁵⁸.

Tanto **Sen** quanto **Blackburn** demonstram preocupação quanto ao uso que é feito da teoria dos jogos, principalmente devido ao pressuposto de racionalidade embutido no princípio de maximização da utilidade esperada. Aquilo que um agente considera como racionalidade pode não ser a mesma

¹⁵⁵ Blackburn, 1998, p. 189.

¹⁵⁶ Como mencionado no texto que faz referência à Nota 153.

¹⁵⁷ Idem, p. 190.

¹⁵⁸ Ibidem.

coisa que um teórico, de modo nem sempre sistemático, idealizou em seu modelo. Um terceiro ponto de vista, o de um observador que adota tal modelo como a representação perfeita do comportamento humano, não pode desconhecer o contexto e as simplificações que o teórico precisou fazer.

Essa discussão mostra a dificuldade de se representar na teoria dos jogos em geral, e no dilema do prisioneiro em particular, toda a riqueza da ação humana. A teoria descreve apenas uma parte da história, e ela pode evoluir não só devido à repetição, que corrige os defeitos, mas também à descoberta, por pensamento ou por observação, de outros interesses além daqueles considerados inicialmente. O conflito proposto pelo dilema do prisioneiro, que apresenta uma solução de equilíbrio onde cada indivíduo tem um incentivo para não cooperar, e que mostra essa mesma solução dominada por um resultado cooperativo, é sugestivo do tipo de problema com que estamos lidando: a racionalidade não comporta uma definição fixa. Qualquer definição será normativa em um primeiro momento, mas, por depender de elementos contingentes, estará sempre sujeita a revisão. Afinal, que natureza da racionalidade é essa?

CAPÍTULO 3 – NOZICK E A NATUREZA DA RACIONALIDADE

- 3.1 *As teorias evidencial e causal da decisão*
- 3.2 *O problema de Newcomb e os dois princípios de decisão*
- 3.3 *Uma questão de princípios*
- 3.4 *A teoria da ação de Nozick e o dilema do prisioneiro*
- 3.5 *De volta à racionalidade*

Até aqui vimos a teoria dos jogos tratar da tomada de decisão em situações estratégicas, isto é, situações em que os resultados individuais dependem das escolhas dos outros envolvidos. No presente capítulo, cuja parte inicial é dedicada à teoria da decisão, vamos examinar situações nas quais “a escolha de um indivíduo não afeta nem é afetada pelas escolhas de outros indivíduos”¹, porque estará voltada para a busca da melhor ação para alcançar determinado fim.

A discussão sobre a racionalidade, desenvolvida até aqui, vai ser completada com a contribuição do trabalho do filósofo Robert **Nozick** (New York, 1938-2002)², que também se interessou por outras áreas, como a filosofia política e a epistemologia³. Mais conhecido por suas idéias libertárias, apresentadas em *Anarchy, State and Utopia* (1974), **Nozick** as rejeitou parcialmente mais tarde, em pelo menos duas oportunidades: a primeira em *The Examined Life* (1980), e depois em *The Nature of Rationality* (1993). Em ambas, “o repúdio ao libertarismo é baseado na importância de *expressar com-*

¹ Ver Craig, verbete “*Decision and Game Theory*”, assinado por C. Bicchieri.

² Nozick teve o último livro (*Invariances*) publicado no final de 2001, poucos meses antes de seu falecimento.

³ Honderich (p. 30) inclui Nozick entre os praticantes da filosofia analítica, com a ressalva de que não há muita análise literal em seu trabalho.

promissos e significados simbólicos”⁴, conceitos ausentes naquele primeiro livro.

Em *The Nature of Rationality* (1993), nosso principal objeto de pesquisa, **Nozick** aborda a racionalidade da decisão, de cuja teoria procura superar os impasses. A proposta⁵ é combinar as duas visões teóricas disponíveis até então, a *causal* e a *evidencial*, e mais um terceiro elemento, o valor *simbólico* das ações. A soma dessas três parcelas, cada uma ponderada de acordo com a importância que o agente lhe atribui no contexto da ação, forma um novo conceito, o do *valor-decisão*, que pretende ser uma descrição completa, e por isso normativa, da decisão racional. É que, por ser completa, ela inclui os valores – éticos ou não – que o agente quer simbolizar por meio da ação.

Finalmente, a natureza da racionalidade vai ser explicada a partir de uma perspectiva evolucionária em que a razão, ao moldar e controlar suas próprias funções, inclui em si a própria Natureza.

⁴ Gaus, p. 106.

⁵ Ver Apêndice 3, “A genealogia da teoria da ação de Nozick”.

3.1 As teorias evidencial e causal da decisão

Vimos no Capítulo 2 que a teoria dos jogos, ou da decisão estratégica, estuda situações em que estão envolvidas ações interdependentes de dois ou mais agentes racionais que, individual e interativamente, procuram maximizar a utilidade própria pela escolha de uma dentre várias alternativas disponíveis, o que leva o jogo a uma solução de equilíbrio.

No presente capítulo vamos nos afastar – pelo menos inicialmente – dos jogos de competição/cooperação, para examinar a teoria da decisão proposta por **Ramsey**⁶ e desenvolvida por **Savage**, que se aplica a contextos não-estratégicos. Essa teoria descreve o comportamento isolado de um agente que em situações nas quais estão presentes à sua escolha, de um lado, um conjunto de ações alternativas e, de outro, um conjunto de possíveis estados⁷ do mundo, considera as incertezas sobre as conseqüências ou resultados de um ato. Essas características estão reunidas na Figura 3.1:

Matriz de decisão				
Ações	Condições ou Estados			
	S ₁	S ₂	...	S _n
a ₁	r ₁₁	r ₁₂	...	r _{1n}
a ₂	r ₂₁	r ₂₂	...	r _{2n}
...				
a _i	r _{i1}	r _{i2}	...	r _{in}
...				
a _m	r _{m1}	r _{m2}	...	r _{mn}

Figura 3.1 – A matriz de decisão, onde os r_{ij} representam os resultados das ações.⁸

⁶ Como já citado no Capítulo 1, Nota 3 e no Capítulo 2, Nota 28, a mesma teoria foi desenvolvida também por von Neumann e Morgenstern, independentemente de Ramsey.

A teoria dos jogos, vista anteriormente, pode ser considerada uma extensão para contextos estratégicos da teoria da decisão que vai ser aqui examinada.

⁷ Vamos considerar as expressões “resultados” e “conseqüências” como sinônimas. Da mesma forma, “estados”, que são partições de possibilidades do mundo, são descritos por “condições”, e assim também serão tratados como intercambiáveis.

⁸ Gärdenfors e Sahlin, p. 3.

A decisão sobre qual ação adotar é orientada pela regra bayesiana básica de que “deve ser escolhida uma alternativa que tenha a máxima utilidade esperada”⁹, cuja aplicação depende de dois principais tipos de fatores:

- Os nossos *desejos*, que determinam os *valores* ou *utilidades* [que atribuímos aos] possíveis resultados de nossas decisões.
- As nossas *informações* ou *crenças* sobre como é o mundo e como nossas possíveis ações influenciarão o mundo, que determinam as *probabilidades* dos possíveis resultados.¹⁰

Nessa teoria clássica da decisão, a medida de utilidade de um resultado reflete não somente as preferências do agente em termos de ordem, mas também “as diferenças em termos de valor *numérico*”¹¹. Da mesma forma, um agente, a partir de suas informações e crenças a respeito de uma dada situação, atribui probabilidades¹² únicas para cada estado do mundo, e essas probabilidades serão independentes da ação escolhida.

Essas medidas de utilidades e probabilidades – originadas dos desejos e crenças do agente – serão combinadas no conceito de “utilidade esperada”, que é a média das utilidades das possíveis conseqüências de um ato, ponderadas pelas probabilidades de ocorrência de cada estado. A nova equação da utilidade esperada, a ser maximizada, seria:

$$UE(a_i) = p(s_1) \cdot u(r_{i1}) + p(s_2) \cdot u(r_{i2}) + \dots + p(s_n) \cdot u(r_{in}) = \sum_j p(s_j) \cdot u(r_{ij}),$$

onde $\sum_j p(s_j) = 1$.

* * *

⁹ Gärdenfors e Sahlin, p. 5.

¹⁰ Idem, p. 1. Não estão considerados outros fatores, tais como o risco.

¹¹ Idem, p. 3.

¹² Como discutido no item 1.1 e mencionado na Nota 6 deste capítulo, a teoria original de Ramsey trabalha com as probabilidades subjetivas do agente.

O surgimento do problema de Newcomb¹³ revelou a deficiência da teoria clássica em lidar com situações em que os estados não são independentes das ações ¹⁴. Por exemplo, “obter cara no lançamento de uma moeda é independente de apostar em cara. (...) [Mas] tirar boas notas não é independente de estudar” ¹⁵. Além da dependência direta entre condições e ações, há casos de dependência indireta devida à existência de uma causa comum às ações e às condições, ou a algum tipo de antecipação da ação, tal como acontece na interação entre um governo e os agentes de mercado.

A mudança que permitiu considerar a probabilidade condicional de cada estado, dada cada ação, foi proposta por **Jeffrey**, que “revolucionou a estrutura” ¹⁶ da teoria da decisão ao introduzir uma matriz de probabilidades, paralelamente à matriz de utilidades, conforme a Figura 3.2:

Matriz de utilidades				
Ações	Condições ou Estados			
	S ₁	S ₂	...	S _n
a ₁	u ₁₁	u ₁₂	...	u _{1n}
a ₂	u ₂₁	u ₂₂	...	u _{2n}
...				
a _i	u _{i1}	u _{i2}	...	u _{in}
...				
a _m	u _{m1}	u _{m2}	...	u _{mn}

Matriz de probabilidades				
Ações	Condições ou Estados			
	S ₁	S ₂	...	S _n
a ₁	p ₁₁	p ₁₂	...	p _{1n}
a ₂	p ₂₁	p ₂₂	...	p _{2n}
...				
a _i	p _{i1}	p _{i2}	...	p _{in}
...				
a _m	p _{m1}	p _{m2}	...	p _{mn}

Figura 3.2 – A matriz de utilidades e a matriz de probabilidades ¹⁷

¹³ O problema de Newcomb, que provocou um novo desenvolvimento da teoria da decisão, foi proposto por Nozick em 1969 e será examinado no próximo item.

¹⁴ Um estado ou resultado *R* é considerado independente de um ato *A* quando a probabilidade condicional de *R*, dado *A*, for igual à probabilidade simples de *R*, isto é, $p(R/A) = p(R)$. A diferença entre $p(R/A)$ e $p(R)$ mede o grau de dependência de *R* em relação a *A*.

¹⁵ Resnik, p. 15.

¹⁶ Schmidt e Wright, p. 284.

¹⁷ A Figura 3.2 é uma generalização baseada em Jeffrey, pp. 5-6. Jeffrey prefere utilizar a expressão “desejabilidade”, em vez de “utilidade”, devido às “associações filosóficas enganosas – notadamente com o hedonismo de Bentham” (Jeffrey, p. 21) – que o termo “utilidade” tem. Entretanto, vamos considerar os dois termos sinônimos e manter o termo “utilidade”, de uso mais comum.

A cada resultado correspondem agora uma utilidade e uma probabilidade. A nova equação da utilidade esperada a ser maximizada será:

$$UE(a_i) = p(r_{i1}/a_i) \cdot u(r_{i1}) + p(r_{i2}/a_i) \cdot u(r_{i2}) + \dots + p(r_{in}/a_i) \cdot u(r_{in}) = \sum_j p(r_{ij}/a_i) \cdot u(r_{ij}),$$

onde $\sum_j p(r_{ij}/a_i) = 1$.

Dado esse passo, as propostas para capacitar o modelo clássico a enfrentar situações complexas seguiram duas abordagens: a teoria *evidencial* da decisão e a teoria *causal* da decisão, que divergem quanto ao tipo de probabilidade condicional a ser considerada nos cálculos. Essa escolha determina qual “utilidade esperada” será de fato maximizada.

Na primeira abordagem, desenvolvida por **Jeffrey**, a *teoria evidencial da decisão* propõe que $p(R/A)$ – a probabilidade de um resultado R condicionada à realização da ação A – seja o grau de crença do agente em R , na suposição de que A é verdadeira¹⁸. A teoria evidencial, inspirada em **Hume**, nega a possibilidade de uma relação de causalidade necessária entre eventos, pois as evidências indicariam apenas regularidades estatísticas ou correlações. A escolha pode ser descrita como uma “preferência por *notícias*”¹⁹ que resultariam das possíveis ações, e a utilidade esperada seria o “valor de notícia”²⁰ da respectiva ação. Em outras palavras, dizer que a utilidade esperada de A é maior que a de B “significa que o agente receberia as notícias de que A é verdadeira melhor que receberia as notícias de que B é verdadeira”²¹.

¹⁸ Essa “suposição que A ” seria uma questão de fato, não subjuntiva, tal como “*admitir que uma evidência é verdadeira*”, como propôs Ramsey, citado por Sahlin, p. 13. Ver Capítulo 1, Nota 16 e Figura 1.2.

¹⁹ Resnik, pp. 116-7.

²⁰ Fitelson, p. 2, citação de James Joyce.

²¹ Jeffrey, p. 82.

Na segunda abordagem, a *teoria causal da decisão* propõe que o peso seja não a probabilidade condicional simples²², mas a probabilidade condicional *causal* de um resultado R , dada a ação A , que significaria:

- o grau pelo qual um agente julga ser o caso de que A causalmente provoque R ;
- o grau de crença em R , na suposição subjuntiva de que A seja verdadeira²³;
- o grau de crença em uma proposição contrafactual ‘Eu faço $A \square \rightarrow R$ ’, que se lê ‘Se eu fizesse A , então aconteceria R ’, e que é representada por $p(A \square \rightarrow R)$ ²⁴.

As variantes da teoria causal têm em comum “a idéia de que a racionalidade de uma decisão depende do valor comparativo dos efeitos que a decisão *causaria* se fosse tomada, (...) consideradas as verdades prováveis das hipóteses causais relevantes”²⁵. É necessário, pois, conhecer a provável conexão causal entre a ação e cada possível consequência.

As recomendações das teorias evidencial e causal podem diferir da recomendação da teoria clássica, como exemplificado a seguir:

Suponha que haja uma correlação entre comer uma barra de chocolate (A) e dormir bem (R), isto é, $p(R/A) > p(R/não-A)$. Entretanto, suponha também que você saiba que essa correlação é devida à presença prévia de um hormônio oculto (H) que, independentemente, leva ao mesmo tempo a comer uma barra de chocolate e a ter um sono profundo, e que, portanto, ‘apaga’ qualquer associação direta entre essas ocorrências, isto é, $p(R/A\&H) = p(R/H)$ e $p(R/A\&não-H) = p(R/H)$. (...)

Nesse caso, a conclusão natural é que comer uma barra de chocolate (A) por si só não leva você a dormir bem (R). Esses dois eventos são efeitos conjuntos de uma causa comum, o hormônio (H). Comer uma barra de chocolate é um *sintoma* de que você vai dormir bem (desde que seja um sintoma de que você tem o H), mas não é a *causa* de um sono profundo.²⁶

²² Uma crítica às probabilidades condicionais simples é que elas “medem sinais potenciais”, mas deveriam medir “os potenciais prováveis das coisas como causas” (Sobel, 1994, p. 152), isto é, a tendência da ação em produzir ou prevenir um dado estado (Resnik, p. 112).

²³ As duas primeiras descrições são de James Joyce (2004), citadas por Fitelson.

²⁴ Gibbard e Harper, pp. 134-6. Nessa versão não se faz distinção “entre um ato que o agente pode realizar e a proposição que diz que ele está prestes a realizá-la. (...) Um *resultado* de um ato é uma proposição simples que, pelo que o agente sabe, expressa *todas* as consequências daquele ato que ele avalia” (idem, p. 137).

²⁵ Campbell, p. 31.

²⁶ Papineau, p. 168.

Traduzimos aqui “*Mars Bars*”, marca de uma barra de chocolate e caramelo, simplesmente por “barra de chocolate”. Esse exemplo antecipa algumas características do problema de Newcomb, que vai ser examinado no próximo item.

As situações desse exemplo estão reproduzidas na Figura 3.3:

	Você tem o hormônio H	Você não tem o hormônio H
Ação não-A: você não come a barra de chocolate	Você dorme bem.	Você não dorme bem
Ação A você come a barra de chocolate	Você come a barra de chocolate e dorme bem.	Você come a barra de chocolate e não dorme bem.

Figura 3.3 – A matriz de resultados do exemplo da “barra de chocolate”

A teoria clássica da decisão recomenda fazer a ação A, pois, como $p(R/A)$, o peso a ser aplicado à utilidade de R, dada a ação A, é maior que $p(R/não-A)$, o peso a ser aplicado à utilidade de R, dada a ação não-A, então a máxima utilidade esperada será obtida pela ação A.

As teorias evidencial e causal da decisão contrariam essa recomendação da teoria clássica. Apesar de produzirem recomendações semelhantes entre si, as duas novas teorias têm razões diferentes para a escolha.

A teoria causal da decisão recomenda não fazer a ação A. Ela

requer que ‘dividamos nossa classe de referência’ por todas as combinações de presença e ausência de outras possíveis causas do desejado resultado R, e que então ajamos segundo a média ponderada do efeito que a ação A faz a R dentro de cada célula dessa partição. (...) Agentes racionais necessitam considerar todas as diferentes maneiras nas quais A pode causalmente influenciar R, e então tomar a média ponderada desses diferentes efeitos que A pode causar a R, com os pesos correspondendo às probabilidades de cada tipo de efeito.²⁷

A teoria evidencial da decisão também recomenda não fazer a ação A.

Mas ela requer uma situação de “conhecimento total”, pela qual

os agentes sempre conhecerão algo extra (K) sobre si mesmos (...) que faz desaparecer as correlações espúrias entre A e R. (...) Suponha, por hipótese, que você sabia que tinha o hormônio H (ou que não tinha). Então, pelo princípio da total evidência, você deveria agir conforme o efeito probabilístico que A produz em R, considerado H (ou não-H). E, desde que todos concordam que tal valor é zero (...), a teoria evidencial da decisão evita recomendar que você coma a barra de chocolate.²⁸

²⁷ Idem, p. 170.

²⁸ Idem, p. 176.

Um outro exemplo em que as condições não são independentes das ações acontece quando o governo de um país cogita em expandir a base monetária para tentar diminuir o desemprego²⁹. Podem ocorrer quatro situações diferentes, em função de como o mercado – conjunto de pessoas e outros agentes econômicos – se comportará em relação à medida cogitada pelo governo, e que estão representadas na Figura 3.4:

	<i>O mercado prediz não-expansão (reação: nenhuma)</i>	<i>O mercado prediz expansão (reação: aumento de preços)</i>
<i>O governo não expande a base monetária</i>	Nada acontece (segundo melhor)	Recessão = não-expansão com inflação (pior resultado)
<i>O governo expande a base monetária</i>	Diminui o desemprego (melhor resultado)	Inflação (terceiro melhor)

Figura 3.4 – Um exemplo em que as condições não são independentes das ações ³⁰

A teoria evidencial recomendaria expandir a base monetária, pois considera que as expectativas do mercado já estariam formadas; então, qualquer que fosse a predição do mercado, é melhor diminuir o desemprego que manter a situação atual, e é melhor ter inflação que recessão.

A teoria causal³¹ recomendaria não expandir a base monetária, pois considera que é muito alta a probabilidade de o mercado acertar o que governo vai fazer; então, é preferível continuar como está a ter inflação.

²⁹ Broome, pp. 220-2.

³⁰ Idem, p. 221. A matriz foi adaptada a este trabalho. Este exemplo não tem a mesma simetria dos jogos estratégicos examinados no Capítulo 2. Aqui a ação do governo não depende necessariamente das ações do mercado; mas as expectativas do mercado, isto é, as condições, parecem depender do que o governo vai fazer.

³¹ A teoria clássica acompanha a teoria causal na recomendação de não expandir, pois supõe a independência das condições às ações e aplica a maximização da utilidade esperada.

3.2 O problema de Newcomb e os dois princípios de decisão

O *problema de Newcomb*³² apresentado abaixo tem um enunciado semelhante aos exemplos discutidos no item anterior:

Um ser, em cujo poder para prever corretamente você tem grande confiança, vai prever qual opção você vai escolher na seguinte situação. Há duas caixas, uma transparente e outra opaca. A caixa transparente contém \$1 mil; a caixa opaca pode conter \$1 milhão ou nada. Você tem de escolher uma das duas ações: (1) pegar o que estiver nas duas caixas, ou (2) pegar somente o que estiver na caixa opaca. Além disso, você sabe, e o ser sabe que você sabe, e assim por diante, que, se o ser prever que você pegará o que estiver nas duas caixas, ele não colocará o \$1 milhão na caixa opaca; se o ser prever que você pegará somente o que estiver na caixa opaca, ele colocará o \$1 milhão nela. Primeiro, o ser prevê a escolha que você vai fazer; em seguida, ele põe o \$1 milhão na caixa opaca ou não, dependendo da previsão que ele fizer; e então você faz a sua escolha.³³

A matriz de resultados desse problema é apresentada na Figura 3.5:

	<i>O ser prevê que você pega apenas uma caixa</i>	<i>O ser prevê que você pega as duas caixas</i>
<i>Você pega apenas uma caixa</i>	\$1.000.000	\$0
<i>Você pega as duas caixas</i>	\$1.001.000	\$1.000

Figura 3.5 – A matriz de resultados do problema de Newcomb³⁴

Durante quase duas décadas, esse problema foi visto como um conflito entre dois argumentos que levam a resultados diferentes: o do princípio da dominação e o do princípio da utilidade esperada.

O primeiro argumento, o do *princípio de dominação*, recomenda pegar as duas caixas, pois você ganha \$1 mil a mais qualquer que seja a previsão que o ser fizer. Supõe-se não haver nenhuma causalção retroativa em operação, isto é, a opção que você fizer agora não influencia a ação passada do

³² Esse problema foi criado em 1960 por William A. Newcomb, um físico da University of California (Nozick, 1993, p. 41). Foi inicialmente publicado em Nozick, 1969.

³³ Nozick, 1993, p. 41. As expressões originais “caixa B1” e “caixa B2” foram substituídas por “caixa transparente” e “caixa opaca”, respectivamente.

³⁴ Gardner, p. 416.

ser. As ações acontecem em uma seqüência determinada: o ser faz a predição, em seguida coloca ou não o \$1 milhão na caixa opaca, e aí você escolhe.

O segundo argumento, o do *princípio da utilidade esperada*, recomenda pegar apenas a caixa opaca. Esse critério maximiza o cálculo das utilidades esperadas das ações, em que as utilidades, aqui tomadas como iguais aos valores monetários, são multiplicadas pelas probabilidades condicionais de cada resultado, dada a ação. Admitindo-se uma probabilidade alta (por exemplo, 0,9) de o ser predizer corretamente, a opção de pegar apenas a caixa opaca terá a maior utilidade esperada³⁵, conforme o cálculo a seguir:

$$UE \text{ (você pega apenas uma caixa)} = (0,9 \times \$1.000.000) + (0,1 \times \$0) = \$900.000.$$

$$UE \text{ (você pega as duas caixas)} = (0,1 \times \$1.001.000) + (0,9 \times \$1.000) = \$101.000.$$

Este segundo argumento é mais bem visualizado na Figura 3.5', que é equivalente à figura anterior, mas apresenta uma nova partição dos estados do mundo em que não há evidência de qual ação é a dominante.

	<i>O ser prediz corretamente</i>	<i>O ser não prediz corretamente</i>
<i>Você pega apenas uma caixa</i>	\$1.000.000	\$0
<i>Você pega as duas caixas</i>	\$1.000	\$1.001.000

Figura 3.5' – A matriz de resultados do problema de Newcomb ³⁶, conforme uma nova partição dos estados do mundo

³⁵ Na definição da teoria clássica da decisão (item 3.1), que é a base para esse segundo argumento, a *utilidade esperada* de uma ação é o somatório das utilidades de cada resultado, multiplicadas pelas respectivas probabilidades condicionais, dada a ação.

No caso particular em foco,

$$UE(A) = p(R_1/A) \cdot u(R_1) + p(R_2/A) \cdot u(R_2),$$

onde $UE(A)$ é a utilidade esperada da ação A , $p(R_i/A)$ é a probabilidade condicional de R_i dada a ação A , e $u(R_i)$ é a utilidade desse resultado R_i . Ver também Sainsbury, p. 57.

³⁶ Schmidt e Wright, p. 283.

De um modo geral, a aplicação dos dois argumentos pode levar a escolhas erradas. No primeiro argumento, a aplicação do princípio de dominação dá um resultado falso se os estados não forem independentes das ações. Se houver essa dependência, isto é, se a escolha de uma ação determinar qual estado é obtido, deve-se utilizar o outro método, em que estão presentes as probabilidades condicionais dos estados, dadas as ações. O erro da aplicação do raciocínio de dominação nesse caso é exatamente ignorar a existência das probabilidades condicionais dos estados.

O segundo argumento, o da utilidade esperada, que aplica a teoria clássica da decisão, pode também levar a resultados errados se as ações e os estados estiverem relacionados a uma causa comum, como no exemplo da barra de chocolate³⁷, ou se o *sentido da explicação* for inverso ao que as probabilidades condicionais indicarem, como será visto na análise a seguir.

* * *

O tratamento que **Nozick** dá ao problema de Newcomb evolui em três momentos, segundo **Lacey**³⁸. Inicialmente, em sua tese de doutorado (1963), preferiu a solução da maximização da utilidade esperada, porque o conteúdo das caixas não é probabilisticamente independente da escolha. Mais tarde, no artigo “Newcomb’s Problem and Two Principles of Choice” (1969), reapresentado em *Socratic Puzzles*, **Nozick** preferiu a solução dominante, porque passou a distinguir dependência probabilística de influência causal. Finalmente, em *The Nature of Rationality* (1993), ele aceita que “o conflito é realmente entre a utilidade esperada evidencial e a causal, não entre a utilidade

³⁷ O exemplo da barra de chocolate é apresentado no texto que faz referência à Nota 26.

³⁸ Lacey, p. 158.

esperada e a dominação”³⁹, e trata tal conflito como dois componentes do “valor-decisão”, um novo conceito que será descrito mais à frente.

No tratamento inicial, de 1963, a questão era se a escolha afetaria ou não as probabilidades dos estados dentre os quais a escolha é feita. **Nozick** sugeriu na ocasião, segundo **Lacey**, que “se e somente se os estados em questão forem probabilisticamente independentes das escolhas que eu fizer, deveria fazer a ação dominante (...). Caso contrário, deveria seguir a utilidade esperada dos resultados”⁴⁰. Embora estivesse confiante em aplicar esse critério a outros exemplos nos quais havia uma clara relação causal entre a escolha feita e o estado a ser obtido, **Nozick** estaria, no caso do problema de Newcomb, “impressionado com a força do argumento dominante”⁴¹.

No segundo tratamento, de 1969, no artigo em que publica pela primeira vez o problema de Newcomb, **Nozick** admite que “o argumento dominante é de fato a abordagem correta”⁴². A razão para a mudança é que a questão crucial diria respeito não à *dependência probabilística*, mas à noção de *influência causal*. A distinção entre esses conceitos aparece nos casos em que dois eventos têm uma causa comum, e por isso são probabilisticamente dependentes, mas nenhum deles tem influência causal sobre o outro. Nesse artigo de 1969, de cuja subsequente discussão surgiram as abordagens evidencial e causal no âmbito da teoria da decisão, **Nozick** procura estabelecer em que casos o princípio de dominação pode ser aplicado⁴³.

³⁹ Idem, p. 159. Referência à proposta original de Gibbard e Harper (p. 146).

⁴⁰ Idem, p. 143. Referência à tese de doutorado de Nozick, *The Normative Theory of Individual Choice*, obra não disponível para consulta.

⁴¹ Idem, p. 144. Mesma referência citada na nota anterior.

⁴² Lacey, p. 144.

⁴³ Ver no item 3.1 uma discussão a respeito das teorias evidencial e causal da decisão.

Um primeiro passo no argumento, compatível com a análise de 1963, é que “não se deve aplicar o princípio de dominação a uma situação na qual os estados não sejam probabilisticamente independentes das ações”⁴⁴. Isso é demonstrado nos casos em que a ação dominante não é a melhor opção:

Exemplo: em uma corrida entre dois cavalos, temos os estados S_1 = o primeiro cavalo vence, e S_2 = o segundo cavalo vence, bem como as ações A_1 = aposto no primeiro cavalo, e A_2 = aposto no segundo cavalo. A minha probabilidade de que o primeiro cavalo vença é de 0,2, e a de que o outro vença é de 0,8. Assim, a utilidade esperada de A_1 será $0,2 \times \$50 + 0,8 \times (-\$5)$ [= \$6], e a utilidade esperada de A_2 será $0,2 \times (-\$6) + 0,8 \times \49 [= \$38]. Então, a ação A_2 , de maior utilidade esperada, será a escolhida. Mas posso também definir uma segunda partição, que introduz uma dependência probabilística dos estados às ações, na qual S_3 = eu tenho sorte, e S_4 = não tenho sorte, e essa nova partição faz surgir uma ação dominante (A_1), mas esta claramente leva a um resultado pior.

1ª partição		
Ação	S_1	S_2
A_1	ganho \$50	perco \$5
A_2	perco \$6	ganho \$49

2ª partição		
Ação	S_3	S_4
A_1	ganho \$50	perco \$5
A_2	ganho \$49	perco \$6

Figura 3.6 – Exemplo da corrida de cavalos⁴⁵

Em outro exemplo de dependência dos estados, já há uma ação dominante disponível, mas novamente ela não é a ação recomendável:

Exemplo: em uma roleta, temos os estados S_1 = o preto vence, e S_2 = o vermelho vence, bem como as ações A_1 = aposto no preto e A_2 = aposto no vermelho. O dono do cassino ordena que o preto vença a próxima rodada se eu escolher A_1 e que o vermelho vença se eu escolher A_2 . Como sei que há essa dependência dos estados à escolha que eu fizer, não posso escolher a ação dominante A_1 , como mostrado na Figura 3.7 (1ª partição). Uma partição alternativa seria S_3 = o operador obedece ao dono, e S_4 = o operador não obedece ao dono. Se eu considerar a probabilidade de S_3 suficientemente alta, a melhor opção será fazer A_2 , e não a ação dominante A_1 .

1ª partição		
Ação	S_1	S_2
A_1	ganho \$10	ganho \$100
A_2	ganho \$5	ganho \$90

2ª partição		
Ação	S_3	S_4
A_1	ganho \$10	ganho \$100
A_2	ganho \$90	ganho \$5

Figura 3.7 – Exemplo da roleta⁴⁶

⁴⁴ Nozick, 1969, p. 55.

⁴⁵ Idem, pp. 52-3.

⁴⁶ Idem, p. 54.

Em um segundo passo da análise de 1969, **Nozick** destaca o caso particular em que os estados, embora não sejam probabilisticamente independentes das ações, já estejam fixados e determinados:

Exemplo: *P* sabe que *S* ou *T* é seu pai, mas não sabe qual. *S* morreu de uma doença hereditária, e *T* não. Sabe-se que essa doença é geneticamente dominante e que a mãe de *P* não tinha esse gene (...). Se *S* for o pai, *P* morrerá dessa doença; se *T* for o pai, *P* não morrerá dessa doença. Além disso, há uma teoria (...) sobre a transmissão genética de uma tendência de preferir tarefas de uma vida intelectual. *S* tinha essa tendência, que é geneticamente dominante. (...) *P* vai decidir se (*A*) vai seguir uma carreira acadêmica ou (*B*) vai se tornar um atleta. Ele prefere a vida acadêmica, mas nem tanto. O princípio de dominação manda fazer *A*, mas como os estados não são probabilisticamente independentes das ações, não se deve usar tal princípio. Por outro lado, a utilidade esperada de *B* é maior que a de *A*, o que levaria *P* a fazer a seguinte afirmação maluca: “Estou fazendo *B* porque assim é menos provável que eu morra da doença mortal”. O fato é que qual deles é o pai já está fixado e determinado há muito tempo. A ação escolhida afeta legitimamente a estimativa da probabilidade dos dois estados, mas qual estado é obtido não depende nada da ação.

Matriz de utilidades		
Ação	<i>S</i> é o pai de <i>P</i>	<i>T</i> é o pai de <i>P</i>
<i>A</i>	-20	100
<i>B</i>	-25	95

Matriz de probabilidades		
Ação	<i>S</i> é o pai de <i>P</i>	<i>T</i> é o pai de <i>P</i>
<i>A</i>	0,9	0,1
<i>B</i>	0,1	0,9

Figura 3.8 – Exemplo da tendência hereditária ⁴⁷

Assim, nas situações em que os estados, apesar de não serem probabilisticamente independentes das ações, já estejam fixados e determinados, “parece legítimo usar o princípio da dominação” ⁴⁸. Mas, mesmo nesses casos, quando o princípio de utilidade esperada recomenda fazer uma ação e o princípio de dominação recomenda outra, permanecem as dúvidas sobre qual ação escolher, como é mostrado no seguinte exemplo:

Exemplo: Introduzindo um pouco de reflexividade no exemplo anterior, suponha que haja duas tendências hereditárias:

- (1) Uma tendência em pensar que o princípio da utilidade esperada deveria ser usado em tais situações. (Se *P* tiver essa tendência, ele estará no estado *S*₁).
- (2) Uma tendência hereditária em pensar que o princípio da dominação deveria ser usado em tais situações. (Se *P* tiver essa tendência, ele estará no estado *S*₂.)

⁴⁷ Idem, pp. 56-7.

⁴⁸ Idem, p. 58.

A mãe de P tem genes neutros. Um dos homens que podem ser o pai de P tem os genes da primeira tendência. O outro homem que também pode ser o pai de P tem os genes da segunda tendência. Sabe-se que P tem uma das duas tendências, mas não se sabe qual. Uma doença mortal está se espalhando e há duas vacinas [A e B] que são efetivas contra a doença. Os efeitos colaterais da vacina B para cada pessoa são piores que os da vacina A , e cada vacina tem efeitos colaterais piores nas pessoas que estiverem em S_2 . Além disso, P sabe, a partir das escolhas das pessoas *nessa* situação, quais são as probabilidades de ter cada uma das tendências, dado que ele tem uma das duas. Que ação P deveria escolher, dadas as matrizes abaixo?

Uma primeira resposta é que se P escolher A , muito provavelmente S_2 é obtido, e P ganhará 4. Se P escolher B , muito provavelmente S_1 acontecerá, e P ganhará 8. Então, ele escolhe B . Mas uma segunda resposta diz que já está fixado e determinado qual estado S_1 ou S_2 vai acontecer. O que P fizer não provocará um ou outro estado.

Matriz de utilidades		
Ações	S_1	S_2
A	10	4
B	8	3

Matriz de probabilidades		
Ações	S_1	S_2
A	0,1	0,9
B	0,9	0,1

Figura 3.9 – Exemplo das vacinas ⁴⁹

Pelos dados do exemplo, quem estiver no estado S_1 terá uma forte tendência a optar pela ação B , por ser esta a de maior utilidade esperada, apesar de haver uma pequena probabilidade de escolher o outro princípio; e quem estiver no estado S_2 terá uma forte tendência a optar pela ação A , por ser esta a ação dominante. Mas não se pode dizer o contrário, isto é, quem escolhe a ação B está no estado S_1 e quem escolhe a ação A está no estado S_2 . Esse exemplo revela a importância do sentido causal em uma explicação:

Se a explicação tiver o sentido de x para y , uma teoria explicativa correta falará de uma probabilidade condicional $p(y/x)$. (...) [Por outro lado], dado que a linha explicativa vai de x para y (por x ser parte da explicação de y) e não de y para x , (...) a probabilidade $p(x/y)$ será uma *verossimilhança*. ⁵⁰ (...) A observação das verossimilhanças dos estados, dadas as ações, talvez nos dê a ilusão de controle sobre os estados.

(...) Quando os estados já estiverem fixados e determinados, e a teoria explicativa descrever uma influência que vá dos estados às ações, de forma que as probabilidades dos estados, dadas as ações, sejam verossimilhanças, então se o princípio de dominação se aplicar, ele deve ser aplicado. ⁵¹

⁴⁹ Idem, pp. 58-9.

⁵⁰ Idem, p. 60. O conceito de verossimilhança é também tratado no Apêndice 1, itens 2 e 3.

⁵¹ Idem, pp. 60-1.

No exemplo das vacinas, a linha explicativa vai dos estados às ações, e assim as probabilidades condicionais corretas seriam do tipo $p(\text{ação}/\text{estado})$. Entretanto, a Figura 3.9 apresenta as probabilidades $p(\text{estado}/\text{ação})$ que, no caso, são verossimilhanças⁵², pois descrevem um sentido causal que vai das ações aos estados, oposta à linha explicativa original. Como mencionado anteriormente neste item, esta é a falha do argumento da utilidade esperada e, por extensão, da teoria clássica da decisão⁵³.

A conclusão de **Nozick**, antes de levar o raciocínio ao problema de Newcomb, é que “o fato crucial não é se os estados já estão fixados e determinados, mas se as ações influenciam ou afetam que estado é obtido”⁵⁴. As possibilidades examinadas nos exemplos estão resumidas na Figura 3.10:

<i>Descrição da situação</i>	<i>Há uma ação dominante disponível</i>		<i>Não há uma ação dominante disponível</i>	
<i>As ações influenciam qual estado é obtido.</i>	(I)	Maximize a utilidade esperada	(II)	Maximize a utilidade esperada
<i>Não há influência das ações nos estados. As probabilidades condicionais diferem de uma ação a outra.</i> ⁵⁵	(III)	[Newcomb]	(IV)	
<i>Não há influência das ações nos estados. As probabilidades condicionais são iguais para todas as ações.</i>	(V)	Execute a ação dominante (ou maximize a utilidade esperada)	(VI)	Maximize a utilidade esperada

Figura 3.10 – Recomendações da 2ª análise de Nozick⁵⁶

⁵² Ver Apêndice 1, item 3: a verossimilhança é uma “tendência à necessidade” ou um sinal de relevância estatística na relação entre uma ação e um estado.

⁵³ Segundo Schmidt e Wright, p. 291, “aplicar o raciocínio da utilidade esperada quando se tem apenas uma suspeita (...) é falacioso, mas não necessariamente uma má idéia. (...) Ele pode ser uma boa heurística”.

⁵⁴ Nozick, 1969, p. 63.

⁵⁵ Como visto em 3.1, o texto de Nozick (1969) é anterior ao de Jeffrey (1983), e neste as probabilidades dos estados podem ser diferentes para cada ação.

⁵⁶ Ibidem. Gibbard e Harper (p. 146) sugerem incluir uma quarta linha, a de influência causal sem dependência probabilística (citado também em Nozick, 1993, p. 188, n. 2).

Quase todas as situações do quadro anterior são bem resolvidas pela teoria clássica da decisão, que manda maximizar a utilidade esperada. A exceção é a linha do meio, em que está o problema de Newcomb. Nesse caso, as probabilidades condicionais indicam uma tendência da ação em influenciar a obtenção de um determinado estado. Mas isso pode não ocorrer, pois “se a nossa escolha de pegar apenas uma caixa *tornasse* mais provável que o ser poria o \$1 milhão nessa caixa, então isso faria parte da utilidade global da ação”⁵⁷. Assim, “se *soubermos* que não há tal influência causal, a dependência probabilística [que é um sinal dessa influência causal] torna-se irrelevante”⁵⁸.

Naquele momento, **Nozick** recomendou que, no caso do problema de Newcomb, “dever-se-ia escolher a ação dominante, e ignorar as probabilidades condicionais que não indicam uma influência”⁵⁹. Ou seja,

se você acredita (...) que haja causação retroativa, que a sua escolha causa o dinheiro estar lá ou não, que isso causa [o ser] ter feito a predição que ele fez, então não há problema. Você deve tomar somente o que estiver na caixa [opaca]. (...) Mas suponha que (...) não haja causação retroativa, que o que você realmente decidir fazer não afeta o que o ser predisse no passado, que o que você realmente decidir fazer não é parte da explicação de por que ele fez tal predição. (...) [Nesse caso,] acredito que você deva pegar o que estiver nas duas caixas.⁶⁰

Para **Nozick**, o problema de Newcomb não é um caso claro de influência ou de não-influência da ação sobre o estado, mas envolve uma “*ilusão* de

⁵⁷ Schmditz e Wright, p. 286.

⁵⁸ Idem, p. 288.

⁵⁹ Nozick, 1969, pp. 63-4.

⁶⁰ Idem, p. 65.

Gardner (p. 417) faz a seguinte leitura dessa citação: “se você acreditar no determinismo absoluto e se o ser predisser o seu comportamento com uma precisão inquestionável”, você deve pegar apenas uma caixa. Mas, se o ser não tiver tal perfeição – como foi pressuposto no enunciado do problema – e “se você acreditar que possui pelo menos uma quantidade mínima de livre-arbitrio”, você deve pegar as duas caixas.

influência”⁶¹. A razão para isso talvez seja que, “embora a explicação sobre o que está na caixa opaca não se refira diretamente à nossa escolha, ela o faz indiretamente, ao mencionar as crenças do ser a respeito dessa escolha”⁶².

Mesmo se a probabilidade de o ser predizer corretamente for um pouco maior que 0,50, a opção de pegar apenas uma caixa continuará ainda a ter maior utilidade esperada que a opção de pegar as duas caixas. **Lacey** observa que a discussão apresentada por **Nozick** reforça uma relação que já era – por sua própria definição – sabida, a de que “a utilidade esperada não é um tipo de utilidade, mas um amálgama de utilidade e probabilidade”⁶³.

* * *

O terceiro tratamento que **Nozick** dá ao problema de Newcomb aparece em *The Nature of Rationality*, de 1993. Como mencionado anteriormente, ele agora aceita a sugestão de **Gibbard** e **Harper** de que “o conflito é realmente entre a utilidade esperada causal e a evidencial, não entre a utilidade esperada e a dominação”⁶⁴. Essa proposição tem origem nas seguintes definições:

$$UEC(A) = \sum_j p(A \square \rightarrow R_j) \cdot u_j \quad \text{e} \quad UEE(A) = \sum_j p(R_j/A) \cdot u_j,$$

onde a *utilidade esperada causal* é calculada com probabilidades de contrafactuais, e a *utilidade esperada evidencial* é calculada com probabilidades condicionais⁶⁵.

⁶¹ Nozick, 1969, p. 67.

⁶² Lacey, p. 144.

⁶³ Idem, p. 145. O resultado persistirá enquanto a probabilidade for maior que 0,5005.

⁶⁴ Idem, p. 159. Referência à proposta original de Gibbard e Harper (pp. 135 e 150).

⁶⁵ Gibbard e Harper, p. 137. A notação foi adaptada a este trabalho.

As teorias evidencial e causal da decisão foram inicialmente discutidas no item 3.1.

De um lado, $UEC(A)$ “é uma medida da eficácia esperada de um ato em produzir os estados de coisas que um agente deseja; ela mede o valor esperado das conseqüências de um ato”⁶⁶. Por outro lado, $UEE(A)$ mede a receptividade da notícia de que se está prestes a realizar a ação A , e tal notícia tende a ser bem-vinda não só porque é provável que o ato traga conseqüências desejáveis, mas também porque a decisão por sua execução é uma evidência de que “o mundo está em um estado desejado”⁶⁷.

Gibbard e **Harper** defendem a idéia de que $UEC(A)$, que suporta a escolha das duas caixas, é “a genuína utilidade esperada de A ”⁶⁸, enquanto a noção de $UEE(A)$, que sustenta a escolha de se pegar apenas a caixa opaca,

parece ser ajudada por uma certa confusão. Suponha que ‘ X é um ato corajoso, então vou fazê-lo’: estaria *mostrando* que *já* sou corajoso (...) ou tentando *me tornar* corajoso? Talvez ambos tenham alguma utilidade, mas são [coisas] muito diferentes.⁶⁹

Repetindo o cálculo do início deste item, mas ampliando-o agora com a ótica dos modelos evidencial e causal, temos a seguinte comparação:⁷⁰

a) **Teoria evidencial da decisão**, que usa probabilidades condicionais simples:

$$\begin{aligned} UEE \text{ (pegar apenas a caixa opaca)} &= (0,9 \times \$1.000.000) + (0,1 \times \$0) = \$900.000. \\ UEE \text{ (pegar as duas caixas)} &= (0,1 \times \$1.001.000) + (0,9 \times \$1.000) = \$101.000, \end{aligned}$$

onde 0,9 é a probabilidade de o ser predizer corretamente.

b) **Teoria causal da decisão**, que usa probabilidades condicionais contrafactuais:

$$\begin{aligned} UEC \text{ (pegar apenas caixa opaca)} &= (\$1.000.000 \mu) + [\$0 (1-\mu)] = \$1.000.000 \mu. \\ UEC \text{ (pegar as duas caixas)} &= (\$1.001.000 \mu) + [\$1.000 (1-\mu)] = \$1.000.000 \mu + \$1.000, \end{aligned}$$

onde μ é probabilidade de haver \$1.000.000 na caixa opaca. Então, se o resultado é causalmente independente das ações, $p(A \square \rightarrow R) = \mu$ para qualquer ação.

⁶⁶ Idem, p. 144.

⁶⁷ Idem, p. 143.

⁶⁸ Idem, pp. 137 e 145.

⁶⁹ Lacey, p. 146.

⁷⁰ Gibbard e Harper, pp. 150-1.

Os resultados em (b) confirmam que o modelo causal, tomado inicialmente como aplicação do princípio de dominação⁷¹, é também justificado pela maximização da utilidade *causal* esperada. Os resultados dos cálculos em (a) retratam o argumento, defendido pelo modelo evidencial, de que

as pessoas que escolhem apenas uma caixa tendem a se tornar ricas, enquanto as que escolhem as duas caixas, não. Seria uma boa notícia para você saber que é o tipo de pessoa que pega apenas uma caixa, porque significaria que o ser pôs o milhão na caixa, e você provavelmente está prestes a ficar rico.⁷²

Nessa terceira discussão, de 1993, apesar de aceitar a distinção entre as utilidades evidencial e causal, **Nozick** não concede uma prevalência de uma utilidade sobre a outra, mas combina as duas visões, dando um certo peso a cada uma, para formar um novo conceito, o de *valor-decisão*.

O ponto inicial do raciocínio é que a nossa intuição pode variar com a quantia que é colocada na caixa transparente⁷³: quem preferisse a opção de pegar as duas caixas poderia se sentir desconfortável caso o valor na caixa transparente fosse muito reduzido para, por exemplo, alguns centavos. Por outro lado, quem preferisse a opção de pegar apenas a caixa opaca ficaria tentado a mudar se o valor na caixa transparente se tornasse muito elevado, próximo, por exemplo, ao \$1 milhão.

Ninguém teria uma “*completa* confiança no argumento escolhido”⁷⁴. Essas diferentes doses de confiança podem ser representadas, por exemplo, por pesos que variam entre 0 e 1, e que somam 1. Uma pessoa daria um peso W_c para multiplicar a utilidade causalmente esperada de uma ação A , e

⁷¹ Segundo Gibbard e Harper (pp. 148-49), haveria ainda duas versões de dominação, correspondentes a cada tipo de utilidade, ambas adotadas sem distinção por Nozick.

⁷² Hurley, p. 65. A pergunta *evidencial*, que desafia a recomendação da teoria causal e do princípio de dominação, seria: “Se você é tão esperto, por que não é rico?” (ibidem).

⁷³ Nozick, 1993, pp. 44-45.

⁷⁴ Idem, p. 45.

um peso We para multiplicar a utilidade evidencialmente esperada dessa ação A ⁷⁵. A soma desses dois produtos será o valor-decisão da ação A , ou $VD(A)$, que guiará a escolha da ação que tiver o máximo VD :

$$VD(A) = Wc \cdot UEC(A) + We \cdot UEE(A). \text{ } ^{76}$$

Esse novo modelo, apesar de ter nascido no bojo da aplicação do problema de Newcomb a uma situação em que a quantia colocada na primeira caixa oscilava radicalmente, não revela uma questão de insegurança em qual princípio adotar. Ao contrário, a equação sinaliza que

ambos os princípios são legítimos e cada um deve receber o seu devido valor. Os pesos, então, não são medidas de incerteza, mas medidas da força legítima que cada princípio tem. Assim, trata-se de uma teoria *normativa* que dirige a pessoa à escolha de uma ação com máximo valor-decisão. ⁷⁷

Limitado provisoriamente aos dois princípios formulados pela teoria causal e pela teoria evidencial da decisão, o conceito de valor-decisão vai ser ampliado mais à frente com a adição de outro princípio: o da *utilidade simbólica*.

⁷⁵ Nozick esclarece que esses pesos não são fixos para cada tipo de utilidade, mas que podem variar conforme a situação de decisão que se enfrenta (idem, p. 54).

⁷⁶ Idem, p. 45.

⁷⁷ Ibidem.

3.3 Uma questão de princípios

A racionalidade é uma característica dos homens, que podem controlar sua conduta por meio da utilização de princípios. A racionalidade é descrita pela teoria da evolução como um traço dentre outros; e a teoria da decisão estuda a ação racional em um contexto amplo.

A teoria da ação de **Nozick** parte da importância que os princípios têm no nosso dia-a-dia. Ao produzir razões, os princípios assumem funções intelectuais e nos justificam perante os outros. Quando permitem prever e coordenar ações, eles assumem funções interpessoais e garantem aos outros a nossa conduta. Ao orientar nossa própria conduta, assumem funções intrapessoais e nos obrigam a agir de uma determinada maneira. Ao fundamentar nossa própria ação, os princípios definem nossa identidade.

Os princípios são rótulos gerais para agrupar as ações que podem ser tratadas de uma mesma maneira. Essa generalização aponta para diversas funções, como exemplificado na Figura 3.11:

FUNÇÕES ←	PRINCÍPIOS ← (rótulos que agrupam ações)	AÇÕES
• <i>Intelectuais</i> : justificam-nos perante os outros.	Testam juízos, produzem razões e convencimentos. Evitam injustiças comparativas.	
• <i>Interpessoais</i> : garantem aos outros nossa conduta frente a tentações.	Dão motivos para prever condutas e produzir ações coordenadas.	
• <i>Intrapessoais</i> : limitam-nos (como <i>ataduras</i>) a agir segundo princípios.	Ajudam a orientar nossa própria conduta e a vencer tentações.	
• <i>Pessoais</i> : definem nossa própria identidade.	Fundamentam nossa ação pessoal: os princípios são componentes internos de nossa identidade.	

Figura 3.11 – As funções intelectuais, interpessoais, intrapessoais e pessoais dos princípios.

78

⁷⁸ Figura baseada em Nozick, 1993, pp. 4-14. É um esquema simplificado, pois um mesmo princípio pode ter mais de uma função.

Esse modo de relacionar ações com princípios vai ser aplicado a dois temas essenciais para o argumento de **Nozick**: como vencer as tentações e como considerar os custos incorridos (*sunk costs*).

A principal função intrapessoal que **Nozick** discute⁷⁹ é a relativa aos princípios que se contrapõem às tentações, ou que ajudam a superar obstáculos. A idéia básica é que, tanto psicologicamente quanto economicamente, “descontamos” o futuro⁸⁰, isto é, a utilidade que uma recompensa futura tem para nós no presente é menor do que quando a recebemos. A maneira que a teoria da decisão desconta o futuro segue uma curva exponencial, cuja característica principal é subtrair uma proporção constante da utilidade de cada período, calculada pela taxa de desconto.

As pesquisas de George **Ainslie**, que influenciaram o pensamento de **Nozick**⁸¹, mostram que nem sempre uma curva exponencial reproduz o comportamento humano, especialmente quando tendências opostas produzem oscilações temporárias nas preferências. Um exemplo extremo⁸² é o de uma pessoa que sacrifica a recompensa maior de uma vida longa e saudável na abstinência de um vício, em troca da satisfação de um desejo cuja recompensa é mais baixa, mas que tem utilidade maior no momento de sua realização. **Nozick** procura generalizar para dois projetos concorrentes, com re-

⁷⁹ Idem, pp. 14-21.

⁸⁰ O conceito de desconto foi introduzido no texto associado às Notas 100-103, Capítulo 2.

⁸¹ Nozick (1993, p. 184, n. 16) se refere a obras anteriores de Ainslie (“Specious Reward: A Behavioral Theory of Impulsiveness and Impulse Control”, *Psychological Bulletin* (1975) e “Beyond Microeconomics”, in Jon Elster, *The Multiple Self* (1986)). Neste trabalho, preferimos acompanhar a obra mais recente de Ainslie, *Breakdown of Will* (2001), não só por ser uma atualização daquela pesquisa, como por reforçar o argumento desenvolvido por Nozick.

⁸² Ainslie, p. 29. Outro exemplo menos dramático é o conflito entre a decisão de longo prazo de fazer uma dieta e a vontade momentânea de tomar um sorvete (idem, p. 42).

compensas e prazos diferentes⁸³, cuja oscilação de preferências é representada por curvas de desconto que se cruzam, como na Figura 3.12:

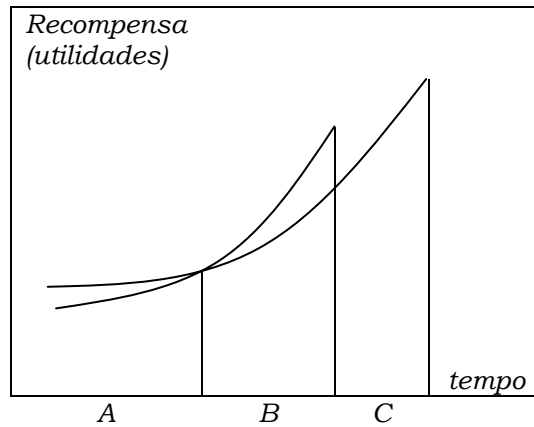


Figura 3.12 – Curvas de desconto que se cruzam. ⁸⁴

No ponto inicial das curvas, que corresponde ao momento em que uma decisão é tomada, a utilidade descontada da recompensa de longo prazo se mostra maior do que a da recompensa imediata. Essa superioridade indica a preferência do agente pela primeira opção, a de longo prazo, e justifica a escolha racional a ser feita. Entretanto, na medida em que se aproxima o momento de realização da recompensa de curto prazo, o valor descontado desta passa a ser o maior, representando uma *tentação* para que seja revista a decisão tomada inicialmente. ⁸⁵

Destacam-se na figura três períodos: o primeiro (A), quando a utilidade da recompensa de longo prazo prevalece; o intermediário (B), que é o *período da tentação*, em que sobressai a recompensa imediata; e o terceiro período (C), quando a preferência de longo prazo volta a prevalecer e se estabiliza.

⁸³ Nozick, 1993, p. 15.

⁸⁴ Idem, p. 16.

⁸⁵ O paradoxo da toxina, citado por Blackburn, tem estrutura semelhante ao que está sendo apresentado aqui. Ver texto que faz referência à Nota 154, Capítulo 2.

É possível que as curvas exponenciais de desconto de duas recompensas se cruzem, mas para isso é necessário que haja diferenças não só entre os valores das utilidades envolvidas, mas também entre as respectivas taxas de desconto⁸⁶, o que nem sempre é factível para uma mesma pessoa⁸⁷. **Ainslie** vê vantagens em substituir as exponenciais por funções hiperbólicas, porque elas representariam melhor um tipo de comportamento “contra o próprio interesse (*self-defeating*)”⁸⁸, como é o caso de seguir as tentações.

A inclinação das curvas hiperbólicas é mais acentuada que a das exponenciais, pois o seu decaimento é inversamente proporcional ao período em que o desconto é aplicado⁸⁹. Isso equivale a dizer que as utilidades descontadas crescem rapidamente quando próximas de sua realização. Ao depender inversamente da “demora” da recompensa, as curvas hiperbólicas indicam “que você tende a preferir, temporariamente, recompensas menores e precoces, a outras maiores e tardias, durante o tempo em que as primeiras forem iminentes”⁹⁰. Entretanto, por não trabalharem com taxas de desconto, elas criam, em casos de comparação entre pessoas diferentes, uma “rigidez que necessita correção”⁹¹. Mesmo assim, no julgamento de **Ainslie**, “ao explicar a sedução das escolhas contrárias ao próprio interesse”, o desconto hiperbólico “dá à teoria da utilidade o elemento que lhe faltava”⁹².

⁸⁶ Ainslie, p. 29. Ver no Apêndice 4 uma comparação entre essas curvas.

⁸⁷ A taxa de desconto que uma pessoa adota é um valor que depende, entre outras coisas, da sua expectativa de vida, isto é, de quanto ela valoriza o futuro.

⁸⁸ Idem, pp. 30 e 206, n. 6. É citado Hume, *Essays, Moral, Political, and Literary*, “a impaciência por riqueza torna os homens mais pobres no longo prazo”.

⁸⁹ Idem, p. 35.

⁹⁰ Idem, p. 38.

⁹¹ Idem, p. 207.

⁹² Idem, p. 39. O desconto hiperbólico também explicaria a sedução da *akrasia*, que é o comportamento sem autocontrole.

Assim, “as pessoas podem ter uma variedade de preferências contraditórias que se tornam dominantes em diferentes pontos, de acordo com o seu *timing*”⁹³. Um agente que “desconta hiperbolicamente” é visto como uma “sucessão de estimadores” que têm conclusões diferentes – isto é,

no decorrer do tempo, esses estimadores mudam sua relação uns com os outros, desde a cooperação em um objetivo comum à competição em objetivos mutuamente exclusivos. O Ulisses que planeja o encontro com as sereias deve tratar o Ulisses que as ouve como uma pessoa à parte, ou para ser influenciada ou, se isso não for possível, para ser evitada. (...) [A] mente [de uma pessoa] mantém uma multidão de processos de busca de recompensas que cresceram para sobreviver em contradição uns com os outros, mas que se toleram.⁹⁴

O interesse por uma recompensa de longo prazo só se manterá se a pessoa “for capaz de *sustentar um compromisso* que prevenirá a outra recompensa [de curto prazo] de se tornar dominante”⁹⁵. Isso significa, referindo-se à Figura 3.12, que “para resistir à tentação no intervalo de tempo *B*, [é necessário] *comprometer-se* no intervalo de tempo *A* a buscar a maior recompensa durante *C* e durante *B*”⁹⁶.

Ainslie descreve quatro táticas⁹⁷ que levam a um comprometimento com as escolhas de longo prazo: (1) tornar fisicamente impossível escolher a tentação, (2) evitar informações que alterem as preferências, (3) preparar as emoções para que elas não provoquem alteração nas escolhas e (4) tomar uma resolução. As três primeiras têm um alcance limitado e parecem não ter

⁹³ Idem, p. 40.

⁹⁴ Idem, pp. 40-41.

⁹⁵ Idem, p. 43. Grifo nosso.

⁹⁶ Nozick, 1993, p. 21. Referência a uma menção de Ainslie.

Esse argumento é semelhante ao de Sen (1987, p. 57), que considera o *compromisso* uma das capacidades essenciais da condição de agente do homem. Ver texto que faz referência à Nota 87, Capítulo 1.

⁹⁷ Ainslie, pp. 74-78.

Nozick (1993, p. 17) cita os quatro mecanismos que Ainslie propõe, em obra anterior – ver Nota 81 –, para resistir à tentação: (1) tornar impossível a tentação, como fez Ulisses ao se amarrar ao mastro, (2) introduzir um castigo, (3) impedir os benefícios da tentação e (4) adotar um princípio geral de conduta.

a mesma força da quarta proposta, que “transforma escolhas individuais em uma questão de princípios”⁹⁸, ou seja, apela à universalidade para “unir ações particulares sob uma regra comum”⁹⁹.

O “calcanhar de Aquiles” dessa tática é que mesmo uma escolha categórica não afasta completamente a atratividade das recompensas menores: é preciso, ainda, que “a pessoa esteja consistentemente motivada a seguir [a escolha]”¹⁰⁰. O problema, para **Ainslie**, é que enquanto “interesses de longo prazo definem princípios, interesses de curto prazo descobrem exceções”¹⁰¹. Situações de barganha, onde os interesses de cada recompensa interagem para dominar, podem envolver “países, indivíduos ou interesses *dentro de um indivíduo*” empenhados em uma “guerra limitada”, na qual os agentes “compartilham alguns, mas não todos os objetivos”¹⁰². Enfim, é uma situação que tem todos os elementos estruturais de um dilema do prisioneiro repetido¹⁰³ – ou de um “modelo intertemporal de barganha”¹⁰⁴.

Para enfrentar esse impasse, **Nozick** reconhece que um princípio geral de conduta estabelece não uma igualdade, mas uma “*correlação* entre as utilidades das várias ações”¹⁰⁵ de uma mesma classe. Mas não apenas isso: um princípio tem a capacidade de indicar que “se você fizer uma ação particular de uma classe, então você fará todas [da mesma classe]”¹⁰⁶.

⁹⁸ Ainslie, p. 79.

⁹⁹ Idem, p. 80. Para Ainslie, trata-se do conceito de universalidade, de Aristóteles.

¹⁰⁰ Idem, p. 86.

¹⁰¹ Idem, p. 88.

¹⁰² Idem, p. 90. Grifo nosso. Esse processo de barganha “pode ser tudo aquilo que unifica uma pessoa” (idem, p. 43), e é o que Ainslie chama de “vontade” (idem, p. 94).

¹⁰³ O dilema do prisioneiro repetido foi apresentado no item 2.4.

¹⁰⁴ Idem, pp. 92 e 125.

¹⁰⁵ Nozick, 1993, p. 17. Nozick não acompanha a distinção usual na filosofia, que considera os “princípios como mais profundos e menos mecânicos que regras” (ibidem).

¹⁰⁶ Ibidem.

A aposta de **Nozick** é que

amarrar a utilidade deste ato [particular] à desutilidade de todos aqueles atos [semelhantes] no futuro, pode ajudá-lo a atravessar o período *B* de tentação; para você, a utilidade deste [ato] particular está agora alterada. ¹⁰⁷

(...) Essa alteração de utilidades é o resultado do exercício do nosso poder e habilidade de fazer uma ação *valer por* ou *simbolizar* outras [que o princípio exclui]. (...) A adoção do princípio forja a conexão entre as ações, de modo que o castigo por violar o princípio neste momento se torna a desutilidade de violá-lo sempre. ¹⁰⁸

O fato de podermos alterar as utilidades significa que podemos “remodelar nossos desejos” ¹⁰⁹,

mas nós não podemos fazer isto muito freqüentemente (...). Se violarmos um princípio particular que adotamos, não teremos nenhuma razão para pensar que na próxima vez será diferente (...), a não ser que possamos distinguir essa ocasião das posteriores, e também que tenhamos razões para acreditar que essa distinção será importante para nós *depois*, de tal modo que não iremos ser indulgentes mais uma vez e formular outra distinção à qual, mais tarde, não iremos aderir ainda de novo[.] Então, fazer a ação neste momento nos leva a esperar que continuaremos a repeti-la. ¹¹⁰

Os efeitos dessa violação podem se estender a outros contextos, e prejudicar a nossa credibilidade em seguir princípios, por exemplo, no âmbito dos deveres profissionais. Adotar “um (meta)princípio de nunca violar qualquer princípio” ¹¹¹ não seria uma boa solução, pois a ocorrência de uma violação dessa proibição só iria amplificar o problema.

Entretanto, a escolha de quais princípios adotar constitui uma meta-decisão que requer certos cuidados. Por ser

ela mesma um ato que afeta os vínculos de probabilidade entre outras ações, (...) devem-se considerar não só os benefícios resultantes dessa escolha, mas também as probabilidades de violação e os futuros efeitos que essa violação teria. Poderia ser melhor adotar um princípio pior (...), mas mais fácil de seguir. ¹¹²

¹⁰⁷ Idem, pp. 17-18.

¹⁰⁸ Idem, pp. 18-19. Ver também Nozick, 1993, p. 26, com outras palavras.

¹⁰⁹ Idem, p. 138.

¹¹⁰ Idem, p. 19.

¹¹¹ Idem, p. 20.

¹¹² Ibidem.

Há ainda outro ponto em que ações e princípios se relacionam: os custos incorridos (*sunk costs*). A consideração desses *custos*, que correspondem às decisões tomadas no passado e às ações já realizadas, é um reforço aos métodos propostos anteriormente por **Ainslie** para resistir à tentação no período *B* – e que significa, como já mencionado, “*comprometer-se em A a buscar a maior recompensa [que é a de longo prazo] em C e B*”¹¹³.

No cálculo econômico, uma decisão considera apenas as conseqüências presentes e futuras das ações, ignorando os atos e os custos incorridos, que não dependem mais de qualquer decisão. Esta regra parece adequada para a maximização dos benefícios monetários, mas não serve como princípio geral de decisão, pois não consideramos irrelevantes nossos compromissos assumidos com outros.

* * *

Ao criticar o conceito de racionalidade como consistência interna, **Sen** já havia percebido que tal consistência só adquire significado no próprio contexto das ações interdependentes. Tal fato pede um novo modelo que considere: (a) o contexto estratégico, que influencia o agente e é influenciado por este; e (b) a liberdade do agente, capaz de estabelecer objetivos, compromissos e valores, e que mantém aberta a possibilidade de ações não catalogadas nas observações empíricas relacionadas com as preferências individuais¹¹⁴. A percepção de **Sen** certamente influenciou a teoria da ação que **Nozick** viria a propor.

¹¹³ Ibidem. Ver texto relacionado à Nota 96.

¹¹⁴ Ver a crítica da consistência interna nos textos que fazem referência às Notas 55-ss, 75, 87 e 95-ss, Capítulo 1, bem como o resumo apresentado na Figura 2.3 do Capítulo 2.

A utilidade simbólica de uma ação é a propriedade, mencionada anteriormente¹¹⁵, de uma ação *valer por* outras, “em virtude não da sua natureza intrínseca nem do que ela causa, mas daquilo que ela simboliza”¹¹⁶. Para **Nozick**, a função simbólica seria descrita não pelo “entrelaçamento dos dois fios de conexão entre fazer o ato agora e repeti-lo no futuro”, mas pela “simbolização de um fato adicional, não coberto por aqueles dois fios, (...) [e] que por si mesmo afeta a utilidade de ações e resultados alternativos”¹¹⁷.

A simbolização é um fato novo e importante que a teoria da decisão deve considerar. Da mesma forma que a teoria tradicional “acredita que há uma imputação retroativa de utilidade através de uma conexão (...) causal”¹¹⁸, ela deve aceitar também que a utilidade da situação que a ação simboliza seja imputada retroativamente à ação através de conexões simbólicas.

A teoria freudiana descreve certas ações e sintomas aparentemente irracionais que possuem um significado que não é óbvio, mas que simbolizam algo mais. No entanto, segundo **Nozick**,

a existência desse significado simbólico não pode explicar, sozinha, a ocorrência ou persistência de tais ações ou sintomas. Temos que considerar, além disso, que aquilo que essas ações e sintomas simbolizam (...) tem alguma utilidade ou valor (...) para a pessoa; e mais ainda, que essa utilidade (...) daquilo que está sendo simbolizado [possa ser] retroativamente imputada à ação ou ao sintoma, dando-*lhe* assim uma utilidade maior do que aparentava ter.¹¹⁹

Foi mencionado que adotar princípios faz algumas ações valerem por outras, mas o significado simbólico vai além: ele “pode conectar uma ação

¹¹⁵ Ver texto que faz referência à Nota 108.

¹¹⁶ Lacey, p. 96.

¹¹⁷ Nozick, 1993, p. 26.

¹¹⁸ Idem, p. 27.

¹¹⁹ Idem, p. 26. Um exemplo freudiano: “se lavar as mãos simboliza libertação da culpa, então um ato de lavar as mãos assume a utilidade não de estar com as mãos fisicamente limpas, mas de estar livre da culpa” (Lacey, p. 96. Citação de Nozick, 1993, p. 28). A utilidade simbólica pode não ter efeito causal no mundo físico, mas é provável que tenha no mundo psicológico e que esse efeito não esteja confinado a apenas uma mente (Lacey, p. 97).

com outras coisas que não (um grupo de) ações como, por exemplo, ser um certo tipo de pessoa ou com a efetivação de um certo estado de coisas”¹²⁰.

No campo da ética, “seguir princípios éticos, em adição aos propósitos particulares que eles visam, tem também um significado simbólico”¹²¹. Uma pessoa que age moralmente mostra-se, ao menos na circunstância, como se fosse um “membro do reino dos fins, um legislador livre e racional”¹²², que é uma das muitas maneiras pelas quais uma ação ética pode significar simbolicamente. Em geral, a utilidade simbólica dos gestos de grandeza “se incorpora à utilidade simbólica da ação, (...) [e] assim esses significados simbólicos tornam-se parte das razões para se agir eticamente”¹²³.

A proposta de **Nozick** é que, por vivermos

em um rico mundo simbólico, parcialmente cultural e parcialmente fruto de nossa própria criação individual (...), é necessário ampliar a teoria da decisão para incorporar tais conexões simbólicas e detalhar as novas estruturas que elas introduzem.¹²⁴

Como não há uma fórmula que represente adequadamente a utilidade simbólica, ela deve ser tratada separadamente das utilidades evidencial e causal. A falta de uma teoria da utilidade simbólica leva **Nozick** a “reservar para ela um lugar na estrutura dessa teoria mais geral de decisão”¹²⁵.

¹²⁰ Nozick, 1993, p. 33.

¹²¹ Idem, p. 29.

¹²² Ibidem. Referência a Kant, *Fundamentação da Metafísica dos Costumes*.

¹²³ Idem, pp. 29-30.

¹²⁴ Idem, p. 32.

¹²⁵ Idem, p. 34.

3.4 A teoria da ação de Nozick e o dilema do prisioneiro

As teorias de decisão examinadas anteriormente levam em conta apenas as utilidades evidencial e causal das ações¹²⁶. A aplicação dessas teorias ao problema de Newcomb mostrou que não se tem uma confiança absoluta em nenhum dos princípios em que as teorias se baseiam, levando **Nozick** a propor uma nova função a ser maximizada, obtida pela combinação dos dois critérios, ponderados pelos índices de confiança que se tem em cada um deles. Avançando em sua proposta, **Nozick** entende que aquelas duas utilidades, a evidencial e a causal, são insuficientes para representar toda a riqueza da ação humana, e que uma teoria mais ampla deveria incluir a utilidade simbólica, tratada como um componente separado dos demais.

Aquela fórmula do valor-decisão (*VD*), apresentada anteriormente¹²⁷, será então ampliada para incluir “a utilidade simbólica (*US*) de uma ação, que incorpora a utilidade dos vários resultados e ações *simbolizados* pelo ato, e seu respectivo peso *Ws*”:

$$VD(A) = Wc \cdot UEC(A) + We \cdot UEE(A) + Ws \cdot US(A).^{128}$$

A utilidade simbólica, diferentemente das outras, não entra na fórmula como uma *utilidade esperada*, pois não há probabilidades envolvidas. Diz **Nozick**: ela “não é um tipo diferente de utilidade (...), [mas] um tipo diferente de *conexão* – simbólica – para os tipos comuns de utilidade”¹²⁹.

¹²⁶ Tanto a utilidade evidencial quanto a utilidade causal de uma ação seriam determinadas pelas conexões probabilísticas com seus respectivos resultados (idem, p. 48).

¹²⁷ Ver item 3.2.

¹²⁸ Ibidem. A generalidade da fórmula permite que se inclua qualquer outro princípio de decisão cuja utilidade seja plausível.

¹²⁹ Ibidem.

Christensen discorda de que a utilidade simbólica seja uma *conexão* entre a ação e a utilidade comum, pois “o que é simbolizado pode nem mesmo existir; ao contrário, ela é uma *fonte* de valor, o qual se origina na simbolização e não no que é simbolizado” ¹³⁰.

Mas há uma outra visão a esse respeito, a de **Schuessler**. Os eleitores que participam de um processo amplo de votação, tal como os consumidores de um produto de massa, têm um comportamento parcialmente explicado “pelo desejo de estabelecer, para si próprios e para os outros, o que eles são” ¹³¹ ao fazerem a sua escolha. Essa necessidade de expressar uma identidade por meio da participação é um dos motivadores da ação coletiva, e difere daquela motivação instrumental, orientada para o resultado, tomada como premissa pelas teorias dos jogos e da escolha racional.

No caso eleitoral, por exemplo, por ser praticamente nula a probabilidade de um voto individual influenciar o resultado, o benefício instrumental de votar é “superado pelo custo de participar da eleição” ¹³², e mesmo assim as pessoas participam. Esse “paradoxo da participação” é um reflexo da incapacidade de o princípio instrumental explicar todas as instâncias da ação coletiva. Por isso, **Schuessler** dissolve o paradoxo ao adotar também

¹³⁰ Lacey, p. 150. Na opinião de Lacey, o papel que Nozick dá à utilidade simbólica em sua teoria da decisão não foi bem recebido.

Gaus (p. 126), por exemplo, aponta alguns erros na explicação simbólica da ação: (1) as razões simbólicas, em um contexto comunicativo, parecem ser razões orientadas a objetivos. Uma ação baseada unicamente em razões simbólicas parece irracional. (2) A relação de instanciação parece ser o modelo mais correto do que a relação simbólica para descrever a relação entre um princípio e uma ação. (3) As razões simbólicas são derivadas de razões orientadas para objetivos, e ambas são derivadas de razões guiadas por princípios.

¹³¹ Schuessler, p. xii.

¹³² Idem, p. ix.

uma suposição básica de que os indivíduos participam de ações coletivas não necessariamente para produzir resultados, mas agem, ao invés, frequentemente, para assim expressar quem eles são, pela adesão a tais resultados.¹³³

Nesse sentido, **Schuessler** conjuga dois contextos, um instrumental, voltado para os resultados das ações, e um contexto de adesão, em que as ações *simbolizam* ou *expressam* desejos¹³⁴:

No contexto instrumental de produção, o custo de participação é excessivo, relativamente ao valor esperado do resultado (...). No contexto expressivo de adesão, em contraste, o custo é baixo, relativamente ao valor da participação.¹³⁵

No segundo caso, há uma lógica “de identificação (...) no lugar de uma lógica de produção”, que é semelhante à descrição da *condição de agente*, em que **Sen** coloca a identidade social entre as múltiplas motivações para a ação¹³⁶ e apresenta o homem como capaz de “estabelecer objetivos, compromimentos (...) [e] valores”¹³⁷. Para **Schuessler**, tais valores emergem

não dos resultados, mas do próprio processo de participação (...). Em termos econômicos, enquanto os resultados ou estados do mundo são ‘bens públicos’, os retornos expressivos ou simbólicos da participação são ‘bens privados’.¹³⁸

Schuessler parte do âmbito privado para sustentar que a “ambigüidade simbólica”, ou a “capacidade dos símbolos em gerar significados múltiplos”¹³⁹, é um dos determinantes do valor expressivo, pois envolve as interpretações dos “leitores” desses símbolos¹⁴⁰ em um processo de co-produção no qual “os valores simbólicos emergem da prática social”¹⁴¹.

¹³³ Idem, p. 5.

¹³⁴ O contexto simbólico ou expressivo é também chamado de “emocional” e “litúrgico” (Schuessler, p. x).

¹³⁵ Idem, pp. 5-6.

¹³⁶ Sen, 1999, pp. 6 e 14-17. Ver textos que referenciam as Notas 75-76 e 80, Capítulo 1.

¹³⁷ Sen, 1987, p. 57. Ver texto que faz referência à Nota 87, Capítulo 1.

¹³⁸ Schuessler, p. 17.

¹³⁹ Idem, p. 31.

¹⁴⁰ Idem, p. 55.

¹⁴¹ Idem, p. 63.

No caso de uma eleição, por exemplo, a produção do significado de um voto “é parcialmente compartilhada (...) e parcialmente pessoal” ¹⁴². Se, de um lado, os eleitores instilam significado – e, com isso, valores individuais – ao escolherem seus candidatos, estes, de outro lado, ao “preservarem tal ambigüidade, (...) preservam adicionalmente a capacidade de diferentes eleitores os escolherem por diferentes razões” ¹⁴³.

Segundo **Schuessler**, a adesão acontece em dois níveis. No primeiro, cada participante cria não um resultado, mas uma relação com o resultado da ação; no segundo, cada participante estabelece uma relação com o grupo que faz a mesma escolha. Dessa forma, a escolha é “uma forma de expressão, com uma conexão simbólica” ¹⁴⁴ que identifica o agente e fornece um meio de adesão para uma “relação entre o individual e o coletivo” ¹⁴⁵.

* * *

As dificuldades encontradas na solução do problema de Newcomb demonstraram a falta de uma confiança absoluta nos princípios em que se baseiam as teorias da decisão, e deram a **Nozick** a oportunidade de desenvolver uma teoria da ação que inclui a utilidade simbólica ao lado das utilidades evidencial e causal. Seu objetivo é aplicar tal teoria ao dilema do prisioneiro, lembrado a seguir, cuja questão crucial é a possibilidade de haver cooperação racional no âmbito social sem que haja coerção.

¹⁴² Idem, p. 56. Referência a Hans Georg Gadamer (1975 e 1976).

¹⁴³ Idem, p. 59. Essa *ambigüidade simbólica* presente na escolha é usada estrategicamente pelo marketing de um candidato, para ampliar o leque de possibilidade de votos. E também pode ser aproveitada de maneira negativa, para tentar reduzir a ambigüidade do candidato adversário e com isso provocar-lhe a perda de eventuais votos (ibidem).

¹⁴⁴ Idem, p. 52.

¹⁴⁵ Idem, p. 160.

Na situação geral representada pelo dilema do prisioneiro, a ação dominante é a não-cooperativa, em que ambos os prisioneiros pegam 10 anos, e não a ação mais cooperativa, que lhes daria apenas 2 anos de prisão. A racionalidade individual deixa cada parte pior do que se a escolha tivesse sido a ação dominada, mais cooperativa, como se vê na Figura 3.13.

Representação em anos de prisão		Prisioneiro 2	
		Coopera	Não coopera
Prisioneiro 1	Coopera	2, 2	12, 0
	Não coopera	0, 12	10, 10

Representação em utilidades		Prisioneiro 2	
		Coopera	Não coopera
Prisioneiro 1	Coopera	2, 2	0, 3
	Não coopera	3, 0	1, 1

Figura 3.13 – A matriz de resultados de um dilema do prisioneiro (reprodução das Figuras 2.2a e 2.2b do Capítulo 2)

Para **Nozick**, o dilema do prisioneiro e o problema de Newcomb¹⁴⁶ são paralelos, pois envolvem argumentos que levam a ações discrepantes:

- um argumento baseado no princípio da dominação, interpretado de modo compatível com a teoria causal da decisão, [que maximiza a utilidade causal esperada, calculada com probabilidades condicionais de contrafactuais]. (...) A teoria causal da decisão recomenda realizar a ação dominante, [não-cooperativa];
- e um outro argumento baseado na consideração do que cada ato indicaria (...), de modo compatível com teoria evidencial de decisão. O argumento pelo qual no dilema do prisioneiro você deve esperar que a outra pessoa aja como você, embora sua ação não afete causalmente o que ela fizer, é fornecido pelo princípio de maximização da utilidade evidencialmente esperada, onde as probabilidades condicionais [simples] não precisam representar qualquer influência causal. (...) A teoria evidencial da decisão recomenda realizar a ação cooperativa se você pensa que a outra parte é relevantemente similar a você.¹⁴⁷

¹⁴⁶ Ver uma comparação de opiniões a esse respeito no Apêndice 2, “O dilema do Prisioneiro é um problema de Newcomb?”.

¹⁴⁷ Nozick, 1993, pp. 51-52. No caso da teoria evidencial, poderia ser dito ainda de cada agente que “tenha um conhecimento comum de que ambos são agentes racionais e que cada um prefere [a ação cooperativa] à [ação dominante]” (idem, p. 52). Hurley (p. 66) complementa: a cooperação também será considerada uma notícia boa, mesmo que ela não seja mútua.

A Figura 3.14 sumariza para o dilema do prisioneiro o mesmo tipo de “experiência” que **Nozick** aplicou na solução do problema de Newcomb para testar a firmeza das intuições e dos princípios de decisão, conforme as utilidades envolvidas sofressem variações extremas¹⁴⁸. Na primeira alternativa, o prêmio pela ação cooperativa é muito maior do que para a ação dominante, e o prêmio da ação desigual é apenas um pouco maior que o da ação cooperativa, o que nos leva a concluir nesse caso, diferentemente da situação-padrão, que “a cooperação é a escolha racional”¹⁴⁹. Na segunda alternativa, o prêmio da cooperação é próximo ao da escolha dominante, enquanto que para a ação desigual é bem diferente daquele da cooperação, o que nos leva a optar, como na situação-padrão, pela ação dominante.

Primeira situação alternativa		Prisioneiro 2	
		<i>Coopera</i>	<i>Não coopera</i>
Prisioneiro 1	<i>Coopera</i>	1.000, 1.000	0, 1.001
	<i>Não coopera</i>	1.001, 0	1, 1

Segunda situação alternativa		Prisioneiro 2	
		<i>Coopera</i>	<i>Não coopera</i>
Prisioneiro 1	<i>Coopera</i>	3, 3	-200, 500
	<i>Não coopera</i>	500, -200	2, 2

Figura 3.14 – Variações na matriz de resultados de um dilema do prisioneiro¹⁵⁰ (os valores estão representados em utilidades)

¹⁴⁸ Ver texto que faz referência à Nota 73.

¹⁴⁹ Nozick, 1993, p. 53.

¹⁵⁰ Ibidem.

Nota-se que a “Segunda situação” não cumpre a condição adicional proposta por Poundstone para caracterizar o dilema do prisioneiro, qual seja “a recompensa [da cooperação] deve ser maior que a média da tentação [de não-cooperar] e do prejuízo do otário [que coopera, e propicia o ganho do outro]” (Poundstone, 1992, p. 127). Ver Capítulo 2, Nota 121.

Essas simulações comprovam que a decisão muda conforme mudam os valores particulares da matriz de resultados, e isso está compatível com o princípio de maximização do valor-decisão, em que a pessoa dá um certo peso a cada uma das utilidades esperadas que estão envolvidas. Entretanto,

em quais pontos precisamente a decisão de uma pessoa mudará com a variação da utilidade vai depender da confiança que ela tiver em cada um desses princípios (isto é, de quais pesos ela implicitamente atribui a eles), assim como das probabilidades que ela atribui à ação da outra pessoa de ser idêntica à dela própria.¹⁵¹

Hurley dá uma nova explicação de como o princípio evidencial opera, a partir de uma questão deixada por **Nozick** no artigo de 1969¹⁵². Supondo, de modo contrafactual, que um gene possa ser a causa comum de câncer e da vontade de fumar¹⁵³, por que não há nesse caso o mesmo apelo evidencial a favor de não fumar, ou seja, a mesma variação de intuição a favor de pegar apenas a caixa opaca no problema de Newcomb, quando o valor na caixa transparente é drasticamente reduzido? Enquanto **Nozick** sugeria naquele *paper* que a diferença estaria em uma “ilusão de influência”¹⁵⁴, **Hurley** quer ir além para revelar o motivo que faz surgir essa ilusão, ou esse peso, a favor de pegar apenas uma caixa. É que a explicação evidencial falha quando o problema de Newcomb é um caso particular de dilema do prisioneiro¹⁵⁵.

¹⁵¹ Nozick, 1993, p. 53.

Nozick chama a atenção de que mesmo se as ações tiverem probabilidade 1 de serem iguais, e mesmo se o agente der um peso maior para o princípio evidencial que para o princípio causal, ele não irá necessariamente realizar a ação cooperativa. Isso porque o valor-decisão depende também dos valores das utilidades, que podem ser tão discrepantes, como ilustrado pela “Segunda alternativa” da Figura 3.14, a ponto de anular o efeito favorável à cooperação, dado pelos pesos e probabilidades mencionados (*idem*, pp. 53-54).

¹⁵² Nozick, 1969, p. 66.

¹⁵³ Hurley, pp. 66-67. O exemplo de Hurley tem a mesma estrutura do exemplo da barra de chocolate (ver texto que faz referência à Nota 26): em ambos haveria uma causa comum que provoca dois efeitos independentes entre si.

¹⁵⁴ Nozick, 1969, p. 67. Ver texto que faz referência à Nota 61.

¹⁵⁵ Hurley (p. 68) apresenta uma visão inversa àquela de Lewis (pp. 251-2) referenciada no Apêndice 2. Ver Nota 146.

A proposta de **Hurley**, que pretende aprimorar o modelo do valor-decisão, requer a substituição do componente evidencial pelo princípio cooperativo, para justificar os casos em que houver “uma força causal que provoque um resultado mutuamente preferível dentre os atos possíveis”¹⁵⁶, mesmo que tal força não seja suficiente para provocar uma ação cooperativa. No caso do dilema do prisioneiro, os jogadores “têm o poder causal de provocarem *juntos* o segundo melhor resultado para ambos”¹⁵⁷; no problema de Newcomb haverá também uma ilusão de influência *coletiva* se os envolvidos tiverem a mesma ordem de preferências de um dilema do prisioneiro.

* * *

A inclusão da utilidade simbólica nos cálculos poderia ser pensada simplesmente como um aumento fixo na utilidade de cada ação, o que dispensaria a adição de uma parcela de utilidade simbólica separada. Contudo,

o valor simbólico de um ato não é determinado apenas por *aquela* ato. O significado do ato pode depender de quais outros atos estão disponíveis e com quais prêmios, e de quais atos estão também disponíveis para a outra parte ou partes. O que o ato simboliza é algo que ele simboliza quando feito *naquela* situação particular, em preferência *àquelas* alternativas particulares.¹⁵⁸

Ou seja: o valor simbólico de um ato não é apenas uma linha traçada isoladamente entre os estados iniciais e os resultados, mas “pode depender da matriz de decisão como um todo”¹⁵⁹. **Nozick** propõe que “em situações de dilema do prisioneiro deveria entender-se que uma ação tem uma utilidade

¹⁵⁶ Hurley, p. 68.

¹⁵⁷ Ibidem.

¹⁵⁸ Nozick, 1993, p. 55.

¹⁵⁹ Ibidem.

Gaus (p. 130 n. 37) lembra que “para que essa ‘solução’ ao dilema do prisioneiro seja plausível, a utilidade simbólica de uma pessoa (sendo uma pessoa cooperativa) não deve ser refletida na matriz de resultados (caso contrário, naturalmente, não seria mais um dilema do prisioneiro)”.

por si própria”¹⁶⁰, que seria uma utilidade *simbólica*. Não se trata de um tipo comum de utilidade, aplicada a uma ação em vez de a um resultado, mas de um tipo diferente de conexão, pois

em certas situações de dilema do prisioneiro, executar a ação dominada (...) pode ter valor simbólico para uma pessoa. [Tal ação] pode projetá-la como sendo uma pessoa cooperativa em interação com outras, disposta a participar de empreendimentos conjuntos de benefício mútuo. Cooperar em tal situação pode então ser agrupado com outras situações de cooperação que não são caracterizadas como de dilema do prisioneiro. Consequentemente, não cooperar nesta situação particular de dilema do prisioneiro pode ameaçar a cooperação naquelas outras oportunidades.¹⁶¹

A conclusão de **Nozick** é que “nossas respostas ao dilema do prisioneiro são governadas, em parte, pela nossa visão do tipo de pessoa que queremos ser e pelas maneiras que queremos nos relacionar com os outros”¹⁶². Em uma situação particular vamos considerar todas aquelas utilidades, probabilidades e pesos já mencionados, mas vamos também atender às “circunstâncias factuais particulares que dão origem à matriz [de resultados], circunstâncias nas quais uma ação pode vir a ter seus próprios significados simbólicos, que não são devidos simplesmente à estrutura da matriz”¹⁶³.

Ao abrigar todas essas considerações¹⁶⁴, o princípio do valor-decisão abre espaço não só para representar por meio das escolhas o tipo de pessoa que queremos ser, como também para mostrar

não simplesmente uma possível explicação *psicológica* do porquê (algumas) pessoas se desviam da racionalidade, mas que a utilidade simbólica é um legítimo componente do procedimento *racional* de decisão dessa pessoa.¹⁶⁵

¹⁶⁰ Nozick, 1993, p. 56.

¹⁶¹ Ibidem. O argumento de Nozick não restringe o significado da utilidade simbólica à cooperação, pois a pessoa pode, ao contrário, escolher a ação dominante para demonstrar “que ela é racional e não age influenciada por sentimentalismo” (idem, p. 57).

¹⁶² Ibidem.

¹⁶³ Ibidem. Comparar com a crítica de Blackburn, segundo a qual, em uma situação empírica, a teoria dos jogos é incapaz de produzir recomendações. Ver texto que faz referência à Nota 147, Capítulo 1, e a discussão no item 2.6.

¹⁶⁴ O tema do dilema do prisioneiro repetido foi apresentado nos itens 2.4 e 2.5.

¹⁶⁵ Nozick, 1993, p. 57.

O objetivo de **Nozick** é aplicar aos princípios éticos os resultados obtidos na discussão do dilema do prisioneiro. Uma ação que vale por todas as que são restringidas por um mesmo princípio, como os códigos deontológicos, adquire uma desutilidade simbólica que, mesmo não sendo um impedimento absoluto, é uma barreira que tem o peso de todo o grupo. Dessa forma, “ao se agrupar as ações em um princípio que as proíbe – ‘não matarás’ –, [substitui-se] o cálculo utilitarista (ou egoísta) em separado dos custos e benefícios de *uma* ação” ¹⁶⁶, por um cálculo que inclua o peso daquela desutilidade simbólica do grupo inteiro, que a ação passa a carregar consigo.

Por outro lado, um princípio que considera que “uma ação ética pode simbolizar (e expressar) um ser racional que dá a si mesmo leis, um legislador do reino dos fins”, tem uma utilidade simbólica que, “incorporada à utilidade simbólica e então ao *valor-decisão* da ação, torna-se parte das razões para o agir ético” ¹⁶⁷. Enfim, por analogia, uma pessoa que adota o princípio de maximizar o valor-decisão, pode ser levada a executar ações éticas.

Ao propor inicialmente o conceito de valor-decisão, **Nozick** atribuiu a ele um caráter normativo, pois combinava a força de vários princípios no direcionamento da decisão de uma pessoa. Agora, ao ligar o valor-decisão à ética, **Nozick** completa a idéia de transformar a teoria clássica da decisão, que poderia ser pensada originalmente como descritiva de um comportamento instrumental, em uma estrutura inteiramente normativa.

¹⁶⁶ Idem, p. 62.

¹⁶⁷ Ibidem.

3.5 De volta à racionalidade

Vimos anteriormente¹⁶⁸ que os princípios tinham uma função causal que era instrumental, pois “a noção de conexão (probabilística) causal captura a idéia central da racionalidade instrumental: a conexão meios-fins”¹⁶⁹. No decorrer do argumento, àqueles aspectos causais foram incorporados os conceitos evidenciais e simbólicos, para juntos formarem um novo princípio, o de valor-decisão. Entretanto, ao ligar o *VD* à ética, **Nozick** mostrou que uma teoria da decisão, inicialmente descritiva de um comportamento compatível com a função causal, não seria mais puramente instrumental, pois agora agregaria uma estrutura normativa, trazida pelos objetivos simbólicos¹⁷⁰.

A visão de uma teoria da decisão inicialmente descritiva, que no final se revela normativa, talvez seja ingênua, pois a instrumentalidade já traria em si, desde o começo, elementos de normatividade. Diz **Hampton**:

a concepção instrumental da razão está recheada sutilmente por uma variedade de compromissos normativos, incluindo alguns compromissos com a natureza do [que é o] bem para uma pessoa instrumentalmente racional, tal que pensar sobre a melhor maneira de alcançar nossos fins pressupõe certos compromissos normativos sobre o que esses fins deveriam ser. Isso significa, paradoxalmente, que a razão instrumental não pode ser interpretada como verdadeiramente instrumental! Pode-se dizer que os economistas e outros cientistas sociais que se utilizam da teoria da utilidade esperada “fazem ética o tempo todo”.¹⁷¹

¹⁶⁸ Referência ao início do item 3.3.

¹⁶⁹ Nozick, 1993, p. 133. Também orientada aos objetivos, a teoria evidencial “não reivindica que a melhor ação seja um instrumento para alcançar nossos fins” (Gaus, p. 112).

¹⁷⁰ Nozick chama a atenção de que “os significados simbólicos podem não ser todos bons, assim como os desejos e as preferências podem também não ser. O problema é que uma teoria da racionalidade não pode *excluir* os significados simbólicos. Tal inclusão, entretanto, não garante um conteúdo bom ou desejável. Para isso, precisaríamos desenvolver uma teoria de quais significados simbólicos e de quais preferências e desejos seriam admissíveis, usando-a para constranger aqueles significados particulares e desejos que poderiam ser introduzidos em uma teoria mais formal de racionalidade” (Nozick, 1993, p. 30, nota).

¹⁷¹ Hampton, p. 13.

Putnam (pp. 30-31) acrescenta que, do ponto de vista de filósofos pragmatistas como Pierce, James e Dewey, os juízos normativos são essenciais para a prática da ciência, e não se referem apenas ao que chamamos de moral ou ético. Eles incluem julgamentos de coerência, plausibilidade, razoabilidade, simplicidade e até de “o que deve ser”.

Nozick concorda com esse comentário, que alarga a noção que se tem do papel instrumental da racionalidade. Para ele, “nem sempre a racionalidade instrumental se destina à realização de alguma coisa *a mais*, algo completamente distinto”¹⁷², pois realizar a própria ação pode estar entre os objetivos de uma ação, que assim pode ter um valor por si mesma.

A discussão desenvolvida até aqui nos mostra que a função instrumental é uma parte comum em qualquer teoria da decisão. Segundo **Nozick**, “a racionalidade instrumental é a teoria *default* (...), a questão é saber se ela abrange *toda* a racionalidade”¹⁷³.

Há um entrelaçamento entre princípios e processos de decisão: de um lado, “usamos a razão para escolher quais princípios de decisão seguir, (...) e [de outro] decidimos quais princípios racionais usar”¹⁷⁴. Se, por hipótese,

a *natureza* da racionalidade pudesse ser pensada de modo inteiramente instrumental, o seu *valor* não o poderia, porque tal racionalidade seria avaliada em parte por si mesma, (...) e assim teria um valor *intrínseco*. (...) Além disso, se suficientemente utilizado, [tal instrumento] pode se tornar uma extensão de nós mesmos e ser assimilado dentro de nós como uma parte importante de nossa identidade.¹⁷⁵

Entretanto, **Nozick** não aceita que a racionalidade instrumental, por mais larga que seja, defina toda a nossa racionalidade:

se os seres humanos forem simplesmente seres *humeanos*, isso pareceria diminuir nossa estatura, pois o homem é o único animal que não se contenta em ser simplesmente um animal. (...) É simbolicamente importante para nós que nem todas as nossas ações visem à satisfação de nossos desejos. Os princípios, como vimos, nos dão meios para controlar e remodelar nossos desejos.¹⁷⁶

¹⁷² Nozick, 1993, p. 133.

¹⁷³ *Ibidem*.

¹⁷⁴ *Idem*, p. 135.

¹⁷⁵ *Idem*, p. 136.

Nozick cita Heidegger em seu apoio: “ferramentas instrumentais, usadas com frequência suficiente, podem se tornar extensões de nós mesmos; nossas fronteiras se estendem por meio delas, até os seus *fins*, à medida que interagimos com o mundo” (*ibidem*).

¹⁷⁶ *Idem*, p. 138. Nozick faz referência ao seu próprio texto, citado na Nota 109.

Os princípios, entretanto, não teriam a força que Kant depositou neles, a de produzirem ações simplesmente pelo respeito que lhes seria devido (*ibidem*).

É possível que, ao produzir significados simbólicos em adição ao que as ações causam, estejamos indo além do “nexo humeano” meios-fins ou, pelo menos, *simbolizando* tal intenção. Essa “metafunção” dos princípios – a de atuar sobre outras funções que os próprios princípios têm – mostra que “seguir princípios pode também ter uma utilidade simbólica”¹⁷⁷.

Hampton considera que, “ao se aceitar que os desejos podem estar sujeitos a padrões normativos, viola-se a visão humeana de que a razão é uma mera escrava das paixões”¹⁷⁸. Nesse sentido, **Nozick** notou que a introdução de axiomas formais de consistência na teoria da utilidade esperada, apesar de ser “um minúsculo passo além de **Hume**”¹⁷⁹, abre a possibilidade para se considerar outros padrões normativos, além dos mencionados – por exemplo, padrões morais – aos quais nossos desejos teriam de se sujeitar.

Apesar dos esforços, **Nozick** não consegue propor uma “teoria da racionalidade substantiva”¹⁸⁰ capaz de contrariar as opiniões de **Hume** e **Blackburn** a respeito do papel limitado da razão na escolha de objetivos e desejos racionais¹⁸¹. **Nozick** anteriormente havia proposto que “os princípios são dispositivos com funções específicas, projetados para trabalhar alinhados com (e modificar) os objetivos que são dados”¹⁸², e assim

nosso procedimento racional nos capacita a modificar esses [objetivos] em uma extensão significativa, mas gradativamente, por meio de mudanças individualmente pequenas, que podem, quando repetidas, produzir uma mudança cumulativa grande.¹⁸³

¹⁷⁷ Idem, p. 139.

¹⁷⁸ Hampton, 1998, p. 244.

¹⁷⁹ Nozick, 1993, p. 140.

¹⁸⁰ Ibidem.

¹⁸¹ Ver texto que faz referência às Notas 148-152, Capítulo 1.

¹⁸² Nozick, 1993, p. 163. Referência à Nozick, 1993, p. 21.

¹⁸³ Nozick, 1993, p. 163.

Para **Nozick**, a busca frustrada de uma teoria dos objetivos e desejos mostrou que a racionalidade instrumental não exaure todo o domínio da racionalidade. E o mais importante é que a introdução do “valor-decisão transcende aquela estrutura largamente instrumental” ¹⁸⁴.

Ao revelar alguns componentes e princípios que estão em jogo na decisão em geral, que de outra forma talvez não fossem notados, **Nozick** aponta o caminho para desenvolver uma teoria da ação racional que estaria além de uma simples teoria da decisão, pois esta trata apenas “da melhor ação” ¹⁸⁵.

Nozick considera que

é natural pensar a racionalidade como um processo direcionado a objetivos, [que] se aplica tanto à racionalidade da ação quanto à racionalidade da crença. (...) Nesta concepção instrumental, a racionalidade consiste na busca efetiva e eficiente de objetivos, fins e desejos. A respeito dos objetivos propriamente ditos, uma concepção instrumental tem pouco a dizer. ¹⁸⁶

A visão biológica de **Nozick** – cuja base é a evolução das espécies pela seleção natural de mutações – pressupõe, no caso da racionalidade, que “a crença verdadeira é uma estratégia de sobrevivência” ¹⁸⁷, e que

o nosso interesse original pela verdade teve uma base instrumental. As verdades nos serviam melhor que as falsidades, e melhor que a ausência de crenças na hora de lidarmos com os perigos e as oportunidades do mundo. Não era necessária uma verdade perfeitamente precisa, mas apenas uma crença que fosse *suficientemente verdadeira* para dar (mais) resultados desejáveis quando agíssemos de acordo com ela. O necessário e desejável era que fosse “uma verdade servível” e, para isso, não era preciso que a crença fosse perfeitamente verdadeira. ¹⁸⁸

¹⁸⁴ Ibidem.

¹⁸⁵ Idem, p. 65.

¹⁸⁶ Idem, p. 64. Apesar de a racionalidade da crença não ter sido tematizada neste trabalho, alguns comentários sobre o assunto foram acrescentados para compor o quadro dos conceitos envolvidos.

¹⁸⁷ De Sousa, p. 11, interpretando o que diz Nozick, 1993, p. 68.

¹⁸⁸ Nozick, 1993, p. 68. Nozick complementa: “A base instrumental de nosso interesse pela verdade não é que as pessoas desejassem acreditar em verdades porque reconheciam que isso poderia ser instrumentalmente útil, mas, por causa da utilidade de se acreditar em verdades aproximadas, se produziu uma seleção evolucionária por algum tipo de interesse na verdade da crença, como, por exemplo, certa curiosidade por descobrir verdades” (ibidem).

De Sousa coloca a linguagem humana a serviço da visão biológica de **Nozick**. Ele duvida que os nossos desejos e instintos tenham autoridade para perseguir a verdade, mas defende a possibilidade de, por meio da linguagem, controlarmos nossos impulsos naturais, pois “temos uma *escolha* de como confrontar os objetivos da verdade com outros objetivos”¹⁸⁹. Essa capacidade da linguagem permite ainda que “o número de possíveis pensamentos explícitos (...) seja *superastronômico*”¹⁹⁰ e que por isso “a cultura humana seja *virtualmente infinita*”¹⁹¹, o que nos leva a pensar que a racionalidade, apesar de ser moldada pela sociedade (“as pessoas não nascem racionais”¹⁹²), tem uma base evolutiva e não exclusivamente cultural.

A idéia de **Nozick** é que racionalidade de uma ação ou de uma crença é uma função tanto (1) da sensibilidade às razões que temos, como (2) da confiabilidade do processo que gerou essas razões¹⁹³. Ações e crenças têm, por sua vez, certos objetivos, tais como a verdade e a satisfação de desejos que possivelmente são mais bem atendidos se consideradas essas razões.

Segundo **Nozick**, encontramos na literatura duas maneiras de explicar a relação entre as razões e as hipóteses que constituem nossas crenças:

em uma visão a priori, (...) as razões são coisas que a nossa mente tem a capacidade de reconhecer. (...) Uma segunda visão, a factual, sustenta que uma razão é uma evidência para uma hipótese quando ela está em uma certa relação factual contingente para tal hipótese.¹⁹⁴

¹⁸⁹ De Sousa, p. 12.

¹⁹⁰ Idem, p. 16. De Sousa explica que um “número superastronômico” é superior ao número de partículas existente no universo, estimado em 10^{80} (ibidem).

¹⁹¹ Idem, p. 8.

¹⁹² Nozick, 1993, p. 124.

¹⁹³ Idem, p. 107.

Nozick teria uma estratégia de combinar explicações internalistas (a relação entre os elementos do sistema de crenças) e confiabilistas (Gaus, p. 107).

¹⁹⁴ Nozick, 1993, pp. 107-8.

O conflito entre essas duas visões abre caminho para mais um passo na direção de uma teoria da ação racional. O fato de aceitarmos as leis da lógica e de as reconhecermos como normas é uma propriedade que pode ser pensada em termos evolucionários. **Nozick** se utiliza dessa perspectiva para sugerir a combinação da visão factual com a visão *a priori*. Uma conexão factual poderia dar uma base diferente para nossa ação ou crença, tal que

agir segundo *razões* envolve o *reconhecimento* de uma conexão na relação estrutural entre conteúdos. Tal reconhecimento poderia ter sido ele mesmo útil e por isso selecionado. A propriedade de uma certa conexão factual *parecer* auto-evidentemente evidencial para nós poderia ter sido selecionada e favorecida, pois a ação a partir dessa conexão factual, se acontecer de fato, em geral fortalece a adaptação.¹⁹⁵

Lacey resume: “o que a evolução seleciona são criaturas para quem certas relações factuais parecem auto-evidentes, ou mais que factuais”¹⁹⁶.

Em outras palavras, **Nozick** enfatiza mais a *conexão* do que o *factual*:

há uma conexão factual, e houve uma seleção a favor de organismos para quem esse tipo de conexão pareceu válido, por apreender esse tipo de conexão e, por tal apreensão, conduzir a certas crenças adicionais, inferências e assim por diante. Há uma seleção por reconhecer como válidos certos tipos de conexão que *são* factuais, isto é, por chegarem a nós parecendo-nos *mais* do que apenas factuais.¹⁹⁷

Para **Nozick**, isso explica por que certos modelos aproximados, como o pensamento indutivo e a geometria euclidiana, parecem auto-evidentes e satisfatórios para nós.

* * *

A descrição de **Nozick** para a relação entre a racionalidade e a evolução é análoga à apresentada por **Kant** para a relação entre a experiência e os conceitos. Diz **Kant**:

¹⁹⁵ Nozick, 1993, p. 108.

¹⁹⁶ Lacey, p. 154.

¹⁹⁷ Nozick, 1993, pp. 108-09.

Há apenas dois caminhos pelos quais podemos pensar uma relação *necessária* entre a experiência e os conceitos de seus objetos: ou a experiência torna esses conceitos possíveis ou esses conceitos tornam a experiência possível. A primeira suposição não se sustenta com respeito às categorias (nem com respeito à intuição sensível pura), pois desde que elas são conceitos *a priori*, e então independentes da experiência, atribuir-lhes uma origem empírica seria um tipo de *generatio aequivoca*. Resta, portanto, só a segunda suposição – um sistema (como se fosse) de *epigenesis* da razão pura – isto é, que as categorias contêm, por parte do entendimento, as condições de possibilidade de toda a experiência em geral.¹⁹⁸

Descritos esses dois caminhos, um proveniente da razão pura e outro de origem empírica, **Kant** antecipa uma via intermediária, em que

as categorias não são nem (...) os primeiros princípios a priori de nosso conhecimento *pensados espontaneamente*, nem derivam da experiência, mas são disposições subjetivas do pensamento, implantadas em nós desde o primeiro momento da nossa existência, e assim arranjadas pelo nosso Criador de modo que o seu uso esteja em harmonia completa com as leis de natureza, nas quais a experiência acontece – um tipo de *sistema pré-formado* da razão pura.¹⁹⁹

Mas **Kant** não aceita essa terceira via, porque “a *necessidade* das categorias, que é parte essencial desse conceito, teria então que ser sacrificada”²⁰⁰. A opção de **Nozick**, dispondo de uma perspectiva evolucionária²⁰¹ que **Kant** não teve, é exatamente essa terceira via²⁰²: a “concordância entre as duas visões não é uma harmonia afortunadamente pré-estabelecida. O que faz as duas visões combinarem é a evolução”²⁰³. Entretanto, a explicação evolucionária reverte o sentido da *revolução copernicana* de **Kant**, pois

é a *razão* que é a variável dependente, moldada pelos fatos, e sua dependência dos fatos explica a correlação e a correspondência entre eles. A razão nos fala da realidade porque a realidade molda a razão, ao selecionar aquilo que parece “evidente”.²⁰⁴

¹⁹⁸ Kant, 1787, B166-67, p. 174. (§27 “Dedução dos Princípios do Entendimento”).

¹⁹⁹ Kant, 1787, B167-68, pp. 174-75.

²⁰⁰ Idem, p. 175.

²⁰¹ A teoria da evolução surgiu em 1859, cinqüenta e cinco anos após a morte de Kant.

²⁰² Lembramos que, para Monod, a seleção natural atua *sobre* os produtos do acaso – as mutações –, e só deles se alimenta; mas a seleção atua em um domínio de condições muito rigorosas, do qual o acaso foi excluído (Monod, p. 137).

²⁰³ Nozick, 2001, p. 123.

²⁰⁴ Nozick, 1993, p. 111. Essa via intermediária foi também defendida por Peirce (1839-1914) em sua trajetória de “um kantismo puro que, simplesmente, foi forçado gradativamente para o pragmatismo” (Tiercelin, p. 54).

* * *

A pesquisa de **Nozick** sobre a racionalidade trilhou dois caminhos, um da decisão e outro da crença. No primeiro, a questão da decisão racional, conforme discutido como tema deste Capítulo, produziu um princípio – o que determina maximizar o valor-decisão – que nos levou além do domínio da decisão meramente instrumental, ao incluir, ao lado da utilidade causal, também a utilidade evidencial e a simbólica. O outro caminho de pesquisa, apenas mencionado, mostrou que a racionalidade das crenças depende “das razões que as tornam críveis, e da sua geração por um processo confiável em produzir crenças verdadeiras”²⁰⁵.

A perspectiva evolucionária de **Nozick**, ao apresentar uma nova explicação da natureza da racionalidade, mostra que

a racionalidade é uma adaptação evolucionária com propósito e função delimitados. Ela foi selecionada por e projetada para trabalhar alinhada com fatos duradouros que se mantiveram durante o período da evolução humana. (...) Muitos dos tradicionais e intratáveis problemas filosóficos, resistentes a uma resolução racional, podem resultar de uma tentativa de estender a racionalidade além dessa função delimitada. Dentre esses se incluem os problemas da indução, de outras mentes [etc.].²⁰⁶

Apesar de a racionalidade instrumental ter sido vista como “uma ferramenta poderosa, (...) parte significativa de qualquer concepção de racionalidade, (...) ela não enquadra totalmente a racionalidade substantiva dos objetivos e desejos”²⁰⁷. Foi preciso desenvolver “uma noção de racionalidade que fosse além até mesmo da noção instrumental mais larga para incluir o

²⁰⁵ Nozick, 1993, p. 176. Como mencionado na Nota 186, a questão da crença racional não foi tematizada neste trabalho.

²⁰⁶ Ibidem.

²⁰⁷ Ibidem.

simbólico e o evidencial”²⁰⁸. **Nozick** defende que todas essas parcelas que compõem o valor-decisão têm uma origem evolucionária, pois atuam para

fortalecer outros desejos e mantê-los durante períodos de privação, em reforço aos seus próprios objetos individuais ou para permitir que as pessoas coordenem suas ações em situações de dilema do prisioneiro em que, sem eles, a cooperação não ocorreria.²⁰⁹

Isso não significa que a racionalidade vai seguir o mesmo caminho traçado até aqui pela evolução. Por exemplo, os objetivos que serviram à “adaptação inclusiva (*inclusive fitness*) podem agora ser perseguidos mesmo quando conflitarem com tais adaptações”²¹⁰. Para vislumbrar tal recurso, **Nozick** introduz o conceito de *imaginação*, que descreve o que seria a capacidade da razão em criar objetivos. Os princípios que usamos em nossas ações têm bases originalmente evolucionárias, mas, ao serem confrontados e ponderados entre si, podem nos levar além daquelas origens. Dessa forma,

podemos usar nossa imaginação para formular novas possibilidades, sejam elas objetivos, teorias ou planos aventureiros, todas desenraizadas de funções evolutivas específicas do passado. Até mesmo se a própria imaginação, que é a habilidade para inventar novas possibilidades, tiver uma função evolutiva, nós agora podemos usar essa capacidade para quaisquer propósitos que escolhermos.²¹¹

Não precisamos pensar que o molde da nossa racionalidade seja um “individualismo cartesiano”, pois vivemos no meio de indivíduos dotados de uma racionalidade que evoluiu de modo similar à nossa, e, como eles,

²⁰⁸ Idem, p. 177.

²⁰⁹ Ibidem.

²¹⁰ Ibidem. O conceito de “*inclusive fitness*”, atribuído a Hamilton (1964), é aplicado por Axelrod à análise da cooperação no dilema do prisioneiro. Tal conceito define a propriedade de “um indivíduo ter um interesse parcial no ganho do adversário (isto é, ele reconhece os resultados em termos do que é chamado de *adaptação inclusiva*)” (Axelrod, p. 97). No processo evolutivo, esse “reconhecimento” é reforçado se os competidores forem parentes, o que daria ao perdedor a possibilidade de ver os genes comuns sobreviverem.

Podem-se ver semelhanças entre esse conceito e o de *simpatia*, mas este, conforme definido por Hume e Sen, não estaria dirigido apenas a parentes. Ver item 3.2.

²¹¹ Nozick, 1993, p. 177.

estamos predispostos a aprender a linguagem dos outros e também a aprender fatos que os mais velhos nos contam. Estamos predispostos a aceitar o que dizem e a aceitar as correções que fazem do que dizemos, pelo menos até termos trabalhado bastante a linguagem e a informação, até estarmos aptos a fundamentar dúvidas e levantar questões. (...) Uma vez adquirida alguma base na linguagem e em crenças factuais, poderemos também questionar e modificar as crenças que os outros têm. ²¹²

Nesse campo intersubjetivo em que “a linguagem é a manifestação e o veículo da racionalidade, (...) há espaço para a formulação, discussão e desenvolvimento de princípios adicionais da crença racional” ²¹³. É então possível perceber que essa racionalidade que “reformou o mundo, (...) [também levou] a si própria a estender o seu domínio mais além” ²¹⁴, por meio de um controle sobre nossas ações, emoções e assim sobre o mundo todo. Isso significa que ela nos permitiu “transcender nosso estado de meros animais, de fato e simbolicamente” ²¹⁵.

Em suma, ao moldar e controlar suas próprias funções, “a natureza da racionalidade inclui em si a própria Natureza” ²¹⁶. Da nossa parte, individual ou coletivamente, dependendo dos princípios e objetivos que escolhermos, e das ações que nos definem simbolicamente, “nossas vidas podem vir a significar mais do que instrumentalmente elas valem. E, ao significar mais, nossas vidas passam a valer mais” ²¹⁷.

²¹² Idem, p. 178.

²¹³ Idem, p. 179.

²¹⁴ Idem, p. 180.

²¹⁵ Idem, p. 181.

²¹⁶ Ibidem.

²¹⁷ Ibidem.

CAPÍTULO 4 - CONCLUSÕES: NORMATIVIDADE E EQUILÍBRIO

*“Cesse tudo o que a Musa antiga canta,
que outro valor mais alto se alevanta.”
Camões, Os Lusíadas.*

Vimos ao longo da pesquisa que para enfrentar o “desafio humeano”¹ alguns filósofos contemporâneos, como **Sen** e **Nozick**, insistem em achar uma brecha para que outros fatores, como o comportamento moral ou simbólico, sejam reconhecidos como motivadores da ação. Mesmo **Blackburn**, alinhado às teses humeanas, admite² que, no domínio da ética, que é o mesmo da escolha, as paixões individuais governam tudo, mas há um terreno comum onde os homens se entendem racionalmente, surgindo padrões de comportamento que são aprovados ou não, de maneira intersubjetiva.

A estratégia moderada adotada por **Sen** e **Nozick** é a de explorar esse campo de pesquisa – que também coincide com aquele espaço no qual não se sabe determinar “quanto (...) é atribuível ao puro efeito da liberdade e quanto à mera natureza”³ – onde se examinam as interações razão-emoção e as alterações que podem surgir no contexto da decisão. Mesmo assim, é difícil escapar ao desafio humeano *radicalizado*: a própria escolha do que trazer à discussão já não seria também dependente das paixões?⁴

¹ Expressão cunhada por B. Williams (Cullity e Gaut, p. 12) para a atitude de negar a existência de razões normativas categóricas, deixando aos defensores destas o ônus da prova. Ver Introdução.

² Blackburn, 1998, p. 201, e Blackburn, 2001, pp. 111-2 e 129-33.

³ Blackburn, 1995, p. 707. Citação de Kant, 1787, A551/B579.

⁴ Este é um *paradoxo socrático*, no qual o objeto investigado inclui o investigador.

Na opinião de **Sen**, “o comportamento real dos seres humanos é afetado por considerações éticas”⁵, como o dever, a lealdade e a boa vontade, que podem induzir a busca por outros objetivos além do próprio bem-estar, e assim provocar desvios daquele comportamento-padrão visto como racional. Nesse sentido, a função própria do homem não seria a busca da felicidade, “entendida utilitaristicamente como satisfação dos desejos”⁶, mas a possibilidade de o homem desenvolver toda a sua capacidade para se realizar na sociedade em que vive. Nesse caso, a base da ação estaria não nos desejos, mas nas capacidades humanas devidamente desenvolvidas.

Uma visão semelhante é apresentada por **Nozick**, ao notar que a introdução de axiomas de consistência na teoria da utilidade esperada, apesar de ser “um passo minúsculo além de **Hume**”⁷, abre a possibilidade de haver outros padrões normativos aos quais nossos desejos teriam de se sujeitar.

Na opinião de **Nozick**, os princípios em geral justificam, orientam e controlam nossas ações. Podem também nos obrigar a determinada conduta, garantir aos outros os nossos compromissos e definir nossa identidade. Mas os princípios não têm a força categórica que **Kant** lhes atribui, pois eles apenas “nos dão os meios para controlar e remodelar nossos desejos”⁸.

⁵ Sen, 1987, p. 67-8.

⁶ Berti, pp. 274-75. Berti aponta a coincidência das posições de Sen e Aristóteles, com relação ao conceito de “capacidade”.

⁷ Nozick, 1993, pp. 138-40.

O próprio Hume aceita essas preocupações, ao reconhecer em nossas preferências a influência, embora limitada e rara, de um “altruísmo positivo”, formado pelo nosso sentimento de amor e pela nossa capacidade de perceber o sentimento de outros (“simpatia”), que é diferente de um “altruísmo negativo”, redutível ao egoísmo. A ação humana seria resultante de uma combinação de egoísmo e altruísmo positivo, mas este não é tão grande a ponto de que não haja necessidade, em certos casos, de algum tipo de coerção (Taylor, p. 162).

Taylor (pp. 151-4 e 159-60) indica as seguintes passagens em Hume, *Treatise*, 3.2.1.12, 3.2.2.5, 3.2.7.1 e 3.2.7.8.

⁸ Nozick, 1993, p. 138.

Todas essas considerações estão incluídas no *valor-decisão*, que resume a teoria de ação de **Nozick**. O valor-decisão tenta conciliar diferentes princípios de decisão, cujas importâncias relativas variam com o contexto. Assim, para um mesmo agente, a escolha pode mudar conforme o contexto.

O modelo representa a soma de uma parte instrumental com uma parte não-instrumental. Na primeira parte, ele combina dois princípios de maximização da utilidade esperada, o causal e o evidencial⁹. Na segunda, incorpora o princípio da maximização da utilidade simbólica, que é aquilo que o agente quer significar com a sua ação.

Como se vê, uma decisão representada pelo valor-decisão não é a consequência da aplicação de um único princípio, mas de um confronto entre vários princípios. Internamente, como indicado por **Ainslie**, há uma constante disputa entre interesses¹⁰. Externamente, em geral, as pessoas envolvidas em uma decisão têm interesses parcialmente comuns e parcialmente conflitantes, como em um dilema do prisioneiro, e assim participam de uma interação que é mediada pela ética. O significado simbólico de uma ação, definido pelo agente em interação com os outros envolvidos no ato, “carrega” a intenção de uma nova ordem ética, que pode ser sancionada ou não.

O valor-decisão incorpora a posição ética do agente, e por isso a decisão que resulta é *normativa* – ela não define simplesmente que ação o agente escolheria, mas que ação ele *deveria* escolher. Além disso, ela não reflete um cálculo puramente instrumental, pois outros fatores foram considerados.

⁹ A descrição dos modelos de decisão é apresentada no item 3.1.

¹⁰ Ver o texto a partir da referência à Nota 81, item 3.3. Em especial, “interesses de longo prazo definem princípios, interesses de curto prazo descobrem exceções” (Ainslie, p. 88).

Nossa conclusão é que o uso do valor-decisão retrata as seguintes relações entre um agente e os valores éticos, vistas de um ponto que inclui um observador que vai descrever uma determinada ação:

- Se um agente escolhe usar o valor-decisão, então sua decisão leva em conta a questão ética. Ele poderia abrir mão da ética, mas, nesse caso, não precisaria ter usado o valor-decisão, pois sua escolha seria puramente instrumental¹¹.
 - O significado simbólico é julgado de acordo com os valores éticos vigentes na comunidade. O resultado pode ser um simples voto vencido, uma condenação, uma nova ética, ou uma combinação dessas possibilidades¹².
- Se um observador escolhe o valor-decisão para descrever uma determinada decisão, ele faz uso normativo do modelo. Ele pode abrir mão da avaliação ética, mas, nesse caso, não precisaria ter escolhido o valor-decisão.
 - O observador julga o significado simbólico da ação segundo os valores éticos da sua própria comunidade, que pode ser a mesma do agente ou compartilhar valores com ela.
 - Se houvesse um isolamento total, como se fosse um ser de outro planeta, dificilmente tal observador entenderia os fatos, pelo menos de imediato. Essa observação alude à frase: “se um leão pudesse falar, não poderíamos compreendê-lo”¹³, que se refere aos instintos próprios do leão, seu modo de vida e suas relações com os outros seres vivos¹⁴.

A favor do uso do valor-decisão, o argumento acima aponta a dificuldade de se encontrar um observador inteiramente neutro, isto é, que não esteja envolvido com o objeto observado. Uma eventual neutralidade poderia ser apenas fruto da irrelevância do objeto ou do mero desconhecimento. Em

¹¹ Nessa afirmação está subentendido que a escolha de uma ação instrumental é um caso particular de decisão. Ao optar por esse tipo de decisão, o agente demonstra que tem um objetivo de eficácia a alcançar e quer – ou não se importa – que os outros saibam disso.

¹² Esse raciocínio se aplica tanto a uma simples discussão de condomínio, quanto ao caso de um bandido que atira gratuitamente em uma pessoa rendida, e até à situação de *Antígona*, de Sófocles, condenada por transgredir a lei que proibia enterrar seu irmão morto.

¹³ Wittgenstein, II, xi, p. 216.

¹⁴ Em Wittgenstein encontra-se ainda o conceito de *proposições fulcrais* ou *proposições-dobradiças* (“*hinge propositions*”), como, por exemplo, “somos seres humanos”. As proposições fulcrais constituem a armação do nosso pensamento e um pano de fundo de certezas relativamente permanentes, ou seja, o ambiente em que nossa imagem do mundo é construída. Essas proposições não se sustentam com base em evidências, mas são certezas pressupostas em qualquer juízo, a partir das quais justificamos outras crenças. O importante no nosso caso é que tal reconhecimento não é atributo de “alguns indivíduos específicos, mas de todas as pessoas”, o que reforçaria a necessidade de inserção do observador na comunidade (Glock, pp. 73-78. Referências a Wittgenstein, *On Certainty*, §§94-5, 100-3, 136, 138, 308, 341-3, 401-3, 462, 494, 614, 655).

toda descrição de ação humana está implícita a possibilidade de um julgamento, que se revela em termos tais como “irracional”, “boa”, “cruel”, e até no uso de expressões especializadas de alguma ciência, como é o caso de “obsessiva”, “carente”, etc.¹⁵. O “emaranhamento” que **Putnam** constatou para justificar o “colapso da dicotomia fato/valor”¹⁶ é o mesmo que se observa entre a descrição de um fato e o enunciado, ou a presunção, de uma norma.

O valor-decisão responde às principais questões, levantadas na Introdução deste trabalho, que pediam um modelo mais completo que aquele da teoria da decisão então disponível, pois o resultado da maximização agora depende:

- dos pesos que atribuímos a cada princípio de decisão no contexto considerado;
- dos valores (utilidades) que cada ação alternativa significa para nós;
- das expectativas (probabilidades causais e evidenciais) de sucesso de cada ação;
- dos valores simbólicos que queremos atribuir às ações alternativas, que por sua vez dependem de uma visão conjunta de todas elas no contexto considerado.

Uma etapa importante no desenvolvimento do valor-decisão foi o uso que **Nozick** fez do problema de Newcomb. Esse problema desafiava as teorias ao opor justificativas igualmente plausíveis que levavam a soluções diferentes. A aplicação do valor-decisão fez com que a escolha por uma das op-

¹⁵ Morton (p. 92) acrescenta um tipo de explicação que é feita “por meio de pontos de atração” (“*by attractors*”), na qual um observador usa comportamentos-padrão para relacionar a ação de um agente com os resultados que são esperados. Um exemplo é a descrição de que “alguém quebrou a perna porque é desastrado quando em férias” (ibidem). Mas, como “frequentemente não se pode descrever em termos neutros, [um observador] deve ter alguma idéia do que seja a justiça” (idem, p. 93).

¹⁶ Referência ao título de Putnam, *The Collapse of the Fact/Value Dichotomy*. Ele comenta: “É muito mais fácil dizer ‘isto é um julgamento de valor’, que significa ‘isto é apenas uma questão de preferência subjetiva’, do que fazer o que Sócrates tentou nos ensinar, que seria examinar quem somos e quais são as nossas mais profundas convicções, e submetê-las ao teste minucioso de um exame reflexivo. (...) A pior coisa que se pode dizer a respeito da dicotomia fato/valor é que, na prática, ela funciona como um cala-boca ou, pior ainda, como um paralisa-pensamento” (Putnam, p. 44).

ções passasse a depender não só da estrutura de resultados, mas também do grau de confiança que se tinha em cada um dos princípios envolvidos. Esse novo modo de “decidir sobre como decidir”, ou metadecisão, foi então aplicado ao dilema do prisioneiro, com a sobreposição de um valor simbólico de origem ética ao que seria o equilíbrio das utilidades indicado pela teoria dos jogos, o que eventualmente poderia alterar a solução instrumental.

* * *

Tudo o que foi discutido até aqui neste trabalho pressupõe uma certa estabilidade de condições. Entretanto, o ambiente humano, o ambiente ético, é dinâmico. Da mesma forma que a evolução biológica acontece por meio de uma deriva natural das espécies, parece haver também uma *deriva ética*, movida em parte pelo acaso, mas que recebe os impulsos principais ditados pela nossa liberdade e pelas nossas razões conflitantes. Um dos elementos que contribuem para o dinamismo da ética é a possibilidade de transgressão¹⁷, isto é, a ação pode não ocorrer *sempre* da maneira prescrita.

Com o termo “deriva” não estamos nos referindo a um mero desvio de comportamento, mas a uma mudança da força relativa das razões, que impõe uma nova norma conforme os contextos mudem¹⁸. Por exemplo, se os avanços da tecnologia vierem a permitir que o homem viva mil anos, esse fato trará profundas conseqüências no nosso modo de vida¹⁹.

¹⁷ Ver texto que faz referência à Nota 12, e também Nota 135, Capítulo 2, sobre a passagem do contrato social para um novo estado de equilíbrio.

¹⁸ Nozick distingue na ética quatro camadas: a *ética do respeito*, a *ética da responsabilidade*, a *ética do cuidado*, e a *ética da Luz*. Enquanto a primeira camada, a *do respeito*, é obrigatória e comanda a cooperação para benefício mútuo, e por isso é a parte fundamental da ética, as outras dependem de uma opção do agente (Nozick, 2001, p. 280).

¹⁹ Para Hans Jonas (pp. 1-24), “a natureza da ação humana mudou”, e por isso é preciso uma “ética do futuro”. Estamos construindo um mundo em que os aparatos artificiais “en-

Enquanto atribuímos ao acaso as mutações genéticas e outros acontecimentos naturais fortuitos, colhemos frutos da nossa liberdade, tais como a manipulação genética e a inteligência artificial, e mesmo algum novo projeto de sistema de organização social... Essa deriva não é inteiramente ao acaso, porque, em geral, os homens têm alguma coisa em comum que sustenta e dá permanência à ética e que, por causa disso, “pelo menos às vezes é possível coordenar as ações para *benefício mútuo*”²⁰. Mas não sabemos, *a priori*, que escala de prioridades adotaremos.

No limite do impasse ético, a saída política, frequentemente adotada nos regimes democráticos, tem ela mesma os seus problemas, pois não está imune aos conflitos de poder, à falibilidade dos atores e às suas próprias paixões. Como vimos, nos dilemas morais a solução sempre deixa resto²¹.

A metadecisão defendida por **Nozick** desloca o problema da descrição para o nível dos princípios. Mas, ao nos aprofundarmos nos princípios, desembocamos novamente na questão razão *vs.* paixões, naqueles termos em que foram colocados por **Platão** e **Hume**²². Seria esse um mero jogo de espelhos, com reflexos infinitos? Na teoria dos jogos, o problema da reflexividade

golem” a natureza – combustíveis, máquinas, bombas, mercados, telecomunicações, computadores –, e criam poluição e ameaças de extinção das espécies. Ao mesmo tempo, aumenta a distância entre as ações do homem e a sua capacidade de prever conseqüências. Essa ampliação do campo de ação joga a *responsabilidade* para o centro da cena ética, o que faz com que o homem inclua a sua posteridade entre seus deveres e afaste a ética do “confinamento antropocêntrico” em que sempre esteve.

Um dos possíveis desdobramentos teóricos que a tese de Jonas traz é o de provocar uma revisão no modo como nós, os agentes, intuimos a *taxa de desconto* – importante fator do cálculo da utilidade esperada em processos repetitivos –, pois os eventos futuros passarão a ter maior relevância no presente. Como explicado no texto que faz referência à Nota 105, Capítulo 2, se a taxa de desconto for suficientemente baixa, a cooperação pode ocorrer no equilíbrio de um jogo repetido, mesmo que não ocorra no jogo estático.

²⁰ Nozick, 2001, p. 244.

²¹ Barcan Marcus (pp. 127-41), conforme citado na Nota 99 do Capítulo 1 e ao longo do item 2.1.

²² Ver Introdução.

infinita das expectativas dos participantes é resolvido pelo teorema de **Nash**, que estabelece como solução um ponto de equilíbrio.

No nosso caso, **Sen** e **Nozick**, o primeiro por meio da busca da realização social e o segundo pelo uso do valor-decisão, propõem dar mais um giro no que parece ser um equilíbrio circular²³. As paixões podem nos motivar a agir, mas nem sempre são suficientes para prevalecer sobre todas as razões; da mesma forma, as razões, sozinhas, não são tão fortes que garantam que uma ação aconteça. Algumas vezes é preciso um *acordo* para resolver os nossos dilemas do prisioneiro internos²⁴, o que envolve a possibilidade de cooperação ao longo do tempo na escolha que temos de fazer entre “prêmios” tais como satisfações, ameaças de punições e sentimentos de culpa.

Essa oscilação sem-fim configura a multiplicidade e diversidade dos nossos fins, sem que saibamos com antecedência quem vai dar a última palavra, se a razão ou os desejos – ou se há uma instância superior de definição. O fato de a solução nessa disputa nunca ser encontrada, além de nos embalar numa vertigem, nos dá evidências de que talvez a nossa natureza seja mesmo essa... a de seres destinados a um *regresso infinito* para construir histórias dos mundos possíveis. Enfim, procurar pela origem dos fins é como renovar a crença – infundada, mas eficaz para o movimento de equilíbrio – de que haja um pote de ouro no fim do arco-íris.

²³ Nozick lembra que a literatura sobre o equilíbrio reflexivo presume que os princípios tenham uma autoridade manifesta sobre os casos empíricos (Nozick, 1993, p. 194 n. 20). Por ser um equilíbrio, supõe-se que nesse contexto haja também uma força no sentido inverso, que permite a revisão das teorias e princípios a partir dos novos dados empíricos. O equilíbrio reflexivo foi comentado no texto que faz referência à Nota 50, Capítulo 1.

²⁴ Um modelo interno de barganha foi sugerido por Ainslie (pp. 92 e 125), conforme mencionado no texto que faz referência à Nota 104, Capítulo 3.

Apêndice 1 – A REGRA DE BAYES E AS TEORIAS DA DECISÃO

1. Definição

O conceito de *probabilidade condicional*, ou $\Pr(p/q)$, conforme representado por **Ramsey**, trata do “grau de crença em p dado q ”, que “expressa aproximadamente as chances pelas quais o sujeito apostaria agora em p , sendo a aposta válida apenas caso q seja verdadeiro”¹. **Sahlin** exemplifica: “a crença parcial em p , a situação econômica vai melhorar, dado q , o câmbio será desvalorizado em pelo menos 20%”².

A *regra de Bayes*, formulada a partir das idéias de Thomas **Bayes** (1702-1761), estabelece a seguinte relação entre as probabilidades condicionais $\Pr(p/q)$ e $\Pr(q/p)$, e as probabilidades simples, $\Pr(p)$ e $\Pr(q)$:

$$\Pr(p/q) = \frac{\Pr(p) \Pr(q/p)}{\Pr(p) \Pr(q/p) + \Pr(\sim p) \Pr(q/\sim p)} .$$

Essa é apenas uma regra quando se trabalha com a probabilidade como freqüência, mas é a idéia central da probabilidade como representação dos graus de crença de uma pessoa, a tal ponto que “a abordagem da probabilidade como crença é chamada comumente de ‘*bayesiana*’”³.

2. A regra de Bayes como aprendizagem contínua

A regra de **Bayes** assume um sentido dinâmico quando se quer calcular a probabilidade subjetiva de uma hipótese H_j , dada uma nova evidência E , pois ela pode ser revisada ou atualizada a qualquer momento, a partir da probabilidade $\Pr(H_j)$ anterior à nova evidência e da “verossimilhança (*likelihood*) das hipóteses H_1, H_2, \dots, H_n , à luz da evidência E ”⁴. No caso de haver apenas duas hipóteses, a regra de **Bayes** toma a forma:

$$\Pr(H_j/E) = \frac{\Pr(H_j) \Pr(E/H_j)}{\Pr(H_1) \Pr(E/H_1) + \Pr(H_2) \Pr(E/H_2)} .$$

Como o denominador é constante, pode-se dizer que a probabilidade de uma hipótese posterior a uma nova evidência é proporcional à probabilidade anterior à evidência, multiplicada pela verossimilhança da hipótese à luz da nova evidência⁵. Essa seria *a maneira racional* de ajustar nossas crenças à medida que novas experiências aconteçam – desde que “tenhamos um compromisso com o tipo de pessoa que escolhemos ser”, o que parece ser um desfecho surpreendentemente moral para um roteiro lógico⁶.

¹ Ramsey, p. 76.
² Sahlin, p. 36.
³ Hacking, p. 171.
⁴ Idem, p. 174.
⁵ Idem, p. 175.
⁶ Idem, p. 259-60.

3. O conceito de verossimilhança

A verossimilhança (*likelihood*) está associada aos conceitos de probabilidade e causalidade. **Skyrms** explica essa relação:

Quando se diz ‘fumar causa câncer’, não é que fumar seja uma condição *suficiente* para contrair câncer, pois é possível fumar e não contrair câncer, nem que fumar seja uma condição *necessária*, pois é possível contrair câncer sem nunca ter fumado. O significado é probabilístico: fumar aumenta a chance de contrair câncer. (...) Dizemos que fumar tem uma tendência positiva à suficiência, ou que fumar é *estatisticamente relevante* para o câncer, se $p(C/F)$ for maior que $p(C/\sim F)$. Dizemos que fumar tem uma tendência positiva à necessidade, ou que o câncer é estatisticamente relevante para fumar, se $p(F \text{ no passado}/C)$ for maior que $p(F \text{ no passado}/\sim C)$.⁷

[Ex.:] A probabilidade que Ezequiel seja um macaco, dado que é um gorila, é 1. Mas a probabilidade de que ele seja um gorila, dado que é um macaco, é menor que 1.⁸

A verossimilhança é essa *tendência à necessidade*. Uma propriedade importante é que ela não obedece à regra aditiva das probabilidades.

4. A regra de Bayes como teoria da decisão

Na teoria dos jogos, os agentes em geral trabalham com informações incompletas, que envolvem risco e incerteza. Eles constroem *funções de utilidade* para acomodar suas preferências referentes às opções disponíveis, e a essas utilidades estão associadas *probabilidades*, que podem ser *objetivas* ou *subjetivas*. A reunião dessas informações se dá no modelo de *maximização da utilidade esperada*, adotado pela teoria da escolha racional.

5. Uma nova proposta de distinção entre as probabilidades

A distinção entre probabilidades *subjetivas* e *objetivas* é criticada por **Hacking** por dar margem a dúvidas. Ele recomenda que a distinção seja feita entre as probabilidades do *tipo frequência* e as do *tipo crença*⁹.

As probabilidades do *tipo frequência* relatam uma observação factual ou uma tendência, como a proporção de um resultado no lançamento de dados ou o número relativo de peças defeituosas produzidas em um processo. No limite, elas são “idealizadas como uma seqüência infinita de observações”¹⁰.

As probabilidades do *tipo crença* expressam “o grau com que alguém acredita (...) [em uma proposição], levando em conta uma evidência”¹¹. Tais crenças parciais, em um extremo subjetivo, são “probabilidades pessoais”¹² e, em outro extremo, “probabilidades lógicas”, também subjetivas, mas condicionadas a evidências.

Conforme o tipo de probabilidade que se adote, o modelo de maximização da utilidade esperada pode gerar diferentes decisões. Enquanto a probabilidade do tipo frequência é a base da teoria *causal* da decisão, a probabilidade do tipo crença é a base da teoria *evidencial* da decisão.

⁷ Skyrms, 2000, p. 134.

⁸ Idem, p. 130.

⁹ Hacking, 2001, p. 131.

¹⁰ Idem, p. 144.

¹¹ Idem, p. 132.

¹² Hacking prefere “probabilidade pessoal” a “grau de crença parcial” (idem, p. 142).

Apêndice 2 – O DILEMA DO PRISIONEIRO É UM PROBLEMA DE NEWCOMB?

Lewis propõe que um dilema do prisioneiro pode ser visto como um caso particular do problema de Newcomb, “ou melhor, [de] dois problemas de Newcomb lado a lado, um por prisioneiro”¹, como representado abaixo:

	<i>Você pega uma caixa</i>	<i>Você pega as duas caixas</i>
<i>Eu pego uma caixa</i>	Eu ganho \$1.000.000 Você ganha \$1.000.000	Eu ganho \$0 Você ganha \$1.001.000
<i>Eu pego as duas caixas</i>	Eu ganho \$1.001.000 Você ganha \$0	Eu ganho \$1.000 Você ganha \$1.000

Cada um dos envolvidos descreve o seu dilema da seguinte maneira:

- (1) Ofereceram-me mil – é pegar ou largar.
- (2) Talvez me dêem também um milhão; mas isso acontecerá ou não de modo causalmente independente do que eu fizer agora. Nada do que eu fizer agora terá qualquer efeito sobre se eu vou ganhar um milhão ou não.
- (3) Eu ganharei o meu milhão se e somente se você não pegar o seu mil.²

Enquanto os pontos (1) e (2) são descrições claras de um problema de Newcomb, o ponto (3) aparentemente teria que ser substituído por:

- (3') Eu ganharei o meu milhão se e somente se for predito que eu não pegarei o meu mil.³

A estratégia de **Lewis** é demonstrar que (3') equivale a (3). Destaca inicialmente a independência causal presente no ponto (2), já que “‘predito’ não precisa significar ‘predito com antecedência’”⁴. É suficiente que haja um processo preditivo qualquer, como, por exemplo, um computador que sortearia um sinal para liberar ou não o dinheiro. Mas, em vez de um computador, poderia ser usada uma “réplica minha”, com a vantagem de agir “como eu agiria”. Como o processo de predição não precisa ser extremamente confiável, pois mesmo assim o interesse no problema se mantém, a réplica pode ser substituída por uma pessoa qualquer, por exemplo, “você”, o que faz (3') finalmente voltar à condição (3), demonstrando a tese de que o dilema do prisioneiro pode ser um caso particular do problema de Newcomb.

O modo como **Lewis** elimina o que não é essencial e agrega, ao mesmo tempo, condições que precisam ser cumpridas, abre caminho para alguns questionamentos. Um desses critica o requisito de que cada um dos envolvidos pense “quase do mesmo jeito que o outro”⁵, pois essa “quase certeza” vai depender de quais crenças estão por trás dessa convicção. A conclusão é que *nem sempre* um dilema do prisioneiro pode ser considerado um caso particular do problema de Newcomb.

¹ Lewis, pp. 251-2. As posições na matriz foram adaptadas às convenções deste trabalho.

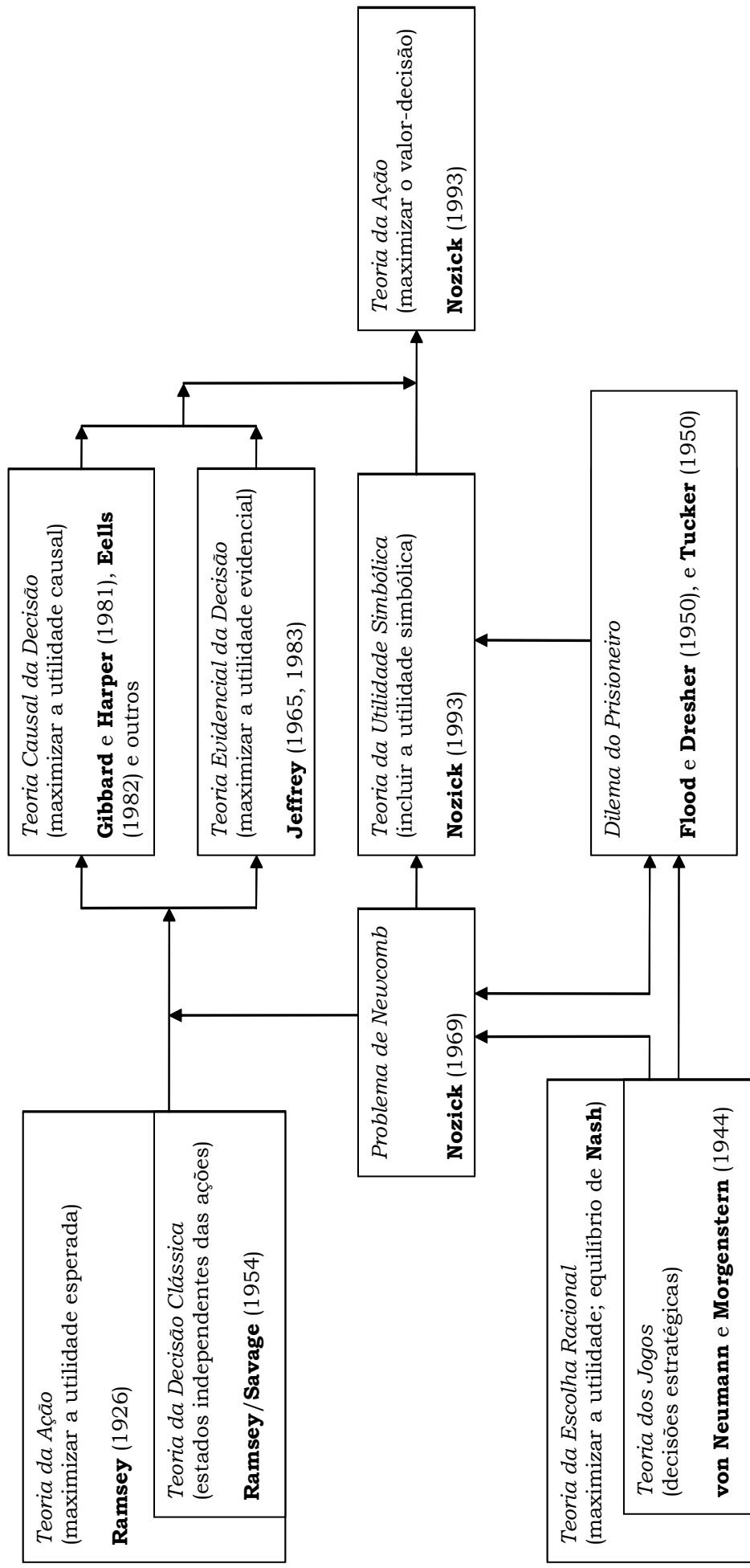
² Idem, p. 252.

³ Ibidem.

⁴ Ibidem. Da mesma forma que os modelos que “‘predizem’ algo que já aconteceu”.

⁵ Sobel, 1985, p. 264.

Apêndice 3 – A GENEALOGIA DA TEORIA DA AÇÃO DE NOZICK



Apêndice 4 – EXEMPLOS DE CURVAS DE DESCONTO QUE SE CRUZAM

Curvas exponenciais

Forma geral: $u_t = u_n \cdot [1 - r]^{(n-t)}$, ¹

onde u_n é o valor da recompensa, n é o instante da recompensa, $0 \leq t \leq n$ é o instante de referência, e $r \geq 0$ é a taxa de desconto.

Parâmetros	Exp 1	Exp 2	Exp 3
Valor da recompensa (u_n)	100,0	60,0	35,0
Instante da recompensa (n)	10	7	4
Taxa de desconto (r)	0,20	0,40	0,40

Tempo	Valores descontados (u_t)		
	Exp 1	Exp 2	Exp 3
0	10,7	1,7	4,5
1	13,4	2,8	7,6
2	16,8	4,7	12,6
3	21,0	7,8	21,0
4	26,2	13,0	35,0
5	32,8	21,6	
6	41,0	36,0	
7	51,2	60,0	
8	64,0		
9	80,0		
10	100,0		

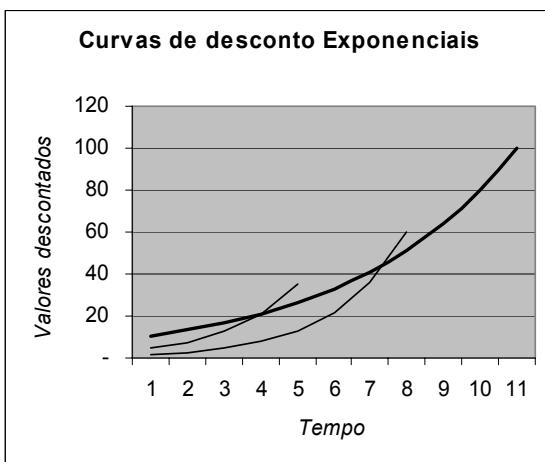
Curvas hiperbólicas

Forma geral: $u_t = u_n / [k_1 + k_2 \cdot (n - t)]$, ²

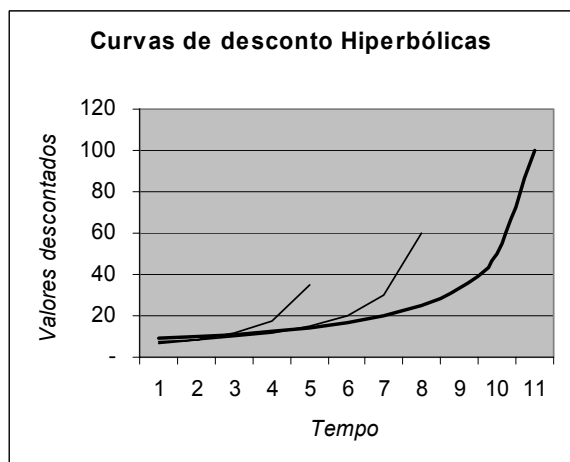
onde u_n é o valor da recompensa, n é o instante da recompensa, $0 \leq t \leq n$ é o instante de referência, k_1 e k_2 são constantes próximas de 1.

Parâmetros	Hip 1	Hip 2	Hip 3
Valor da recompensa (u_n)	100,0	60,0	35,0
Instante da recompensa (n)	10	7	4
Constantes k_1 e k_2	1	1	1

Tempo	Valores descontados (u_t)		
	Hip 1	Hip 2	Hip 3
0	9,1	7,5	7,0
1	10,0	8,6	8,8
2	11,1	10,0	11,7
3	12,5	12,0	17,5
4	14,3	15,0	35,0
5	16,7	20,0	
6	20,0	30,0	
7	25,0	60,0	
8	33,3		
9	50,0		
10	100,0		



Obs.: para essas curvas, as taxas de desconto foram arranjadas de forma a haver cruzamento.



Obs.: o cruzamento das curvas hiperbólicas não depende de taxas de desconto.

¹ Ainslie, p. 29 (notação adaptada).

² Idem, p. 35 (notação adaptada).

REFERÊNCIAS

1. **AINSLIE**, George. *Breakdown of Will*. Cambridge, UK: Cambridge University Press, 2001.
2. **ARISTOTLE**. *The Nicomachean Ethics*. The Loeb Classical Library, ed. by G. P. Goold, vol. XIX. Tradução: H. Rackham. Cambridge-MA e London: Harvard University Press, ©1926/1999.
3. **ARISTOTLE**. *The Complete Works of Aristotle*. The Revised Oxford Translation, ed. by Jonathan Barnes, 2 vols. Princeton-NJ: Princeton University Press, ©1984, 1995.
4. **AXELROD**, Robert M. *The Evolution of Cooperation*. Basic Books, 1984.
5. **BARCAN MARCUS**, Ruth. "Moral Dilemmas and Consistency". In **BARCAN MARCUS**, Ruth. *Modalities: Philosophical Essays*. New York: Oxford University Press, ©1993, 1995, (pp. 125-141).
6. **BARON**, Jonathan. *Thinking and Deciding*. 3rd ed. Cambridge, UK: Cambridge University Press, 2000.
7. **BÄUERLE**, Rainer e **CRESSWELL**, M. J. "Propositional Attitudes". In **GABBAY**, D. e **GUENTHNER**, F. *Handbook of Philosophical Logic*. Vol. IV. Dordrecht, Boston e London: Kluwer Academic Publishers, 1994. (Cap. IV-7, pp. 491-512).
8. **BÊRNI**, Duílio de Avila. *Teoria dos jogos: jogos de estratégia, estratégia decisória, teoria da decisão*. Rio de Janeiro: Reichmann & Affonso Editores, 2004.
9. **BERTI**, Enrico. *Aristóteles no século XX. (Aristotele nel novecento, 1992)*. Tradução: Dion Davi Macedo. São Paulo: Edições Loyola, 1997.
10. **BINMORE**, K. G. *Playing Fair: Game Theory and the Social Contract*. Volume I. Cambridge-MA e London: Harvard University Press, ©1994/2000.
11. **BLACKBURN**, Simon. *Dicionário Oxford de Filosofia. (The Oxford Dictionary of Philosophy, 1994)*. Consultoria da edição brasileira: Danilo Marcondes. Tradução: Desidério Murcho... *et al.* Rio de Janeiro: Jorge Zahar Editor, 1997.
12. **BLACKBURN**, Simon. "Practical Tortoise Raising". In *Mind*, 104, pp. 695-711, 1995.
13. **BLACKBURN**, Simon. *Ruling Passions: A Theory of Practical Reasoning*. New York: Oxford University Press, ©1998, 2000.
14. **BLACKBURN**, Simon. *Being Good: A Short Introduction to Ethics*. New York: Oxford University Press, 2001.
15. **BORRADORI**, Giovanna. *The American Philosopher: Conversations. (Conversazioni americane, 1991)*. Chicago: The University of Chicago Press, 1994.
16. **BRANDÃO**, Jacyntho Lins. "O jogo e o labirinto no Eutidemo". *Revista Filosófica Brasileira*. Nº. 3, 1988, pp. 23-48.
17. **BROOME**, John. "An Economic Newcomb Problem". In *Analysis*, 49, 4, pp. 220-2, 1989.

18. **CAMPBELL**, Richmond. "Introduction: Background for the Uninitiated". In **CAMPBELL**, Richmond e **SOWDEN**, Lanning (eds.). *Paradoxes of Rationality and Cooperation: Prisoner's Dilemma and Newcomb's Problem*. Vancouver: The University of British Columbia Press, 1985, (pp. 3-41).
19. **CHRISTENSEN**, David. "Nozick's *The Nature of Rationality*". *Noûs*, 29(2), pp. 259-274, June 1995.
20. **CLARK**, Michael. *Paradoxes from A to Z*. London e New York: Routledge, ©2002, 2003.
21. **COHEN**, L. Jonathan. *An Essay on Belief and Acceptance*. New York: Oxford University Press, ©1992, 1995.
22. **CRAIG**, Edward (ed.). *Routledge Encyclopedia of Philosophy*, Version 1.0. London: Routledge, 1998.
23. **CULLITY**, Garrett; **GAUT**, Berys. "Introduction". In **CULLITY**, Garrett; **GAUT**, Berys (eds.), *Ethics and Practical Reason*. New York: Oxford University Press, 1998, (pp. 1-27).
24. **DANCY**, Jonathan e **SOSA**, Ernest (eds.). *A Companion to Epistemology*. Oxford-UK, Malden-MA: Blackwell Publishers, ©1992, 2000.
25. **DAVIS**, Morton D. *Game Theory: A Nontechnical Introduction*. Mineola-NY: Dover Publications, ©1983, 1997.
26. **DE FINETTI**, Bruno. "Foresight: Its Logical Laws, Its Subjective Sources" (1937). Trad. Henry E. Kyburg, Jr. In **KYBURG**, Jr., Henry E. e **SMOKLER**, Howard E. (eds.) *Studies in Subjective Probabilities*. 2nd ed. Huntington-NY: Robert E. Krieger Publishing Company, 1980.
27. **DE SOUSA**, Ronald. "Rational Animals: What the Bravest Lion Won't Risk – Reflections on Nozick's *The Nature of Rationality*". Disponível em: <<http://www.chass.utoronto.ca/%7Eesousa/rational.animal.pdf>>. Acesso em: 13/fevereiro/2005.
28. **DIÓGENES LAÉRTIOS**. *Vidas e Doutrinas dos Filósofos Ilustres* (c200 d.C.). Tradução: Mário da Gama Kury. 2.ed. Brasília: Editora Universidade de Brasília, 1987.
29. **DIXIT**, Avinash H. e **NALEBUFF**, Barry J. *Thinking Strategically: The Competitive Edge in Business, Politics, and Everyday Life*. New York e London: W. W. Norton & Company, ©1991, 1993.
30. **ELGIN**, Catherine Z. "Epistemology's Ends, Pedagogy's Prospects". In *Facta Philosophica*, Volume 1, Number 1, pp. 39-54, 1999.
31. **ELSTER**, Jon. "Introduction". In **ELSTER**, Jon (ed.). *Rational Choice*. New York: New York University Press, 1986, (pp. 1-33).
32. **ÉSQUILO**, "Agamêmnon", in *Oréstia*. Tradução: Mário da Gama Kury. 3.ed. Rio de Janeiro: Jorge Zahar Editor, ©1990, 1998, (pp. 17-86).
33. **FITELSON**, Branden. "Review of *The Foundation of Causal Decision Theory*, by James Joyce". *Mind*, 112(447), 2003 (pp. 545-551). Disponível em: <<http://fitelson.org/joyce.pdf>>. Acesso em: 13/fevereiro/2005.

34. **GÄRDENFORS**, Peter e **SAHLIN**, Nils-Eric. "Introduction: Bayesian Decision Theory – foundations and problems". In **GÄRDENFORS**, Peter e **SAHLIN**, Nils-Eric (eds.). *Decision, Probability, and Utility: Selected Readings*. Cambridge-NY: Cambridge University Press, ©1988, 2000, (pp. 1-15).
35. **GARDNER**, Martin. "Newcomb's Paradox". In **GARDNER**, Martin. *The Night is Large: Collected Essays, 1936-1995*. New York: St. Martin's Griffin, 1996, (pp. 412-20).
36. **GAUS**, Gerald F. *Goals, Symbols, Principles*. In **SCHMIDTZ**, David (ed.) *Robert Nozick*. Cambridge-UK: Cambridge University Press, 2001, (pp. 105-30).
37. **GENSLER**, Harry J. *Formal Ethics*. London e New York: Routledge, 1996.
38. **GIBBARD**, Allan e **HARPER**, William L. "Counterfactuals and Two Kinds of Expected Utility" (1978). In **CAMPBELL**, Richmond e **SOWDEN**, Lanning (eds.). *Paradoxes of Rationality and Cooperation: Prisoner's Dilemma and Newcomb Problem*. Vancouver: The University of British Columbia Press, 1985, (pp. 133-58).
39. **GIBBONS**, Robert. *Game Theory for Applied Economists*. Princeton-NJ: Princeton University Press, 1992.
40. **GIBBONS**, Robert. "An Introduction to Applicable Game Theory". In *The Journal of Economic Perspectives*, Volume 11, Issue 1, pp. 127-149, 1997.
41. **GLOCK**, Hans-Johann. *Dicionário Wittgenstein*. Tradução: Helena Martins. Rio de Janeiro: Jorge Zahar Ed., 1998.
42. **HACKING**, Ian. "What's best. Review of Robert Nozick, *The Nature of Rationality*". In: *London Review of Books*, 27 January 1994, pp. 17-18.
43. **HACKING**, Ian. *An Introduction to Probability and Inductive Logic*. Cambridge, UK: Cambridge University Press, 2001.
44. **HAMPTON**, Jean E. *The Authority of Reason*. Cambridge-UK: Cambridge University Press, 1998.
45. **HARDIN**, Garrett. "The Tragedy of the Commons". In: *Science*, 162 (1968), pp. 1243-8. Disponível em: <<http://dicoff.org/page95.htm#b1>>. Acesso: 13/fevereiro/2005.
46. **HAUSMAN**, Daniel M. e **MCPHERSON**, Michael S. *Economic Analysis and Moral Philosophy*. Cambridge, UK: Cambridge University Press, ©1996, 2002.
47. **HONDERICH**, Ted (ed.). *The Oxford Companion to Philosophy*. New York: Oxford University Press, 1995.
48. **HORNSBY**, Jennifer. *Simple Mindedness: In Defense of Naïve Naturalism in the Philosophy of Mind*. Cambridge-MA, London: Harvard University Press, 1997. (Part II: "Agency").
49. **HUME**, David. *A Treatise of Human Nature*. (1739-40). Editado por: David Fate Norton e Mary J. Norton. New York: Oxford University Press, 2001.
50. **HURLEY**, S. R. "A New Take from Nozick on Newcomb's Problem and Prisoner's Dilemma". *Analysis*, 54(2), pp. 65-72, April 1994.
51. **JEFFREY**, Richard C. *The Logic of Decision*. 2nd ed. Chicago e London: University of Chicago Press, ©1983, 1990.

52. **JOHNSON**, James. *Roteiro de aulas do curso ICPSR*. Michigan University, 2002. Material cedido gentilmente pelo prof. Bruno Wanderley Reis.
53. **JONAS**, Hans. *The Imperative of Responsibility (Das Prinzip Verantwortung, 1979; Macht oder Ohnmacht der Subjektivität?, 1981)*. Tradução: Hans Jonas e David Herr. Chicago: The University of Chicago Press, 1984.
54. **KANT**, Immanuel. *Crítica da razão prática. (Kritik der praktischen Vernunft, 1788)*. Tradução: Artur Morão. Lisboa: Edições 70, 1997.
55. **KANT**, Immanuel. *Critique of Pure Reason. (Kritik der reinen Vernunft, 1787)*. Trad.: Norman Kemp Smith. Boston/New York: Bedford/St. Martin's, 1969.
56. **KREPS**, David M. *Game Theory and Economic Modeling*. New York: Oxford University Press, ©1990, 1997.
57. **LABURTHE-TOLRA**, Philippe. **WARNIER**, Jean-Pierre. *Etnologia-Antropologia. (Ethnologie-Anthropologie, 1993)*. Tradução: Anna Hartmann Cavalcanti. Petrópolis-RJ: Vozes, 1997. (Capítulo 12).
58. **LACEY**, A. R. *Robert Nozick*. Princeton-NJ: Princeton University Press, 2001.
59. **LEWIS**, David. "Prisoner's Dilemma Is a Newcomb Problem" (1979). In **CAMPBELL**, Richmond e **SOWDEN**, Lanning (eds.). *Paradoxes of Rationality and Cooperation: Prisoner's Dilemma and Newcomb Problem*. Vancouver: The University of British Columbia Press, 1985, (pp. 251-5).
60. **LUCE**, R. Duncan e **RAIFFA**, Howard. *Games and Decisions: Introduction and Critical Survey*. New York: Dover Publications, ©1957, 1989.
61. **MELLOR**, D. H. "Introduction". In **MELLOR**, D. H. (ed.). *Philosophical Papers: Frank Plumpton Ramsey*. New York: Cambridge University Press, 1990, (pp. xi-xxvii).
62. **MAUSS**, Marcel. *Ensaio sobre a dádiva. (Essai sur le don, 1923-24/1950)*. Tradução: Antônio Filipe Marques. Lisboa: Edições 70, 2001.
63. **MONOD**, Jacques. *O acaso e a necessidade. (Le Hasard et la Nécessité, 1970)*. Tradução: Bruno Palma e Pedro Paulo de Sena Madureira. 3.ed. Petrópolis-RJ: Editora Vozes, 1972.
64. **MORTON**, Adam. *The Importance of Being Understood: Folk Psychology as Ethics*. London e New York: Routledge, 2003.
65. **MYERSON**, Roger B. *Game Theory: Analysis of Conflict*. Cambridge-MA: Harvard University Press, ©1991, 2001.
66. **MOORE**, F. C. T. "Taking the Sting Out of the Prisoner's Dilemma". *Philosophical Quarterly*, 44(175), pp. 223-233, April 1994.
67. **NOZICK**, Robert. "Newcomb's Problem and Two Principles of Choice" (1969). In **NOZICK**, Robert. *Socratic Puzzles*. Cambridge-MA e London: Harvard University Press, ©1997, 2000, (pp. 45-73).
68. **NOZICK**, Robert. *Anarquia, estado e utopia. (Anarchy, State and Utopia, 1974a)*. Tradução: Ruy Jungmann. Rio de Janeiro: Jorge Zahar Editor, 1991.

69. **NOZICK**, Robert. "Reflections on Newcomb's Problem" (1974b). In **NOZICK**, Robert. *Socratic Puzzles*. Cambridge-MA e London: Harvard University Press, ©1997, 2000, (pp. 74-84).
70. **NOZICK**, Robert. *Philosophical Explanations*. Cambridge-MA: Harvard University Press, ©1981, 2000.
71. **NOZICK**, Robert. *Examined Life: Philosophical Meditations*. Touchstone Books, 1990.
72. **NOZICK**, Robert. *The Nature of Rationality*. Princeton-NJ: Princeton University Press, ©1993, 1995.
73. **NOZICK**, Robert. "Letter to the editor: What's best". In: *London Review of Books*, 10 March 1994, pp. 4-5.
74. **NOZICK**, Robert. "Symbolic Utility". In **BASU**, K., **PATTANAİK**, P. e **SUZUMURA**, K. (eds.). *Choice, Welfare, and Development: A Festschrift in Honour of Amartya K. Sen*. New York: Oxford University Press, 1995, (pp. 110-140).
75. **NOZICK**, Robert. *Socratic Puzzles*. Cambridge-MA e London: Harvard University Press, ©1997, 2000.
76. **NOZICK**, Robert. *Invariances: The Structure of the Objective World*. Harvard University Press, 2001.
77. **PAPINEAU**, David. *The Roots of Reason: Philosophical Essays on Rationality, Evolution, and Probability*. New York: Oxford University Press, 2003.
78. **PLATÃO**, Fedro, *Cartas, O Primeiro Alcibiades (Diálogos, volume V)*. Tradução: Carlos Alberto Nunes. Belém: Universidade Federal do Pará, 1975.
79. **PLATÃO**. *Protágoras, Górgias, Fedão*. 2.ed. Tradução: Carlos Alberto Nunes. Belém: EDUFPA, 2002.
80. **PLATÃO**. *República, A*. 8.ed. Tradução: Maria Helena da Rocha Pereira. Lisboa: Fundação Calouste Gulbenkian, 1996.
81. **POUNDSTONE**, William. *Labyrinths of Reason: Paradox, Puzzles, and the Frailty of Knowledge*. New York: Anchor Press/Doubleday, 1988.
82. **POUNDSTONE**, William. *Prisoner's Dilemma: John von Neumann, Game Theory and the Puzzle of the Bomb*. New York: Anchor Press/Doubleday, ©1992, 1993.
83. **PUTNAM**, Hilary. *The Collapse of the Fact/Value Dichotomy and other essays*. Cambridge-MA: Harvard University Press, ©2002, 2004.
84. **RAMSEY**, F. P. "Truth and Probability" (1926). In **MELLOR**, D. H. (ed.). *Philosophical Papers: Frank Plumpton Ramsey*. New York: Cambridge University Press, 1990, (pp. 52-109).
85. **RASMUSEN**, Eric. *Games and Information: An Introduction to Game Theory*. 3rd ed. Malden-MA, Oxford-UK: Blackwell Publishers, ©2001, 2002.
86. **RECANATI**, François. *Relativized Propositions (Draft 2)*. Disponível em: <<http://jeannicod.ccsd.cnrs.fr/documents/disk0/00/00/02/69/index.html>, 2002>. Acesso em: 13/fevereiro/2005.

87. **REIS**, Bruno Pinheiro W. *Reflexões sobre a epistemologia de Popper e o individualismo metodológico*. Belo Horizonte: UFMG, 1988. (Cópia cedida pelo autor).
88. **RESCHER**, Nicholas. *Paradoxes: their roots, range, and resolution*. Peru-Illinois: Open Court, 2001.
89. **RESNIK**, Michael D. *Choices: An Introduction to Decision Theory*. Minneapolis: The University of Minnesota Press, ©1987, 2000.
90. **RICHARDSON**, Henry S. *Practical Reasoning about Final Ends*. New York: Cambridge University Press, 1997.
91. **ROUDINESCO**, Elizabeth e **PLON**, Michel. *Dicionário de psicanálise. (Dictionnaire de la psychanalyse, 1997)*. Tradução: Vera Ribeiro e Lucy Magalhães. Rio de Janeiro: Jorge Zahar Editor, 1998.
92. **SAHLIN**, Nils-Eric. *The Philosophy of F. P. Ramsey*. New York: Cambridge University Press, 1990.
93. **SAINSBURY**, R. M. *Paradoxes*. 2nd ed. Cambridge, UK: Cambridge University Press. ©1995, 2001.
94. **SCHELLING**, Thomas C. *The Strategy of Conflict*. New York: Oxford University Press, 1963. (Capítulo 1, pp. 3-80).
95. **SCHMIDTZ**, David. *Introductions*. In **SCHMIDTZ**, David (ed.) *Robert Nozick*. Cambridge-UK: Cambridge University Press, 2001, (pp. 1-9).
96. **SCHMIDTZ**, David e **WRIGHT**, Sarah. "What Nozick Did for Decision Theory". In *Midwest Studies in Philosophy*, 28 (1), pp. 282-294, 2004.
97. **SCHUESSLER**, Alexander A. *A Logic of Expressive Choice*. Princeton-NJ: Princeton University Press, 2000.
98. **SEN**, Amartya. "Behaviour and the Concept of Preference" (1973). In **ELSTER**, Jon (ed.). *Rational Choice*. New York: New York University Press, 1986, (pp. 60-81).
99. **SEN**, Amartya. "Rational Fools" (1977). In **SEN**, Amartya. *Choice, Welfare and Measurement*. Cambridge-MA e London: Harvard University Press, ©1982, 1999, (pp. 84-106).
100. **SEN**, Amartya. *Sobre ética e economia. (On ethics & economics, ©1987, 1988)*. Tradução: Laura Teixeira Motta. São Paulo: Companhia das Letras, 1999.
101. **SEN**, Amartya. "Internal Consistency of Choice". In *Econometrica*, Vol. 61, N° 3, pp. 495-521, maio 1993.
102. **SEN**, Amartya. "Rationality and Social Choice". In *The American Economic Review*, Vol. 85, Issue 1, pp. 1-24, março 1995.
103. **SEN**, Amartya. *Reason before Identity (The Romanes Lecture for 1998)*. New York: Oxford University Press, 1999.
104. **SEN**, Amartya. "East and West: The Reach of Reason". In *New York Review of Books*, Vol. 47, N° 12, pp. 33-38, julho 2000. Disponível em: <<http://www.mtholyoke.edu/acad/intrel/senreason.htm>>. Acesso em 3/maio/2003.

105. **SEN**, Amartya. *Rationality and Freedom*. Cambridge-MA e London: The Belknap Press of Harvard University Press, 2002.
106. **SHERMAN**, Nancy. *Making a Necessity of Virtue: Aristotle and Kant on Virtue*. New York: Cambridge University Press, 1997.
107. **SHERMAN**, Nancy. *The Fabric of Character: Aristotle's Theory of Virtue*. New York: Oxford University Press, 1989.
108. **SINGER**, Peter. *A Darwinian Left: politics, evolution and cooperation*. New Haven and London: Yale University Press, 2000.
109. **SKYRMS**, Brian. *The Dynamics of Rational Deliberation*. Cambridge-MA e London: Harvard University Press, 1990.
110. **SKYRMS**, Brian. *Choice and Chance: An Introduction to Inductive Logic*. 4th ed. Belmont-CA: Wadsworth/Thomson Learning, 2000.
111. **SKYRMS**, Brian. *The Stag Hunt and the Evolution of Social Structure*. Cambridge, UK: Cambridge University Press, 2004.
112. **SOBEL**, Jordan Howard. "Not Every Prisoner's Dilemma Is a Newcomb Problem". In **CAMPBELL**, Richmond; **SOWDEN**, Lanning (eds.). *Paradoxes of Rationality and Cooperation: Prisoner's Dilemma and Newcomb Problem*. Vancouver: The University of British Columbia Press, 1985, (pp. 263-74).
113. **SOBEL**, Jordan Howard. *Taking Chances: Essays on Rational Choices*. Cambridge, UK: Cambridge University Press, 1994.
114. **SÓFOCLES**, "Antígona". In *A trilogia tebana*. Tradução: Mário da Gama Kury. 8.ed. Rio de Janeiro: Jorge Zahar Editor, ©1989, 1998, (pp. 195-253).
115. **TAYLOR**, Michael. *The Possibility of Cooperation*. Cambridge, UK: Cambridge University Press, 1987 (rev. ed.).
116. **TIERCELIN**, Claudine. "Peirce on Norms, Evolution and Knowledge". *Transactions of the C. S. Peirce Society*, vol. XXXIII, n°1, pp. 35-58, winter 1997. Disponível em: <http://jeannicod.ccsd.cnrs.fr/documents/disk0/00/00/02/15/ijn_00000215_00/ijn_00000215_00.htm>. Acesso em 03/fevereiro/2005.
117. **VAN PARIJS**, Philippe. *Que é uma sociedade justa, O – Introdução à prática da filosofia política. (Qu'est-ce qu'une société juste?, 1991)*. Tradução: Cíntia Ávila de Carvalho. São Paulo: Editora Ática, 1997.
118. **VLASTOS**, Gregory. *Socrates, Ironist and Moral Philosopher*. Ithaca: Cornell University Press, 1991.
119. **WAKKER**, Peter P. *History of the Term Dutch Book*. Disponível em: <<http://www1.fee.uva.nl/creed/wakker/miscella/Dutchbk.htm>>. Acesso em 08/junho/2003.
120. **WILSON**, Edward O. *Consilience: The Unity of Knowledge*. New York: Vintage Books, ©1998, 1999.
121. **WITTGENSTEIN**, Ludwig. *Investigações Filosóficas (Philosophische Untersuchungen, 1953)*. Tradução: José Carlos Bruni. 2.ed. São Paulo: Abril Cultural, 1979.

ÍNDICE REMISSIVO

- ação coletiva, 53, 119
ação dominada, 122, 126
ação dominante, 81, 98, 99, 101, 102, 103, 122, 123, 126
aceitação, 4, 5, 47
acordo, 49, 52, 57, 58, 79, 87, 112, 131, 141, 145
adaptação, 133, 135, 136
Agamenon, 27
agente, ii, xii, 3, 8, 10, 11, 12, 13, 14, 19, 22, 25, 26, 27, 28, 30, 31, 36, 38, 39, 40, 43, 44, 45, 50, 54, 57, 64, 65, 72, 84, 87, 88, 89, 91, 92, 105, 110, 112, 115, 120, 121, 122, 124, 140, 141, 142, 143
Ainslie, 109, 111, 112, 113, 115, 140, 145, 150
altruísmo, 33, 34, 84, 139
anel de Gíges, 54
Antígona, 141, 157
Aquiles, 113
Aristóteles, v, ix, 16, 46, 113, 139, 151
Aristotle, v, 19, 151, 157
asno de Buridan, 28, 49
auto-interesse, i, 1, 16, 17, 21, 22, 25, 29, 31, 32, 43
Axelrod, 136
Barcan Marcus, 27, 48, 49, 50, 144
barganha, 57, 58, 113, 145
Bäuerle, 7
Bayes, i, 11, 72, 73, 146, 147
bem coletivo, 77, 81
bem comum, 54, 76
bem-estar, ix, 1, 8, 24, 25, 26, 27, 39, 58, 64, 65, 68, 84, 139
Bentham, 1, 90
Bêni, 78
Berti, 139
Binmore, 83
Blackburn, i, xii, 1, 2, 23, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 48, 50, 65, 75, 81, 82, 83, 84, 110, 126, 130, 138
Broome, 94
Butler, 32, 33, 37
Campbell, 92
caráter heurístico, 65
carona, 55, 56
causa, 7, 17, 84, 90, 92, 97, 98, 103, 116, 124, 131, 144, 147
causa comum, 90, 92, 97, 98, 124
causação retroativa, 95, 103
causal, i, xi, 7, 8, 35, 86, 87, 88, 91, 92, 93, 94, 97, 98, 101, 102, 103, 104, 105, 106, 107, 116, 117, 118, 121, 122, 124, 125, 128, 135, 140, 147, 148
Cemig, viii, ix
Christensen, 119
ciência, ii, iv, 31, 52, 128, 142
coerência, 13, 14, 15, 43, 44, 128
Cohen, 4, 5, 7
coletivo, 1, 55, 68, 121
compromisso, 58, 112, 146
confiança, 9, 58, 79, 80, 95, 106, 118, 121, 124, 143
consistência, i, ii, 1, 2, 3, 4, 5, 8, 11, 12, 13, 15, 17, 18, 19, 20, 21, 29, 38, 43, 46, 47, 48, 49, 64, 115, 130, 139
contexto, ii, vii, viii, x, xi, xii, 2, 7, 14, 17, 19, 20, 22, 24, 29, 30, 41, 44, 45, 47, 54, 61, 62, 64, 65, 85, 87, 108, 115, 119, 120, 138, 140, 142, 145
contrafactuais, 104, 105, 122
contrafactual, 7, 62, 92, 124
cooperação, xii, 53, 54, 55, 57, 58, 64, 67, 68, 69, 71, 72, 73, 75, 76, 78, 79, 80, 81, 84, 88, 112, 121, 122, 123, 124, 126, 136, 143, 144, 145
coordenação, viii, 77, 79, 80
Craig, 14, 19, 86
crença, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 15, 19, 35, 43, 47, 70, 74, 84, 131, 132, 133, 135, 137, 145, 146, 147
Cresswell, 7
Cullity, v, 138
culpa, 27, 49, 116, 145
cultura, 20, 22, 23, 24, 54, 132
curvas hiperbólicas, 111, 150
Dancy, 7
Davis, 58, 61
Dawkins, 34
decisão, i, ii, v, vi, viii, ix, x, xi, xii, 1, 2, 3, 8, 9, 11, 20, 27, 28, 35, 36, 42, 50, 51, 52, 53, 58, 59, 61, 63, 68, 72, 73, 78, 81, 83, 86, 87, 88, 89, 91, 92, 93, 95, 96, 97, 98, 102, 103, 104, 105, 107, 109, 110, 115, 117, 118, 121, 122, 123, 124, 125, 126, 127, 129, 131, 135, 138, 140, 141, 142, 147
deliberação, ix, 31, 37, 46
dependência, 54, 90, 97, 98, 99, 102, 103, 134
deriva, 143, 144
desconto, i, 55, 70, 71, 109, 110, 111, 144, 150
desejo, v, 7, 8, 32, 35, 109, 119
Dewey, 128

- dilema do prisioneiro, iii, i, ii, xii, 29, 45, 50, 52, 53, 54, 55, 56, 58, 59, 60, 64, 66, 67, 68, 69, 70, 72, 73, 75, 76, 77, 78, 79, 80, 81, 82, 84, 85, 86, 113, 118, 121, 122, 123, 124, 125, 126, 127, 136, 140, 143, 148
- Diógenes, iv
- Diógenes Laértios, iv
- Dixit, 62
- dominação, 23, 95, 97, 98, 99, 100, 101, 104, 106, 122
- Dresher, 52
- Durkheim, 22
- Dutch Book*, 12, 15, 157
- economia, xii, 1, 16, 22, 24, 25, 26, 27, 51, 156
- educação, 41, 84
- egoísmo, 8, 22, 32, 33, 35, 36, 139
- Elgin, 14
- Elster, 67, 109
- empírica, 3, 31, 32, 37, 40, 41, 82, 126, 134
- empírico, 31, 41, 82
- engenharia, 16
- equilíbrio bayesiano, 72, 73, 74
- equilíbrio cooperativo, 80
- equilíbrio de Nash, 52, 57, 66, 68, 69, 72, 73, 77
- equilíbrio reflexivo, 14, 145
- Êsquilo, 27
- estado, ii, ix, 24, 32, 54, 56, 68, 79, 89, 90, 92, 97, 98, 100, 101, 102, 103, 105, 117, 137, 143, 154
- estratégia, 22, 47, 62, 66, 67, 74, 77, 80, 82, 131, 132, 138, 148, 151
- ética, ii, 2, 16, 20, 22, 24, 25, 26, 29, 42, 117, 127, 128, 138, 140, 141, 143, 144, 156
- evidencial, i, 86, 87, 88, 91, 92, 93, 94, 97, 98, 104, 105, 106, 107, 117, 118, 121, 122, 124, 125, 128, 133, 135, 136, 140, 147
- evolução, 33, 34, 131, 133, 134, 135, 136, 143
- expectativa, 22, 45, 58, 111
- expectativas, x, 54, 61, 94, 142, 145
- exponencial, 109
- falácia, 31, 34, 35, 46
- fator de desconto, 70, 71
- felicidade, ix, 26, 32, 35, 37, 39, 139
- filosofia, ii, iv, 1, 34, 35, 52, 86, 113, 157
- fins, vi, vii, ix, xi, 16, 19, 29, 43, 117, 127, 128, 129, 130, 131, 145
- Fitelson, 91, 92
- Flood, 52
- Folk theorem*, 72
- forma extensiva, 59
- forma normal, 59, 60, 61, 66, 67
- Freud, 33
- Friedman, 72
- função, xii, 2, 9, 10, 36, 38, 42, 43, 58, 72, 73, 84, 94, 108, 109, 116, 118, 128, 129, 132, 135, 136, 139
- Gadamer, 121
- Gärdenfors, 88, 89
- Gardner, 95, 103
- Gaus, 87, 119, 125, 128, 132
- Gaut, v, 138
- Gauthier, 83
- gene, 34, 100, 124
- genes, 34, 35, 101, 136
- Gensler, 4, 5
- Gibbard, 62, 92, 98, 102, 104, 105, 106
- Gibbons, 59, 60, 61, 68, 70, 71, 72, 73, 74
- Glock, 141
- grau de crença, 6, 7, 8, 9, 10, 11, 12, 91, 92, 146, 147
- Guyer, 75
- Hacking, 12, 146, 147
- Hampton, 128, 130
- Hardin, 56
- Harper, 62, 92, 98, 102, 104, 105, 106
- Hausman, 77
- Heidegger, 129
- Heil, 7
- heurística, 102
- Hobbes, 33, 54
- homo economicus*, 31, 39, 43
- Honderich, 7, 48, 86
- Hornsby, 34, 35
- Hume, v, 1, 2, 42, 54, 55, 70, 79, 91, 111, 130, 136, 139, 144
- humeano, vi, 130, 138
- Hurley, 106, 122, 124, 125
- identidade, 23, 24, 81, 108, 119, 120, 129, 139
- ilusão, 35, 38, 46, 101, 103, 124, 125
- imputação retroativa, 116
- inclusive fitness*, 136
- independência, 94, 148
- individual, ii, iii, xii, 22, 27, 53, 55, 68, 76, 77, 88, 117, 119, 121, 122, 137
- indução retroativa, 68, 81
- informação, 19, 62, 66, 68, 69, 70, 72, 73, 137
- interesse, ii, vii, xi, xii, 14, 19, 21, 22, 29, 32, 36, 37, 39, 53, 54, 56, 58, 77, 79, 81, 111, 112, 131, 136, 148
- interesse próprio, ii, xi, 22, 37, 56, 58, 82
- James, 91, 92, 128, 152, 154
- Jeffrey, 6, 90, 91, 102
- jogo da coordenação, 55, 70, 79, 80
- jogo de soma não-zero, 58
- jogo dinâmico, 69, 70, 71
- jogo do covarde, 77, 78, 79

- jogo do impasse, 78
jogo empírico, 40, 41, 65
jogo estático, 69, 71, 72, 144
jogo estratégico, 54
jogo não-cooperativo, 57, 58
jogo teórico, 41, 65, 82
Jonas, 143, 144, 154
Joyce, 91, 92, 152
Kahnemann, 28
Kant, v, 46, 49, 117, 129, 133, 134, 138, 139, 157
Kavka, 83
Kenney, 28
Keynes, 5, 6
Kreps, 62, 73
Laburthe-Tolra, 22
Lacey, 97, 98, 104, 105, 116, 119, 133
Lewis, 124, 148
liberdade, 20, 22, 23, 26, 27, 29, 30, 35, 43, 65, 115, 138, 143, 144
linguagem, iv, 33, 35, 132, 137
livre-arbítrio, 103
Luce, 58
matriz de resultados, 59, 61, 62, 63, 75, 81, 93, 95, 96, 122, 123, 124, 125
Mauss, 22
maximização da utilidade esperada, ii, 31, 36, 37, 39, 82, 84, 94, 97, 140, 147
McLennen, 83
McPherson, 77
Mellor, 3, 7
metadecisão, 114, 143, 144
Milgrom, 73
Mill, 1, 31, 32
miopia, 70, 71
Monod, 134
moral, 27, 48, 50, 128, 138, 146
Morgenstern, 3, 51, 88
Morton, 54, 57, 73, 142, 152
mundos possíveis, 10, 145
mutações, 131, 134, 144
Myerson, 60, 70
Nalebuff, 62
não-cooperação mútua, 77, 78, 79
Nash, 57, 68, 69, 72, 73, 77, 78, 80, 145
natureza da racionalidade, 85, 87, 129, 135, 137
nível subpessoal, 34
normativa, 31, 37, 46, 82, 83, 85, 87, 107, 127, 128, 140
normatividade, 4, 5, 65, 128
normativo, 4, 31, 39, 43, 67, 127, 141
notícia, 91, 105, 106, 122
Nozick, iii, i, ii, iii, xii, 2, 52, 59, 60, 86, 87, 90, 95, 97, 98, 99, 100, 102, 103, 104, 106, 107, 108, 109, 110, 112, 113, 114, 115, 116, 117, 118, 119, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 143, 144, 145, 152, 153, 154, 156
observador, 18, 29, 46, 63, 65, 85, 141, 142
otário, 75, 76, 123
paixões, i, x, 1, 31, 32, 42, 130, 138, 144, 145
paradoxo, ii, 37, 45, 46, 47, 48, 83, 84, 110, 119, 138
paradoxo socrático, 138
Pareto, 24, 26, 52, 57, 68, 73, 77, 78, 79
Peirce, 134, 157
peso, 36, 92, 93, 106, 118, 124, 127
Platão, iv, v, 46, 54, 144
política, 31, 51, 54, 86, 144, 157
ponto de equilíbrio, 66, 67, 68, 145
Poundstone, 46, 51, 60, 67, 75, 76, 77, 78, 79, 123
pragmatismo, 134
prática social, 120
preferência, 9, 18, 20, 29, 38, 39, 41, 42, 47, 55, 56, 63, 64, 71, 76, 82, 84, 91, 110, 125, 142
princípio, xii, 4, 5, 19, 21, 23, 31, 32, 36, 37, 42, 48, 49, 50, 82, 84, 93, 95, 96, 97, 98, 99, 100, 101, 106, 107, 108, 112, 113, 114, 115, 118, 119, 122, 124, 125, 126, 127, 128, 135, 140, 142
probabilidade, xii, 2, 3, 5, 6, 8, 10, 11, 12, 13, 15, 43, 70, 71, 72, 74, 80, 90, 91, 92, 94, 96, 99, 100, 101, 104, 105, 114, 119, 124, 146, 147
probabilidade condicional, 10, 90, 91, 92, 96, 101, 146
probabilidade subjetiva, 3, 8, 11, 72, 146
probabilidades condicionais, 92, 96, 97, 102, 103, 104, 105, 122, 146
problema de Newcomb, i, xii, 50, 83, 86, 90, 92, 95, 96, 97, 98, 102, 103, 104, 107, 118, 121, 122, 123, 124, 125, 142, 148
processo de decisão, vii, ix, x, 37
psicologia, xii
Putnam, 128, 142
questão socrática, 16
racionalidade, iii, i, ii, v, vi, xii, 1, 2, 3, 8, 10, 11, 13, 15, 16, 17, 21, 22, 23, 28, 29, 42, 43, 44, 45, 46, 50, 55, 58, 63, 64, 65, 68, 76, 79, 81, 84, 85, 86, 87, 92, 108, 115, 122, 126, 128, 129, 130, 131, 132, 133, 135, 136, 137
racionalidade instrumental, 128, 129, 131, 135
racionalidade prática, v, vi
Raiffa, 28, 58
Ramsey, i, xii, 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19, 20, 29, 38,

- 39, 43, 50, 51, 63, 65, 83, 84, 88, 89,
91, 146, 154, 155, 156
Rand Corporation, 52
Rapoport, 75
Rasmusen, 57, 58, 62, 66, 67, 68, 71, 72,
73, 74
Rawls, xi, 26
razão, v, vi, 2, 21, 22, 24, 26, 28, 29, 37,
39, 42, 43, 44, 50, 82, 83, 87, 98, 104,
114, 128, 129, 130, 132, 134, 136,
138, 144, 145, 154
Recanati, 19
recompensa, 36, 75, 109, 110, 111, 112,
113, 115, 123, 150
regresso infinito, 66, 145
Reis, ii, iii, v, 154
Rescher, 46, 47, 48
Resnik, 13, 90, 91, 92
respeito, v, 10, 12, 14, 15, 23, 38, 41, 42,
48, 49, 61, 69, 72, 79, 89, 98, 104,
119, 122, 129, 130, 131, 134, 142, 143
responsabilidade, 58, 64, 143, 144
revolução copernicana, 134
Richardson, xi
Roberts, 73
Rousseau, 79
Russell, 2, 7, 77
Sahlin, 6, 10, 12, 88, 89, 91, 146
Sainsbury, 8, 46, 47, 48, 96
Samuelson, 18, 20, 65
satisficing, 73
Savage, 3, 88
Schelling, 63, 65
Schmidtz, 90, 96, 102, 103
Schuessler, 119, 120, 121
seleção natural, 131, 134
self-defeating, 111
Selten, 69
Sen, i, xii, 1, 2, 14, 16, 17, 18, 19, 20, 21,
22, 23, 24, 25, 26, 27, 28, 29, 43, 44,
50, 54, 58, 63, 64, 65, 81, 84, 112,
115, 120, 136, 138, 139, 145, 155
Sherman, ix, xi
significado simbólico, 116, 117, 140, 141
significados simbólicos, 87, 117, 126,
128, 130
Simon, 2, 73, 81, 151
simpatia, iv, 21, 23, 136, 139
Skyrms, 55, 57, 68, 70, 79, 80, 147
Slovik, 28
Smith, 21, 154
Sobel, 92, 148
Sócrates, 142
Sófocles, 141
soma zero, 33, 57, 68
Sosa, 7
sunk cost, 109, 115
taxa de desconto, xii, 70, 71, 109, 111,
144, 150
taxas de desconto, 55, 71, 111, 150
Taylor, 54, 55, 56, 70, 77, 78, 79, 139
tentação, 31, 51, 55, 56, 75, 76, 110, 112,
114, 115, 123
teoria da ação, i, vii, 8, 63, 86, 87, 108,
115, 118, 121, 131, 133
teoria da decisão, iii, ii, xii, 2, 3, 47, 86,
88, 90, 98, 108, 109, 116, 117, 119,
128, 129, 131, 142, 147, 151
teoria da escolha racional, 37, 38, 51, 147
teoria da evolução, 33, 108, 134
teoria dos jogos, i, ii, xii, 41, 44, 45, 51,
52, 54, 57, 58, 61, 65, 66, 68, 80, 82,
84, 85, 86, 88, 126, 143, 144, 147
Tiercelin, 134
tit-for-tat, 74
toxina, 83, 84, 110
tragédia, 56
transgressão, 143
Tucker, 52
Tversky, 28
Ulisses, 112
utilidade, ii, xii, 1, 2, 3, 8, 9, 10, 11, 18,
24, 25, 26, 29, 31, 37, 38, 39, 40, 41,
45, 47, 49, 58, 67, 71, 83, 84, 88, 89,
90, 91, 93, 95, 96, 97, 98, 99, 100,
101, 102, 103, 104, 105, 106, 107,
109, 110, 111, 114, 116, 117, 118,
119, 121, 122, 124, 125, 126, 127,
128, 130, 131, 135, 139, 140, 144, 147
valor, ii, v, vi, xii, 8, 9, 10, 23, 24, 26, 27,
50, 61, 65, 70, 71, 75, 82, 87, 89, 91,
92, 93, 98, 105, 106, 107, 110, 111,
116, 118, 119, 120, 124, 125, 126,
127, 128, 129, 131, 135, 136, 138,
140, 141, 142, 145, 150
valor intrínseco, 65, 129
valor presente, 70, 71
valor simbólico, 87, 125, 126, 143
valor-decisão, ii, xii, 87, 98, 106, 107,
118, 124, 125, 126, 127, 128, 131,
135, 136, 140, 141, 142, 145
valores simbólicos, 120, 142
verdade, iv, 4, 5, 8, 9, 10, 21, 31, 32, 82,
131, 132
verossimilhança, 101, 102, 146, 147
von Neumann, 3, 51, 52, 88, 155
Wakker, 12
Warnier, 22
Whitehead, 56
Williams, vi, 138
Wilson, 73
Wittgenstein, 2, 141, 153
Wright, 90, 96, 102, 103