

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA

*Autoridade, incerteza e responsabilidade:*  
ensaios sobre a legitimidade do regulador à  
luz da Grande Crise.

Tese de Doutorado

Ramiro de Ávila Peres

Orientador: Alfredo Storck

Porto Alegre, 16 de novembro de 2018.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA

*Autoridade, incerteza e responsabilidade:*  
ensaios sobre a legitimidade do regulador à  
luz da Grande Crise.

Tese apresentada ao  
Departamento de Pós-Graduação  
em Filosofia da Universidade  
Federal do Rio Grande do Sul  
como requisito parcial à obtenção  
do grau de Doutor em Filosofia

Ramiro de Ávila Peres

Orientador: Alfredo Storck

Porto Alegre, 16 de novembro de 2018.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA

*Autoridade, incerteza e responsabilidade:*  
ensaios sobre a legitimidade do regulador à  
luz da Grande Crise.

Tese apresentada ao  
Departamento de Pós-Graduação  
em Filosofia da Universidade  
Federal do Rio Grande do Sul  
como requisito parcial à obtenção  
do grau de Doutor em Filosofia

**Banca Examinadora:**

---

Prof. Dr. Alfredo Storck – UFRGS

---

Prof. Dr. Paulo Caruso MacDonald – UFRGS

---

Prof. Dr. Nikolay Steffens – UFRGS

---

Prof. Dr. Hélio Alves – UFRGS

---

Prof. Dr. Cláudio Ari Mello – PUC/RS

Porto Alegre, 16 de novembro de 2018.

## Agradecimentos

Aos colegas e amigos do PPG: a amizade, as discussões e o estímulo intelectual. Em especial, Mariana, Patrícia, Paulo e Alfredo: o que houver de bom nestas linhas deve-se a vocês.

Ao Prof. Balthazar Barbosa Filho (*in memoriam*) e ao Prof. Paulo Faria: o pouco que sei sobre futuros contingentes.

À UFRGS: ter sido meu segundo lar há mais de catorze anos.

Ao Banco Central: o apoio – em especial, o afastamento que me permitiu concluir este projeto no prazo.

Aos colegas do Banco Central: a fofoca fraterna, a companhia, o trabalho, o cafezinho e o pastel.

Aos colegas de Altruísmo Eficaz – em especial, a Fernando Moreno: a esperança.

Ao Prof. Emmanuel Picavet: o convite e a acolhida em Paris – e, principalmente, os importantes *insights* contidos em *Choix Rationnel et Vie Publique*.

A minha família – em especial, a minha esposa, Sheyla, e a meu cão, Coelho: o fato de que não sou só um *rational fool*.

"...Ideas! That was their world now! Big questions and big answers, about life, and how you had to live it, and what you were for. New ideas spilled into Darktan's weary head.

And among the ideas, in the middle of his head, he saw the little figure of Dangerous Beans.

Darktan had never talked much to the little white rat or the little female who scurried around after him and drew pictures of the things he'd been thinking about. Darktan liked people who were *practical*.

But now he thought: he's a trap-hunter! Just like me! He goes ahead of us and finds the dangerous ideas and thinks about them and traps them in words and makes them safe and shows us the way through.

We *need* him... we need him *now*. Otherwise, we're all running around like rats in a barrel..."

(Terry Pratchett, *The Amazing Maurice and his educated rodents*, 188-89)

[...] a visiting angel offers to answer truthfully any single question put by the philosophers at an international conference. After much deliberation, the philosophers eventually decide to ask:

(Q4) What is the ordered pair whose first member is the question that would be the best one for us to ask you, and whose second member is the answer to that question?

The angel, who is both witty and well-informed, answers

(A4) The ordered pair whose first member is the question you just asked me, and whose second member is this answer I am giving you.

(Scott & Scott, 1999, 331)

## Sumário

<b>Sumário</b>	<b>6</b>
<b>Resumo</b>	<b>8</b>
<b>Introdução</b>	<b>9</b>
<b><i>Parte I: Paradoxos da Razão e da Autoridade</i></b>	
<b>Capítulo 1: Do desacordo ao paradoxo da autoridade - uma análise da concepção de serviço de autoridade à luz da teoria do “ponto-cego”, de R. Sorensen</b>	<b>17</b>
1 Introdução	18
2 O caráter paradoxal da autoridade	19
2.1 A independência de conteúdo e a “concepção de serviço”	19
2.2 Novos paradoxos	21
3 Paradoxos e <i>pontos-cegos</i>	22
3.1 Racionalidade e justificação	22
3.2 Pontos-cegos	25
3.3 Pontos-cegos da razão prática	26
4 Dissolvendo o paradoxo	27
5 Conclusões	30
<b>Capítulo 2: Interpretando a reivindicação de autoridade legítima: análise da refutação de Joseph Raz à incorporação de normas morais ao direito.</b>	<b>32</b>
1 Introdução	32
2 O argumento de Raz	34
3 A Objeção: Interpretação e Autoridade	37
4 Conclusão: Por que o argumento anti-incorporação falha?	43
<b>Capítulo 3: O que Tartaruga dirá a Aquiles, ou “levando a sério a interpretação tradicional do argumento da batalha naval”</b>	<b>46</b>
<b><i>Parte II: Da Incerteza à Responsabilidade</i></b>	
<b>Capítulo 4: A Escolha de Greenspan e o Elogio da Passividade – considerações sobre decisões da autoridade em contextos de grande incerteza</b>	<b>56</b>
1 Introdução: Ceticismo e Passividade	57
1.1 Incerteza e risco	60
1.1.1 Causas da incerteza: reflexividade, interações sociais e eficiência informacional	63
1.2. Eventos catastróficos	66
1.3 O modelo do agente racional	68
1.4 Porque não é simples fugir de atribuições de probabilidades	71
2 Incerteza e filosofia política normativa	74
	6

2.1 A pervasividade da objeção: teorias da justiça e legitimidade	74
3 Greenspan: um estudo de caso de decisão sob incerteza	84
3.1 Incentivos, coordenação e segurança	85
3.3 Autoridade, autonomia e representação	88
4 Conclusão	92
<b>Capítulo 5: A Crise Moral: a responsabilidade de administradores de instituições financeiras e o argumento da inevitabilidade</b>	<b>94</b>
1 Introdução: a explicação cognitiva e a explicação moral da crise financeira	94
2 A crise inevitável	97
2.1 A tese-alvo: o argumento a partir do inevitável	100
3 Krishna, Shakespeare e Tuld: do fatalismo à teoria da decisão racional	101
3.1 Resultado e considerações de responsabilidade pessoal	101
3.2 Fatalismo absurdo	104
3.3 Fatalismo e cegueira	105
4 Conclusão	108
<b>Capítulo 6: As <i>Eumenides</i> e a Crise: responsabilidade, risco moral e dissuasão no sistema financeiro</b>	<b>110</b>
Introdução	111
1 Distribuindo riscos de acordo com o princípio da diferença	114
A relação entre incentivos e princípios	117
2 Garantia implícita e risco moral	119
Supervisão e risco moral	125
Conclusão	128
<b>Conclusão</b>	<b>130</b>
Racionalidade	131
Incerteza e Autoridade	134
Crise e Política	135
<b>Referências</b>	<b>138</b>

## **Resumo**

A presente tese é composta de seis artigos independentes entre si, agrupados como capítulos em duas seções, de acordo com a similaridade do tema. A primeira parte trata de problemas filosóficos relacionados à teoria da autoridade jurídica de Raz e à teoria da decisão racional. A segunda parte trata, principalmente, da atuação de autoridades regulatórias em situações de incerteza – em especial, usando como exemplo a atuação de Bancos Centrais no contexto da Crise de 2008. Abordam-se, respectivamente: o paradoxo do desacordo (a potencial inconsistência em discordar de uma autoridade legítima), a refutação de Raz à incorporação de conceitos morais ao direito, o problema da incerteza e futuros contingentes, a legitimidade da autoridade em situações de incerteza, a objeção fatalista à reprovação moral por eventos econômicos, e a necessidade de instituir sanções para dissuadir financistas a assumirem riscos excessivos.

Palavras-chave: Filosofia Política, Autoridade, Teoria da Decisão

## **Abstract**

This dissertation is made of six independent articles, grouped as chapters in two parts, according to similarities between their themes. The first part deals with philosophical problems related to Raz's theory of legal authority and to rational decision theory. The second part deals mainly with the performance of regulatory authorities in situations of uncertainty - in particular, using as example the performance of Central Banks in the context of the 2008 Crisis. We discuss, in order: the paradox of disagreement (the potential inconsistency of disagreeing with a legitimate authority), Raz's refutation to the incorporation of moral concepts into law, the problem of uncertainty and future contingents, the legitimacy of authority in situations of uncertainty, the fatalistic objection to moral disapproval for economic events, and the need to impose sanctions to deter finance investors from taking excessive risks.

Keywords: Political Philosophy, Authority, Decision Theory



## Introdução

Quando, em meados de setembro de 2008, o Banco Lehman Brothers se declarou insolvente, precipitando uma crise sistêmica, já se sabia que algo estava a caminho, pois já atingira tanto o banco inglês Northern Rock quanto o americano Bear Sterns. Ainda no ano anterior, o mercado de hipotecas securitizadas havia quebrado, após o estouro da bolha imobiliária. Tais desastres poderiam ter sido evitados por medidas regulatórias ou macroeconômicas adequadas, mas isso é julgar a partir do privilégio da retrospectiva – na época, esses fatos eram considerados quase impossíveis; em sua defesa, as autoridades e os financistas alegaram que não os poderiam ter antecipado – que estavam num contexto de grave incerteza e que havia desacordo sobre que medidas seriam adequadas.

O principal ponto deste trabalho é que é justamente para lidar com situações de incerteza e de distribuição de responsabilidades que precisamos de autoridades; investigamos os possíveis óbices que o desacordo e a limitação cognitiva (i.e., ao levantamento e processamento de informações) podem oferecer à decisão (e, conseqüentemente, à legitimidade de autoridades), e que princípios podem-se utilizar para preencher essas lacunas. O primeiro, sugerido no quarto capítulo, é reconhecer que nossas instituições têm um dever especial (e, frequentemente, impopular) de *prevenir* desastres – que deriva de considerações sobre a teoria da legitimidade e a teoria da justiça. O segundo, nos dois últimos capítulos, aduz a necessidade de atribuição de responsabilidade na definição de incentivos e de punições. Mesmo com bons modelos para agregar informações e projetar as conseqüências de uma decisão, ainda resta uma boa dose de incerteza sobre os resultados; essa incerteza pode constituir um obstáculo à adoção de soluções ótimas – e, por isso, as normas não precisam visar a soluções ótimas, mas a criar um ambiente estável para os agentes. Uma das formas de fazê-lo é lhes atribuindo responsabilidades.

Teóricos da autoridade frequentemente partem do ponto de vista de um juiz ou de um legislador parlamentar; no primeiro caso, apresentam uma concepção de legitimidade baseada na necessidade de arbitrar conflitos (Shapiro, 2000; Nozick, 1990), e, no segundo, numa teoria da democracia representativa (Waldron, 2009). Esses dois tipos de teoria são fundamentais para a filosofia política moderna (e conexos a algumas das considerações dos capítulos II e IV); contudo, as pessoas exercem diferentes papéis identificados como autoridades, como pais, professores e chefes, e pode-se arguir que esse conceito seria melhor elucidado atentando para

as diferentes demandas aplicadas às pessoas nesses papéis distintos (Hershovitz, 2010, 49). Nos dois exemplos paradigmáticos de autoridade jurídica, a incerteza factual (em especial, sobre o futuro) é um tema negligenciado; por isso, este trabalho frisa a atuação de autoridades regulatórias (que abrange não apenas entidades vinculadas a outros poderes do Estado, como o Executivo, mas também entidades às quais outorgamos a função de criar regras e convenções regendo atividades privadas<sup>1</sup>) – assim como o papel de entidades que atuam sem exercer um poder normativo ou coercitivo (autoridades fiscais e monetárias, encarregadas de intervenções macroeconômicas, como Bancos Centrais).

A Crise Financeira de 2008 é o “pano de fundo” desta tese: dez anos depois, a despeito do benefício da visão retrospectiva, ainda se discutem suas causas, consequências, e que medidas podem ser adotadas para prevenir situações afins. Trata-se de um evento que, mesmo marcando profundamente o começo deste século (discutido amplamente em estudos sobre macroeconomia e regulação), talvez já estivesse esquecido pelo público em geral, não fosse o filme de Adam McKay *The Big Short*. A crise trouxe (de volta) à baila a relação entre o mercado financeiro e a incerteza, o desenho de instituições resilientes e o papel de autoridades como garantes da estabilidade social e econômica; no plano internacional, levou ao terceiro acordo de Basileia e a regulações bancárias mais restritivas – incluindo a reformulação de normas de contabilidade, supervisão e resolução de instituições financeiras. Também pode ser associada ao desenvolvimento da *blockchain* (como resposta à desconfiança em relação ao sistema financeiro), uma inovação na área de criptografia, computação descentralizada e *design* de mecanismos – concebida para viabilizar o *bitcoin*, a primeira criptomoeda, que dispensa a necessidade de uma terceira parte para validar transações (Nakamoto, 2008). No limite, ensejou propostas para acabar com a atividade bancária como a conhecemos, por meio da adoção de medidas que reduziriam drasticamente (ou mesmo impediriam) a criação de moeda escritural por instituições privadas (v. Kotlikoff, 2013; Macmillan, 2014).

As últimas linhas dessa tese foram marcadas por eventos não menos imprevistos que a Crise, como o governo de Donald Trump nos EUA; foram revisadas na França, num período em que o inusitado governo Macron enfrentava contestações internas (como uma marcante greve nos transportes públicos), e em que a instabilidade na União Europeia era agravada pelas indefinições sobre o governo italiano (decorrente da vitória nas eleições legislativas de dois

---

<sup>1</sup> Como conselhos profissionais e entidades de classe em geral (as quais, mesmo que tenham suas competências definidas em lei, têm grande independência do governo e tratamento jurídico bastante peculiar), ou entidades de relevância pública às quais o direito atribui poderes e competências especiais (e.g.: ABNT, CGI.br, B&F Bovespa, FGC...), ou entidades independentes de escopo supranacional (como o IFRS).

partidos *outsiders* com agenda incompatível), pelas negociações sobre o *Brexit*, pela crise migratória e pelos desafios da união monetária (herdados da Crise Soberana do Euro). Outrossim, todo o doutorado ocorreu ao longo da profunda crise econômica e política que tem assolado o Brasil desde 2015; mas espero que nosso silêncio sobre estes eventos mais recentes seja suficientemente eloquente – seria *fácil demais* tirar daí exemplos de “riscos de cauda longa” imprevistos, de desafios à teoria da legitimidade, ou mesmo de discussões polêmicas sobre ética, sanções e responsabilidade. Ao fim e ao cabo, a filosofia tem por prática voltar-se para indagações sobre *origens* – e é pelo menos argúvel que estes eventos históricos não teriam ocorrido sem a Grande Recessão a qual, após a Crise, varreu o otimismo da economia e da política mundial<sup>2</sup>.

### Os artigos

A presente tese é composta de seis artigos independentes entre si, agrupados como capítulos em duas seções de acordo com a similaridade do tema – de forma que cada um complementa o artigo anterior e abre alas para o artigo seguinte. Considerando que foram escritos em diferentes momentos e que alguns já foram submetidos a periódicos para publicação, os textos podem por vezes não ter a coesão que consideraríamos ideal, dando a impressão de algo como um romance em cadeia (uma metáfora que emprestamos de Dworkin, 1999); contudo, eles partilham os *leitmotivs* principais (considerações sobre a autoridade a partir da teoria da decisão e da filosofia prática, levando em conta a noção de incerteza), e foram dispostos segundo uma ordem temática. A seção inicial (capítulos um a três), intitulada *Paradoxos da Razão e da Autoridade*, conecta filosofia política e jurídica, a partir de duas objeções filosóficas à teoria da autoridade jurídica de Raz (1979).

O primeiro artigo faz uma releitura de uma objeção à *teoria da autoridade*: que ou concordamos com suas diretrizes (autoridade desnecessária), ou discordamos delas – a autoridade “absurda”: não podemos segui-la sem contradição com nossas próprias crenças. Embora, à primeira vista, o problema pareça um mero jogo de palavras, analiso como, no contexto da teoria de Raz, seria afim ao paradoxo de Moore, caso uma autoridade enunciasse

---

<sup>2</sup> “5000 to 1 pales in comparison with the odds you would have got in 2008 on a future world in which Donald Trump was president, Theresa May was prime minister, Britain had voted to leave the European Union, and Jeremy Corbyn was leader of the Labour Party – which to many close observers of Labour politics is actually the least likely thing on that list. The common factor explaining all these phenomena is, I would argue, the credit crunch and, especially, the Great Recession that followed” (Lanchester, 2018).

um “ponto-cego” do tipo “você deve fazer  $\alpha$ , ainda que acredite que não deve fazer  $\alpha$ ”. Assumindo uma concepção de racionalidade limitada, sugerimos duas maneiras de interpretar e resolver tal paradoxo, do ponto de vista do sujeito: ou como um exemplo de “irracionalidade racional” (em que o sujeito escolhe submeter-se ao juízo da autoridade, sem pensar sobre como este é inconsistente com seus próprios juízos), ou – mantendo a analogia com instrumentos (cujas leituras e indicações também podem conflitar com nossos juízos *prima facie*) – tomando a ordem como evidência em favor de  $\alpha$  e buscando solucionar a inconsistência de forma “dinâmica” (i.e., ou alterando o próprio juízo *prima facie* sobre  $\alpha$ , ou o juízo sobre a legitimidade da ordem)<sup>3</sup>. Além de apresentar brevemente os principais pontos da teoria de Raz – que serve de base para as discussões sobre legitimidade subsequentes – o artigo também introduz algumas discussões sobre a própria noção de racionalidade e sobre atitudes proposicionais (como o conceito de *blindspots*), que serão aproveitadas nos capítulos subsequentes. O paradoxo do desacordo permite melhor compreender o argumento de Greenspan, discutido no quarto capítulo, do ponto de vista do sujeito: a diretriz da autoridade não poderia ser legítima quando o agente tem mais informações que ela. Contudo, tal argumento não faz jus à complexidade das decisões sociais (que impõem restrições especiais aos respectivos problemas de decisão), nem à produção de coordenação por meio de normas e autoridades.

No segundo artigo, vislumbramos como a discussão sobre legitimidade é *pervasiva*; abordamos a eterna questão de filosofia jurídica sobre a relação entre direito e moralidade, esboçando um contra-argumento à refutação, por parte de Raz, da *tese da incorporação* – a ideia de que normas jurídicas podem incorporar conceitos morais. A partir da própria concepção de serviço e da teoria da interpretação raziana, aduzimos que é possível um sistema jurídico em que o sujeito só tenha acesso ao significado da norma jurídica se interpretar a autoridade usando deliberação moral – negando, pois, uma das premissas essenciais do argumento raziano. Além de rediscutir o conceito de legitimidade, tal artigo tem por função “limpar o terreno” – prevenindo a possível objeção chã de que a análise de conceitos morais não é relevante para a

---

<sup>3</sup> Isso não é necessariamente incompatível com a tese da preempção: do ponto de vista do sujeito, a ordem efetivamente substitui suas razões primárias, *desde que a autoridade seja legítima*. Contudo, o que está em questão é precisamente o juízo sobre a legitimidade da autoridade em relação a tal ordem.

teoria ou (mais radical) para a interpretação do direito<sup>4</sup> – para as discussões de ética aplicada à política e ao direito na segunda parte da tese.

O terceiro artigo (publicado em Peres, 2017) é uma espécie de *scherzo* – um movimento curto, em tom de chiste, inserido entre os movimentos longos e sérios de uma sinfonia; no entanto, seu tema é bastante sério – uma discussão filosófica abstrata e geral sobre a relação entre o tema da incerteza e a teoria da ação. Trata-se de um diálogo entre Aquiles e Tartaruga sobre a indeterminação de juízos sobre o futuro – num estilo propositalmente modelado a partir dos precedentes de Lewis Carroll (1895), Hofstadter (Hofstadter & Dennett, 1981) e Blackburn (2010). Apesar de seu caráter literário, pareceu-me a maneira mais adequada de expressar a ideia de que a impossibilidade de prever a própria decisão antes de tomá-la é uma das principais explicações sobre a defectibilidade de nossos juízos sobre o futuro, tornando desnecessário apelar para algum *status* ontológico ou epistêmico privilegiado de futuros contingentes. A tese da indeterminação do futuro deixa uma lacuna na teoria da decisão, porquanto incapaz de explicar como formamos, cotidianamente, juízos (ou, no vocabulário da teoria da decisão, distribuições de probabilidade) sobre futuros contingentes.

A segunda parte – *Da Incerteza à Responsabilidade* – é o cerne da tese, tratando de pontos de aplicação mais imediata. Os três artigos discutem o papel e a responsabilidade de autoridades regulatórias (em especial, de Bancos Centrais) e de agentes econômicos (i.e., dos agentes do sistema financeiro), tendo por pano de fundo a Crise de 2008.

O quarto artigo foi pensado tendo em mente a situação de alta incerteza de autoridades monetárias e regulatórias antes e durante a crise de 2007-2008 – em especial, seguindo Schefczyk (2016), avaliamos o desempenho do Federal Reserve (FED) durante a gestão de Alan Greenspan quanto à “tarefa de estabilização”. Ele se baseia em diversas apresentações realizadas ao longo do doutorado (no XIX Colóquio de Filosofia da Unisinos de 2016, no Encontro de Pós-Graduação em Filosofia de 2017, e no Encontro ANPOF XVIII de 2018) e num resumo publicado em Peres (2018). Apresenta um novo desafio cético à teoria da autoridade, o “elogio da passividade” de Huemer (2012): a tese de que, uma vez que ignoramos o futuro e as consequências de nossa própria decisão (premissa cética), é melhor se abster de agir. Apesar da negligência desse ponto por teorias da razão prática, qualquer teoria da decisão deve atentar para como são formadas nossas crenças e representações sobre estados de

---

<sup>4</sup>Conquanto a separação (conceitual) entre direito e moral no nível da *teoria* do direito seja uma tese corrente entre filósofos positivistas, estes *não* defendem (até onde saiba) que de tal separação se segue que aplicadores do direito não devem observar princípios morais ao interpretar e reconhecer normas jurídicas.

coisas possíveis – e ainda, sobre como esses fatores impactam sobre a decisão. A premissa cética costuma ser usada como crítica a autoridades tecnocráticas, vindo à tonam via de regra, quando se discutem finanças e macroeconomia; contudo, quando vulgarizada, ela prova demais – se não estamos justificados em qualquer representação sobre o futuro, então também não temos fundamento para dizer que é melhor se abster.

Tal debate se volta para o que, no direito, chamamos de decisões *discricionárias* (que *criam* normas e que tem de levar em conta as consequências da decisão), e não decisões *vinculadas* – de reconhecimento/aplicação de normas pré-existentes (em que considerações sobre as consequências não são enfatizadas). Historicamente, a teoria e a prática do direito enfatizam esse segundo tipo de decisão. Numa lição sobre incerteza, foi o próprio direito brasileiro que me surpreendeu ao incluir considerações consequentialistas explícitas na Lei nº 13.655, de 25 de abril de 2018, inserindo alterações importantes sobre direito público na Lei de Introdução às Normas ao Direito Brasileiro (Decreto-Lei nº 4.657, de 4 de setembro de 1942)<sup>5</sup>; os arts. 20 a 30 desta ora obrigam as autoridades que decidirem pela ilegalidade de uma norma ou ato jurídico (ou que decidirem com base em “valores abstratos”) da Administração Pública a terem em conta as consequências da decisão e a regularem seus efeitos, tendo em vista a segurança jurídica<sup>6</sup>. Dessa forma, pode-se dizer que o tema da decisão sob ignorância ganhou relevância para aplicadores do direito público em geral.

No quinto artigo (publicado em Peres, 2016), atacamos a chamada “tese da inevitabilidade”, discutindo a atribuição de responsabilidade *moral* a agentes financeiros por prejuízos decorrentes de crises. Segue-se uma posição e um argumento semelhantes aos de

---

<sup>5</sup> Um exemplo de como juristas pensam (ou deixam de pensar) de modo distinto de profissionais que trabalham com sistemas formais: embora seja uma lei ordinária, a LINDB desempenha um papel-chave na interpretação de todas as normas jurídicas – em especial, sobre como arbitrar conflitos de normas; alterá-la pode mudar, na prática, a maneira como se aplicam normas hierarquicamente superiores (como o direito constitucional) e normas de escopo especial. Visualizando o sistema de normas como uma rede, ela seria um nó central bastante especial.

<sup>6</sup> Não tenho a pretensão de avaliar essa nova lei. Por um lado, seria preciso fazer uma investigação empírica séria (um campo ainda raro para pesquisadores em direito e filosofia) sobre sua efetividade em aumentar a segurança jurídica; por outro, as considerações que ela impõe ao julgador parecem ser exigências razoáveis (em especial, a proteção da boa-fé do administrado) de qualquer decisão que não seja inconsequente - uma situação que já se observava na prática do direito público. Contudo, chama atenção o fato de que: a) a declaração de invalidade é um mecanismo de segurança jurídica, ainda que *ex post* – garante-se a hierarquia e a integridade do sistema excluindo atos inválidos (de modo que, ao menos em tese, restringir a possibilidade de o juiz declarar invalidade implica aumentar a probabilidade de normas conflitantes); b) segurança jurídica é um valor abstrato – e, pela razão em (a) supra, controverso; c) conquanto crie uma restrição adicional às decisões, essa lei *umenta* o escopo das considerações do julgador e as possibilidades de decidir (e de errar). Pelas prescrições anteriores do Código Civil, um julgador só teria duas opções ao analisar a validade de um ato: declará-lo válido (e manter o *status quo*) ou *inválido* (e desfazer seus efeitos); claro, isso implica que a segunda opção é muito custosa, mas pouco *incerta*. Agora, da declaração de invalidade, há vários resultados possíveis, já que o julgador tem a tarefa de modular seus efeitos; saímos de um jogo com dois equilíbrios muito distintos para outro de múltiplos equilíbrios possíveis – donde o processo judicial incorporou algo análogo à teoria da barganha.

Schefczyk (2012)<sup>7</sup>, que ataca a assim chamada *exemption view* de Posner (2010). A diferença é que, ao invés de usar um argumento afim ao do *harmless torturer* (um exemplo de paradoxo indutivo: nossa moralidade do senso comum falharia em reprovar um sádico que se limitasse a causar a suas vítimas doses de sofrimento individualmente imperceptíveis), nosso alvo é uma forma de fatalismo, comum tanto aos personagens do filme *Margin Call* quanto ao General *Arjuna* do *Baghavad-Gita*, um clássico da filosofia e da religião hindu.

No último capítulo (desenvolvido a partir de uma apresentação no *I Encontro Liberdade e Igualdade de Oportunidades* da UFRGS, em dezembro de 2016), transpomos a discussão sobre responsabilidade da ética para o direito, defendendo a implementação de normas e instituições que visem a vigiar e punir agentes que ponham em risco a estabilidade financeira. Primeiro, a fim de evitar a objeção paternalista (de que não caberia ao estado disciplinar o mercado), assumimos a tarefa pouco agradável de justificar medidas como o *bail-out* – a utilização de recursos públicos (ou afins, como os fundos gerais de garantia a depositantes) para salvar instituições financeiras insolventes ou ilíquidas. Tal prática seria, a princípio, justificada pelo *argumento do incentivo*, derivado do princípio da diferença de Rawls: sem essa garantia (implícita ou explícita), os sistemas financeiros seriam mais vulneráveis a crises sistêmicas<sup>8</sup>, com consequências negativas para todos (e proporcionalmente piores para os mais pobres). Abordamos esse argumento a partir da crítica de G. Cohen de que o argumento do incentivo não passa num *teste interpessoal*: ao apresentá-lo, uma pessoa favorecida por ele soaria como fazendo algo afim a uma ameaça – um aspecto frequentemente negligenciado da baixa popularidade de medidas econômicas, mesmo quando se reconhece a solidez dos raciocínios que as fundamentam. A crítica padrão à garantia de instituições financeiras é a de que cria risco moral – incentiva as instituições a assumirem riscos que, de outra forma, não assumiriam. Nossa conclusão é que, ainda que possa ser conveniente, ou até moralmente necessário, proteger os investidores, isso deve ser compensado pelo efeito dissuasório de *sanções*.

## **Alertas**

---

<sup>7</sup> Na época em que tal capítulo foi escrito, desconhecia esse autor (fundamental, posteriormente, para o quarto capítulo).

<sup>8</sup> Assumindo, porém, uma cláusula *coeteris paribus*; o argumento pode não ser aplicável se redesenharmos o sistema financeiro de forma a reduzir seu papel como emissor privado de moeda e a não oferecer risco sistêmico.

Primeiro, a fim de prevenir conflitos de interesse, é preciso destacar que, durante o doutorado, fui servidor do Banco Central do Brasil. As opiniões expressas no texto são apenas minhas, e não podem, em hipótese alguma, ser imputadas a essa instituição.

Segundo, de acordo com um aforisma atribuído ao finado S. Hawkings, cada fórmula subtraída de um texto dobra o número de leitores (Binmore, 2005, x); esse é um dos motivos para o caráter relativamente informal dos capítulos a seguir. Um segundo é uma consideração prática: dentro dos limites da argumentação visada em cada capítulo, aumentar a formalidade implicaria consumir “recursos escassos”, como atenção e tempo. Um dos principais motivos é a própria limitação cognitiva do autor, que (dada a carência de um *background* mais robusto em matemática e lógica) se obriga a admitir que tem uma compreensão limitada e inexata da teoria da decisão racional – inclusive no que tange alguns dos artigos referenciados na bibliografia. Por esse motivo, frequentemente meu argumento acaba com uma deferência à autoridade dos autores; um exemplo importante é o artigo de Press & Dyson (2012): não seria capaz de reproduzir a prova do teorema, nem a equação que permite desenvolver o tipo de estratégia que os autores expõem.

Busquei reduzir essa lacuna parcialmente de diversas formas: discutindo com colegas com maior domínio das ciências exatas em questão, frequentando cursos de curta duração em economia e finanças, e mesmo ingressando num curso superior de tecnologia da informação; isso ficou muito aquém de ser suficiente – não apenas por ter de dividir o tempo com atividades diretamente vinculadas ao doutorado, como o estágio docência e o período que passei como estudante visitante (o “sanduíche”) na Universidade Paris I, mas também porque a profundidade do tema demanda maior talento e *expertise*. Assim como mesmo nossas sociedades democráticas demandam uma divisão de trabalho baseada na especialização (o que implica um *trade-off* com o *desideratum* de igualdade), nossos empreendimentos intelectuais comuns (inclusive a descoberta e justificação dos melhores princípios aplicáveis à organização dessa sociedade) também a demandam (o que, para um membro desse empreendimento, implica um *trade-off* com o *desideratum* de cognição plena); concluí que uma divisão mais racional do trabalho cognitivo em questão me atribuiria o papel de pesquisar (ao invés de *produzir*) contribuições teóricas e formais mais profundas e buscar aplicá-las a discussões filosóficas.



## **Capítulo 1: Do desacordo ao paradoxo da autoridade - uma análise da concepção de serviço de autoridade à luz da teoria do “ponto-cego”, de R. Sorensen**

**RESUMO:** Estuda-se um desafio do anarquismo filosófico à teoria da autoridade jurídica de J. Raz: seguir uma ordem de que se discorda seria irracional, pois equivaleria a agir contra o que se considera mais justificado. Utilizando referências atinentes à teoria da decisão e à epistemologia, além de exemplos sobre ferramentas de auxílio à tomada de decisão do cotidiano (como *experts* e programas), esboçamos duas possíveis respostas, que correspondem a duas concepções instrumentais distintas de autoridade: numa, é justificável para o agente colocar-se numa situação que o leve a agir de forma inconsistente com suas crenças – a ideia de “submeter-se à autoridade”, afim a uma “confiança cega”; noutra, utilizando a teoria dos “pontos-cegos”, delineiam-se restrições que uma teoria da racionalidade impõe à teoria da autoridade.

**Palavras-chave:** autoridade, racionalidade, paradoxo.

**ABSTRACT:** We study a challenge from philosophical anarchism to J. Raz's theory of legal authority: it would be irrational to follow an order with which one disagrees, since it would mean acting against what is considered more justified. Through references from decision theory and epistemology, and deploying examples about tools for assisting in routine decision-making, we sketch two possible answers: first, it may be justifiable to put yourself in a situation that leads to irrationally acting; second, according to the “blindspots” theory, we outline some restrictions from a theory of rationality to the authority.

**Keywords:** authority, rationality, paradox

**Sumário:** 1. Introdução. 2. O caráter paradoxal da autoridade. 2.1. A independência de conteúdo e a “concepção de serviço”. 2.2. Novos paradoxos. 3. Paradoxos e pontos-cegos. 3.1. Racionalidade e justificação. 3.2. Pontos-cegos. 3.3 Pontos-cegos da razão prática. 4. Dissolvendo o paradoxo. 5. Conclusões

## 1 Introdução

Uma teoria da autoridade política precisa dar conta da conexão entre a questão da obrigação política, a da legitimidade da autoridade e o problema da lei injusta. Em outras palavras, ela precisa responder por que razão estamos *justificados* em obedecer a entidades como governos e leis, mesmo quando consideramos que as normas reconhecidas por essas entidades não são justas. Afinal, se a autoridade deve ser capaz de fazer uma *diferença prática* na deliberação do sujeito, ele deve estar obrigado a segui-la mesmo quando ela esteja errada, ou ele não concorde com o teor de suas diretivas – no vocabulário corrente de filosofia política, deve segui-las independentemente de seu *conteúdo*.

O anarquismo<sup>9</sup> fornece uma resposta cética a esse problema: não existe obrigação política – nem, portanto, o problema da lei injusta. Uma das principais inspirações filosóficas dessa teoria são as situações paradoxais a que somos levados pela noção de independência de conteúdo – a exemplo do que Shapiro chama *paradoxo da racionalidade*: “quando as autoridades estão erradas, elas não podem ter o poder de obrigar os outros – quando estão certas, o seu poder de obrigar não tem sentido” (Shapiro, 2000, 2-3, trad. livre)<sup>10</sup>.

Na seção a seguir, estudaremos uma das principais teorias desenvolvidas para responder a esse paradoxo – a “concepção de serviço” (*service conception*) de J. Raz (1979). Após, apresentaremos uma reformulação do problema, para a qual a teoria *raziana* não fornece uma resposta clara. Sugerimos a possibilidade de uma “concessão mínima” ao paradoxo – que um sujeito pode ser vítima de uma forma de irracionalidade ao obedecer a uma ordem da qual discorda, sem que isso implique que a própria aceitação da autoridade é irracional.

Essa solução, contudo, não faz jus às aspirações da concepção de serviço. Para elucidar melhor o paradoxo, vamos compará-lo ao que Roy Sorensen (1988) chama de “pontos-cegos”

---

<sup>9</sup> Para os de índole pragmática, que consideram pouco relevante ou urgente responder aos desafios do ceticismo filosófico, lembro que o anarquista se distingue de suas contrapartes epistemológicas (como os filósofos que põem em dúvida a realidade do movimento, do tempo, da percepção, de outras mentes, etc.) na medida em que sua teoria – para usar a pecha que Dworkin (2003, 292) imputa ao que ele chama “ceticismo externo” – não “deixa tudo como está”: ela serve de fundamento geral para o libertarianismo político e tem implicações práticas genuínas.

<sup>10</sup> “... when authorities are wrong, they cannot have the power to obligate others — when they are right, their power to obligate is meaningless”.

(*blindspots*)– proposições afins a sentenças *mooreanas* (do tipo “*p*, mas não acredito que *p*”), como, e.g., “é racional fazer o que não considero racional fazer”. Caso a comparação seja correta, a conclusão do paradoxo é não uma proposição absurda, mas uma sentença em relação à qual o sujeito racional está impedido de adotar certas atitudes proposicionais. Esclarecemos a teoria de Sorensen na terceira seção.

Por fim, encerramos com uma análise mais detalhada sobre o que está envolvido quando um sujeito discorda de uma ordem da autoridade aceita. Conquanto observemos que o paradoxo da autoridade seja muito mais inofensivo do que a maneira como a literatura filosófica costuma apresentá-lo, sua discussão revela um limite para a aceitação racional de ordens.

## 2 O caráter paradoxal da autoridade

### 2.1 A independência de conteúdo e a “concepção de serviço”

Pode-se começar o estudo filosófico da autoridade com o reconhecimento de um paradoxo: “[...] a razão requer que o sujeito sempre deve agir sob a ponderação das razões de que está ciente. É da natureza da autoridade que esta demanda submissão mesmo quando o sujeito pensa que o que é demandado é contra a razão. Portanto, a submissão à autoridade é irracional” (Raz, 1979, p. 3, trad. livre).

Para responder a esse desafio, Raz desenvolve sua concepção de serviço da autoridade, segundo a qual ordens válidas são razões protegidas: elas são ao mesmo tempo razões para realizar uma ação (razões de primeira ordem) e razões para excluir a deliberação sobre outras razões (razões excludentes, de segunda ordem). Para ser dotada de autoridade, segundo Raz, uma ordem deve mediar entre pessoas e razões; ela deve ser apta a guiar a conduta do agente, sem que ele precise deliberar sobre as razões que justificariam a ação – para isso, ela substitui as razões que o agente já tinha. Essa é a *tese da preempção*: “o fato de que uma autoridade requer a realização e uma ação é uma razão para sua realização, a qual não deve ser adicionada às outras razões relevantes quando se pensa sobre o que fazer, mas deveria substituir algumas delas” (Raz, 1994, 214, trad. livre). Mas, para efetuar tal substituição efetivamente, a norma deve ser baseada nas razões que o sujeito já tem – i.e., sua justificação *depende* das razões para agir de primeira ordem do sujeito, ou a autoridade não seria distinta de simples coerção (ibid.). Essa tese da *dependência* implica a da *preempção*: uma vez que a autoridade já teria

considerado algumas das razões aplicáveis ao sujeito, deliberar sobre estas junto com o comando da autoridade seria "contá-las" duas vezes (Raz, 1986, 58)<sup>11</sup>.

A concepção de serviço tornou-se a teoria da autoridade mais difundida entre filósofos do direito – ao menos no ambiente acadêmico anglo-saxão (Hershovitz, 2010). A noção de preempção explica os requisitos de “peremptoriedade” e de “independência de conteúdo” usualmente associados aos comandos de uma autoridade. Junto com a tese da dependência, as duas teses favorecem uma teoria da legitimidade, centrada na ‘tese da justificação normal’ – segundo a qual uma autoridade é legítima quando o indivíduo a ela sujeito tem maior probabilidade de atender às razões que se aplicam a ele aceitando as diretrizes dela e tentando segui-las do que se buscar seguir essas razões diretamente (Raz, 1986, 53). Essa concepção permitiria dissolver o paradoxo: embora a razão exija que o agente aja de acordo com as razões de que ele está ciente, isso não significa que ela exija que ele aja de acordo com todas essas razões; uma ordem legítima substitui as razões de primeira ordem. Ela também permite fundamentar a tese de que toda autoridade *de facto* reivindica (podemos também dizer que *pretende*) ter autoridade *de jure* – i.e., autoridade legítima (Raz, 1986, 26-7).

Se a autoridade é legítima, isso significa que o sujeito tem obrigação de segui-la, ainda que julgue de modo contrário (Raz, 1986, 62). É importante diferenciar isto da conclusão de, e.g., Hart (1994, 253) de que a autoridade, ao menos em algum ponto, pretende “cortar” a deliberação sobre os prós e contras da ação – o que Raz chama de “submissão do juízo” (*surrender one’s judgement*). Como os comandos da autoridade prática são razões para agir, seria irrelevante (para caracterizar alguém como autoridade) se o sujeito julga de acordo com eles ou se se abstém de fazer um julgamento (Raz, 1986, 39). Em conjunto com a tese da *preempção*, porém, isso parece implicar que o sujeito que aceita uma autoridade legítima e que não se abstém de fazer esse juízo está agindo de forma irracional – pois estaria “reabrindo” uma deliberação encerrada.

---

<sup>11</sup> Nesse ponto, discordamos de Raz: qual o problema de contar duas vezes? Se eu obedecer à autoridade, a dupla contagem não prejudica nossa análise (na verdade, essa parece ser a situação do sujeito que segue a regra, a despeito de seu próprio julgamento – e que Raz considera normal); por outro lado, se, apesar da dupla contagem, eu desobedeço à diretiva, isso significa que as razões em contrário eram muito significativas, pois superaram as razões (superavaliadas) para agir de acordo com a ordem. O ponto é que são justamente as razões em contrário que o comando busca excluir: “These exclusionary reasons do not, of course, exclude relying on reasons for behaving in the same way as the directives requires. [...] Hence the pre-emption excludes only reasons that conflict with the authority’s directive” (Raz, 2009, 144). Logo, o argumento da dupla contagem é uma petição de princípio: é a tese da preempção que torna a dupla contagem desnecessária (porque a preempção exclui as razões contrárias à ordem), e não a irracionalidade da dupla contagem que implica a preempção.

Por outro lado, se o comando da autoridade legítima é, além de uma razão prática, uma razão epistêmica (e parece ser esse um dos pontos cruciais do caráter instrumental da concepção de serviço), então o sujeito age de forma irracional ao discordar dela – de forma semelhante a um investigador que ignora evidência relevante. Não posso beneficiar-me do conhecimento especializado de meu corretor se não aceitar seus conselhos de investimentos, assim como não posso beneficiar-me dum programa que efetue a mesma tarefa. É claro, há situações em que não devo aceitá-los – e.g., o instrumento pode deixar de ser confiável (se meu corretor tem conflito de interesses, ou se o programa apresenta *bugs*...). Essas são as situações que Raz (1986, 62) chama de erros *claros* (e.g., se o resultado da soma de números inteiros for apresentado com uma fração decimal), em contraposição a um erro grande (se o mesmo resultado for um erro de milhares de unidades), porém difícil de detectar; é compatível com a concepção de serviço que ela seja condicionada à ausência de erros claros.

Isso aparentemente inverte o paradoxo: sendo  $\alpha$  uma ordem legítima, não deveríamos perguntar “como pode ser racional fazer  $\alpha$ , se não o considero racional”, mas “como posso não considerar racional fazer  $\alpha$ ?”. Submeter-se à autoridade, na concepção de serviço, não equivale a abdicar de julgar racionalmente, mas a fazê-lo de acordo com a evidência mais relevante.

## 2.2 *Novos paradoxos*

Na Introdução, citou-se o paradoxo da racionalidade, que questionava como alguém pode submeter-se a uma autoridade se, quando discordar dela, a submissão é injustificada, e, quando concordar com ela, a submissão é desnecessária. Essa formulação sugere uma analogia com o que podemos chamar de “falácia do estudante”: “se eu estudar e não passar no teste, é porque sou ignorante e o estudo é inútil; se passar, é porque sou inteligente, e o estudo é desnecessário”. A falácia é rapidamente exposta porque é evidente que o estudo aumenta a probabilidade de passar no exame; ele tem um efeito epistêmico benéfico, na medida em que, no momento do teste, permite fazer um julgamento melhor.

Algo semelhante valeria para a autoridade, segundo a concepção de serviço: de acordo com a tese da justificação normal, o sujeito que obedece a autoridade legítima tem maior probabilidade de agir racionalmente. Mas tentemos reformular o paradoxo de uma forma (talvez) ainda mais primitiva. Suponha que F (uma autoridade) ordene a R que faça  $\alpha$ ; este, porém, *acredita* que não deve fazer  $\alpha$ . Sabendo disso, F expõe a teoria da autoridade *raziana*,

que lhe permite enunciar a R<sup>12</sup>: (i) “você deve fazer  $\alpha$ , ainda que acredite que não deve fazer  $\alpha$ ” e (ii) “você deve fazer  $\alpha$ , mesmo que eu esteja errado”. R, para concordar com (i), deveria concluir que (iii) “devo fazer  $\alpha$ , ainda que acredite que não devo fazer  $\alpha$ ”. O problema aqui surge da manutenção de duas *attitudes* incompatíveis (aceitar uma ordem e discordar dela); embora nenhuma dessas sentenças seja contraditória, a proposição (ii) não deve ser enunciada por F, e R não pode crer em (iii).

### 3 Paradoxos e pontos-cegos

#### 3.1 Racionalidade e justificação

O problema em questão refere-se à relação entre a formação de juízos justificados e ações. Uma possível solução para o paradoxo seria o que D. Parfit (1984, 12) chama de “irracionalidade racional”: situações em que, paradoxalmente, um agente racional, visando ao melhor estado de coisas para si, preferiria agir de maneira inconsistente com suas crenças ou com a própria racionalidade. O objetivo de Parfit é estudar a tensão entre a concepção de racionalidade como auto-interesse e o cumprimento de promessas e ameaças por um indivíduo racional.

Por exemplo, imagine que eu ordene que você me preste um serviço, sob a ameaça de explodir uma bomba que nos matará a ambos (uma estratégia de “destruição mútua garantida”); supondo que sou racional, você decide ignorar a ordem – afinal, não ganharei nada (e perderei muito) cumprindo a ameaça. Mas, nesse caso, *eu sequer poderia ter enunciado a ameaça* – pois deveria prever que você anteciparia a sua incredibilidade (ela seria “conversa fiada” – *cheap talk*). É, pois, racional para mim tornar-me (unilateralmente) um irracional “cumpridor de ameaças”, formando a intenção imutável de as cumprir; enquanto você se beneficiaria tornando-se um irracional “ignorador de ameaças”<sup>13</sup> (Parfit, 1981, 353 – Apêndice A).

---

<sup>12</sup> Apelar para o conteúdo ilocutório da ordem não resolve o problema. Se F enunciar o imperativo “faça  $\alpha$ ” com qualquer uma das sentenças supra estaremos no mesmo paradoxo, na medida em que a autoridade reivindica legitimidade.

<sup>13</sup> No extremo, temos a “resposta de Schelling ao assalto à mão armada” (*Schelling's Answer to Armed Robbery*), que demonstraria que qualquer teoria da racionalidade deve aceitar que, em algumas situações, pode ser racional causar a si mesmo irracionalidade, ainda que temporariamente (Parfit, 1984, 11); trata-se de uma referência a algumas das estratégias para vencer situações de barganha estudadas por Thomas Schelling (2015) no clássico *Strategy of Conflict*. A ideia é que frequentemente pode ser uma vantagem limitar as próprias opções. Suponha que um assaltante invadisse minha casa, ouvisse-me chamar a polícia (a qual demora alguns minutos para chegar) e, por fim, exigisse o ouro de meu cofre, sob ameaça de morte; o problema é que não tenho garantia alguma de que ele não me matará, já que posso identificá-lo – aparentemente, isso seria o mais racional (cf. o jogo *kidnap* em Binmore, 2007, 44). Nesse caso, seria racional tomar uma “poção” que me deixasse temporariamente irracional,

Isso pode lembrar situações polêmicas entre os estudiosos da teoria da decisão causal e os da teoria dos jogos, como o paradoxo da toxina de G. Kavka (1983, 33-4), o “dilema do prisioneiro” iterado um número determinado de vezes, ou situações afins ao jogo do *ultimato*, em que as estratégias adotadas por jogadores em testes de laboratório (e recomendadas pelo senso comum, na medida em que maximizam a utilidade esperada) contrastam com “equilíbrios perfeitos sub-jogo”<sup>14</sup>. Se um agente só é racional quando age da maneira que ele acredita que possa causar a consequência desejada, é concebível que, em algumas situações estratégicas ou nas quais haja “falhas de racionalidade”, pode ser melhor para o agente *não agir de forma racional*.

Um exemplo pode ser a teoria *causal* da autoridade esboçada por Shapiro (2000, 53), voltada a situações de racionalidade limitada – como seria o caso de alguém que contratasse um *personal trainer* rigoroso para obrigá-lo a exercitar-se mesmo quando, por preguiça, preferisse não o fazer. Submeter-se a outrem para evitar *akrasia* seria análogo a instituir uma

---

sob cujo efeito eu adotaria crenças e atitudes manifestamente inconsistentes, e agiria visando a consequências indesejadas – como dizer “por favor, torture-me; odeio dor”. Segundo Parfit (1984, 11) "In the case just described, while this man is still in my house, it would be irrational for me to cause myself to cease to be very irrational. [...] During this period, my acts are very irrational. [...] We should claim instead that, since it was rational for me to cause myself to be like this, this is a case of rational irrationality”.

O exemplo de Parfit pode parecer assaz estranho; mas isso não é tão relevante. Se não estiver convencido, suponha que um sujeito “confiável” ameace matar minha família a menos que eu tome a poção. O cenário é basicamente o mesmo: tomar a poção é uma decisão racional; as ações cometidas sob o efeito da poção, não. Afinal, se pode ser racional causar a si mesmo um estado de privação temporária de qualquer capacidade cognitiva (é isso o que analgésicos e anestésicos fazem), por que não o poderia ser privar-se da própria racionalidade?

<sup>14</sup> Um dilema do prisioneiro é um jogo simétrico entre dois jogadores ( $x$  e  $y$ ), que devem escolher (unilateralmente) uma das seguintes estratégias: cooperar (*dove*) ou não (*hawk*). O jogador  $x$  prefere: a) que  $x$  não coopere e que  $y$  coopere a b) que ambos cooperem, e isto a c) que ninguém coopere, e isto a d) que  $y$  coopere e  $x$  não coopere. Um *nash-equilibrium* é a situação em que a estratégia escolhida por cada jogador é a melhor resposta à estratégia escolhida por outro; já uma estratégia dominante é a que maximiza o *pay-off* do agente, não importa o que o outro faça (Binmore, 2007, 20). No dilema do prisioneiro, só há um *nash-equilibrium*, em que ambos *não cooperam*. O que há de paradoxal é que isso implica uma situação *pior* do que se ambos cooperassem.

Não devemos, porém, confundir esse jogo com sua versão iterada indefinidamente – na qual a cooperação mútua é um *nash-equilibrium*: se  $x$  não cooperar num *round*,  $y$  o “punirá” não cooperando no *round* seguinte (Binmore, 2007, 21). O problema é quando o jogo tem um número *definido* de *rounds*: nesse caso, a indução retrospectiva (*backward induction*) mostra que os jogadores não deveriam cooperar. Então, jogar *hawk* em todos os casos é uma estratégia dominante, e a situação em que ambos sempre jogam *hawk* é um “equilíbrio perfeito sub-jogo”. Prova: suponha que o jogo tem 2 *rounds*; nesse caso, os jogadores sabem que não cooperarão no 2o (porque se reduz ao problema original). Mas isso implica que eles não têm por que cooperar no 1o *round*. Mas se isso vale para dois *rounds*, também vale para três, e assim sucessivamente. Mas imagine que o jogo tem 100 *rounds*: a indução retrospectiva diz que os jogadores vão adotar *hawk* em todos os *rounds*, embora uma estratégia que misture *dove* e *hawk* de acordo com a disposição dos demais de cooperarem retornaria um *pay-off* melhor (Binmore, 1997).

Nesse caso, poderíamos sentir-nos tentados a concluir que se  $x$  e  $y$  cooperassem no primeiro *round*, eles estariam adotando uma decisão irracional do ponto de vista do jogador em *cadaround*, mas acertada quando se considera o jogo de forma geral. Mas, se essa possibilidade não pode ser excluída previamente, então um jogador não sabe se o outro agirá de forma racional e o equilíbrio sub-jogo não estará acessível – donde não temos motivo para chamar de irracional a decisão de cooperar no primeiro *round*.

regra para evitar tomar decisões em contextos nos quais a probabilidade de erro é grande, ou mesmo aceitar um código de conduta a fim de tornar-me confiável e entrar num grupo<sup>15</sup>:

Quando alguém se compromete a seguir uma autoridade para beneficiar-se de suas diretivas, essa pessoa restringe seu “eu” futuro de forma a agir segundo as ordens da autoridade, quaisquer que sejam. O sujeito não escolhe obedecer quando a ordem é emitida – quando ordenado, ele não tem escolha, exceto obedecer<sup>16</sup>.

Nestes casos, o ato de submeter-se é racional (da mesma forma que, nos exemplos de Parfit, tomar a poção é uma ação racional, ou fazer uma ameaça autodestrutiva é racional), mas isso não significa que cada ação subsequente o seja. Aplicando o mesmo raciocínio à teoria da autoridade, a preempção se aplicaria, pois, não às *razões para decidir*, mas à própria decisão, uma vez que é o próprio ato de decidir (aqui considerado como deliberação e escolha) que é substituído pelo comando da autoridade. Se isso estiver correto, podemos dizer que R age racionalmente ao submeter-se ao conjunto de ordens de F (podemos dizer que esse conjunto é um “*global optimum* racional”), ainda que isso não implique que ele aja racionalmente ao submeter-se a cada ordem (pois esta não é compatível com outras de suas atitudes proposicionais). Pode haver uma ordem em relação à qual não é racional submeter-se – mas se R investigar, para uma determinada ordem, se é esse o caso, deixará de submeter-se ao conjunto de ordens de F.

Embora isso forneça uma solução para os paradoxos de F e R, na medida em que aceita um dos “chifres do dilema” – i.e., que R está agindo com uma forma de inconsistência no *caso particular dea* (mas não, de forma geral, por aceitar F como autoridade) – ela é difícil de extrapolar para todos os casos, pois frequentemente pensamos em regras e autoridades como algo semelhante a *instrumentos* para tomar decisões melhores, como parece ser a proposta da concepção de serviço. Além disso, essa solução ignora um aspecto fundamental de nosso paradoxo: no exemplo de F e R, não é a obediência à ordem em si que é irracional, mas a enunciação da conjunção das proposições que expressam que F deve ser obedecido e que a ação ordenada não é devida. Veremos uma alternativa que faz jus a esses pontos.

---

<sup>15</sup> Interessante notar que o argumento do estado de natureza de Hobbes (2002, 94-97) pode ser visto dessa forma: cada indivíduo renunciaria a seu direito a tudo, aceitando um soberano, a fim de deixar de ser uma ameaça para os demais. Isso não é coincidência, já que Parfit apresenta a noção de “irracionalidade racional” ao discutir a teoria da racionalidade como auto-interesse – a qual teria Hobbes como um dos principais expoentes.

<sup>16</sup> When a subject commits himself to following an authority in order to benefit from its directives, that person constrains his future self to act on the demands of the authority, whatever they may be. The subject does not choose to obey when the directive is issued – when ordered, the subject has no choice but to obey” (Shapiro, *ibid.*, trad. livre)



### 3.2 Pontos-cegos

Todos os vertebrados têm “pontos cegos” no campo de visão, causados pela passagem do nervo ótico pela retina; esses pontos são “preenchidos” na formação da imagem pela paisagem (é o *scotoma* do campo de visão). De forma análoga, há sentenças consistentes em relação às quais não podemos ter certas atitudes proposicionais – incluindo crenças e conhecimento. Por isso, essas sentenças foram batizadas de “pontos-cegos” (*blindspots*) por Roy Sorensen (1988).

Pontos-cegos decorrem do caráter reflexivo do auto-conhecimento – i.e., do fato de que temos crenças (ou outras atitudes proposicionais) que têm por referência outras crenças ou outras atitudes proposicionais. O exemplo clássico seriam as *sentenças mooreanas* – assim chamadas em homenagem a George Moore e a sua descoberta da sentença paradoxal “chove, mas não acredito nisso”. Ela não é uma contradição: posso perfeitamente saber que “chove, mas  $x$  não acredita nisso”, desde que  $x$  não seja substituído por “o autor deste artigo” (i.e., eu). Contudo, quando um sujeito não tem opinião sobre o valor de verdade de determinada proposição, ele pode crer numa disjunção verdadeira de “pontos-cegos”; e.g., sei que ou “o número de estrelas é par, e não acredito nisso” é verdadeiro, ou “o número de estrelas é ímpar, e não acredito nisso” é verdadeiro: embora não possa crer em nenhum dos disjuntos especificamente, devo crer na disjunção.

Dessa forma, a noção de “ponto-cego” permite explicar como seres humanos normais podem ter *crenças falibilistas*: se eu fosse um lógico perfeito, deveria acreditar na conjunção de todas as minhas crenças (só assim poderia dizer que meu sistema de crenças é *completo*) – i.e., deveria acreditar que *todas* as minhas crenças são verdadeiras; mas isso contradiz a crença falibilista e modesta de que tenho ao menos uma crença falsa<sup>17</sup>. Se  $\beta$  representa o operador “creio que”, e as letras minúsculas, proposições em que creio, então o conteúdo da “crença modesta” é a seguinte disjunção de pontos-cegos: “( $p$  é falso &  $\beta p$ ), ou ( $q$  é falso &  $\beta q$ ), ou ( $r$  é falso &  $\beta r$ )...”. Como já observamos, porém, não posso crer em nenhum dos disjuntos. Em bom português, podemos dizer que, embora eu creia que *cada uma* de minhas crenças é verdadeira, isso não significa que creia que *todas* (i.e., seu conjunto) sejam verdadeiras.

---

<sup>17</sup> Se o escopo dessa crença (chame-a  $\beta\infty$ ) *étodo* o meu conjunto de crenças (permitindo, pois, auto-referência), então podemos ter um “paradoxo do mentiroso”: ou há ao menos uma sentença do conjunto complementar de crenças que é falsa, ou  $\beta\infty$  o é; mas, nesse caso, ela é verdadeira, e assim *ad infinitum*.

### 3.3 Pontos-cegos da razão prática

Um caso de ponto-cego de relevância prática é a “prova da auto-aniquilação” do utilitarismo, por Bernard Williams (2005, 161 e s.), que ora resumimos:

a) “um utilitarista deve buscar a atitude menos danosa possível que seja capaz de evitar consequências piores” (Williams, 2005, 161);

b) “assim, ele será frequentemente impelido a fazer coisas que, consideradas em si mesmas, são francamente abjetas” (como o que Williams chama de “intervenções preventivas”);

c) se um número muito grande de pessoas for utilitarista, será empiricamente provável uma escalada de “intervenções preventivas” (até porque nem todas as pessoas são utilitaristas; do contrário, viveríamos numa utopia, onde o pensamento moral seria em boa medida desnecessário);

d) esse cenário é pior do que se ninguém for utilitarista.

Conclusão de Williams: “se o utilitarismo estiver certo, e com ele algumas proposições empíricas bastante prováveis, será melhor que as pessoas não creiam no utilitarismo. Se, por outro lado, ele for falso, então certamente será melhor que as pessoas não creiam nele” (Williams, 2005, 165).

De forma análoga, Sorensen (1988, 382) menciona que um utilitarista pode ter dificuldades auto-referenciais ao aprender que um juízo falso sobre uma situação *S* pode levar a melhores consequências; isso proscreveria julgamentos racionais a respeito de *S*. O exemplo de Williams apenas aplica a mesma lógica ao próprio utilitarismo; essa teoria implicaria, nesses casos, uma tese do tipo: “devo agir de modo a maximizar a utilidade geral, mas não posso acreditar nisso (pois implica agir contrariamente à maximização da utilidade geral)”. Essa “fragilidade à auto-referência” não é exclusiva do utilitarismo, mas se aplica a qualquer teoria da decisão que pretenda aplicar-se à *própria atividade de tomar decisões*, já que implica a possibilidade de “*x* deve fazer  $\alpha$ , mas não deve pensar nisso”.

Algo semelhante pode ocorrer com teorias da justificação da decisão jurídica. Seria esse o caso da “pretensão de correção” de R. Alexy, bem como da reivindicação de legitimidade

de Joseph Raz<sup>18</sup> por parte da autoridade jurídica; embora toda autoridade seja falível e esteja em condições de reconhecê-lo<sup>19</sup>, ela não pode dizer, ao enunciar certa decisão, que *esta* decisão é incorreta. Uma “autoridade falibilista” está numa situação análoga à da pessoa que admite que tem uma crença falsa – mas que é incapaz de apontá-la.

O que não significa que não possamos fazê-lo por ela: posso enunciar um de seus pontos-cegos (“*p*, mas você não acredita que *p*”), mas *você* não pode acreditar na proposição enunciada (ao menos não sem a tornar falsa). Sorensen (1988, 161) considera que isso explica a discórdia mesmo em situações idealizadas; certamente, isso implica a ausência de conhecimento comum (Sorensen, 1988, 320) no que for relacionado ao ponto-cego – uma vez que não posso ter conhecimento de algo que é um ponto-cego para mim<sup>20</sup>.

#### 4 Dissolvendo o paradoxo

Retomando o paradoxo inicial: F não pode enunciar “você deve fazer  $\alpha$ , ainda que eu esteja errado” porque esse é um “ponto-cego da autoridade legítima”: sua pretensão de refletir corretamente as razões do sujeito é incompatível com a admissão conjunta de erro. Isso não proscreve a modéstia ao “estilo Ayres Britto”: F pode admitir que existe *alguma decisão sua x* que é errada, mas que, apesar disso, deve ser seguida. Há uma diferença lógico-formal entre os dois casos: enquanto “alguma decisão *x*” aqui é uma variável ligada por um quantificador existencial (“existe um *x* qualquer tal que *x* é uma decisão e...”), “ $\alpha$ ” denota, em nosso exemplo, o que seria uma proposição determinada (i.e., opera substituindo o ‘nome’ de um objeto – trata-se *desta* decisão); i.e., quando dizemos “alguma decisão *x*”, não precisamos *apontá-la, discriminá-la*. A admissão de F de que há alguma decisão *x* que não é correta é análoga ao caso da “crença falibilista”: equivale a uma disjunção de pontos-cegos, do tipo “ou a decisão 1 é errada, e tenho a pretensão de que não seja errada; ou a decisão 2 é errada, e tenho a pretensão de que não seja errada...”.

F também deve abster-se de enunciar “você deve me obedecer, ainda que não acredite nisso”, mas por um motivo diferente (conquanto conexo): ele pretende que sua ordem seja uma razão para R, e essa é uma proposição que *R não pode aceitar*. O problema maior está em R: é

---

<sup>18</sup> Para uma comparação entre as duas pretensões, ver Gardner (2012, 125).

<sup>19</sup> E.g., a afirmação de Carlos Ayres Britto (2001, 10), então Ministro do Supremo Tribunal Federal, de que essa corte tem o direito de errar por último.

<sup>20</sup> De acordo com o Teorema do Acordo de Robert J. Aumann (1976), dois agentes racionais (mais precisamente, *bayesianos*) com conhecimento comum de suas crenças iniciais (seus *priors*) e conhecimento comum das respectivas implicações não podem discordar em suas conclusões (seus *posteriors*).

mais fácil abster-se de produzir um enunciado que uma crença; é por isso que os paradoxos da autoridade costumam ser enunciados do ponto de vista de R, o “súdito”. Se este acredita que F é uma autoridade legítima, mas discorda de F, pode não ser tão simples abster-se de formar uma crença que ele já tem; vale a pena analisar cada uma das possibilidades lógicas em que isso pode ocorrer:

1. F é uma autoridade legítima, e  $\alpha$  é devido: se a concepção de serviço estiver correta, R está sendo irracional em discordar. Ele não está desobedecendo à autoridade (i.e., não viola uma obrigação), mas seria racional (a fim de manter a consistência entre suas crenças e atitudes) rever sua segunda crença (de que  $\alpha$  não é devido).

2. F é uma autoridade legítima, mas  $\alpha$  não é devido: podemos dizer que F está dentro de sua “margem de erro” usual. Nesse caso, F não deixou de refletir as razões de primeira ordem de R. Ainda assim, R está sendo, para tomar de empréstimo outra expressão de Parfit (2011, 34), “menos que racional”<sup>21</sup> ao discordar. Embora suas duas crenças sejam verdadeiras (na legitimidade de F e no erro de  $\alpha$ ), ele está mais justificado em manter a primeira – donde ele “deve racionalmente” (parafrazeando Parfit, *ibid.*) segui-la. É possível discutir se F cria uma obrigação nesse caso; se seu comando realmente substitui as razões de primeira ordem (i.e., se faz “diferença prática”), mas isso é irrelevante para o paradoxo (e para a perspectiva de R)<sup>22</sup>.

Em certo sentido, podemos dizer que a autoridade não comete um erro nesse caso – assim como R não erra ao segui-la (i.e., ele age de acordo com a razão). Se é possível chegar a conclusões teóricas falsas mesmo depois de reunir toda a informação disponível e raciocinar da

---

<sup>21</sup> É interessante distinguir duas noções: a) “R tem razões para fazer  $\alpha$ ”, de b) “R deve racionalmente fazer  $\alpha$ ”. Suponha que R se depara com uma perigosa serpente no deserto; ele acredita que o melhor a fazer é sair correndo – e, dada essa crença, isso é o que ele *deve, racionalmente, fazer*. Infelizmente, R ignora que as serpentes dessa espécie são muito rápidas, mas não atacam indivíduos que ficam parados; nesse caso, embora ele tenha uma razão para não correr, R a desconhece – donde ele deve, racionalmente, correr (v. Parfit, 2011, 34 e s.).

<sup>22</sup> É irrelevante porque o paradoxo não deveria ocorrer: não estou justificado em pensar que a ordem é errada. Apenas sugiro que R pode não ter obrigação porque, se ele não cumprisse a determinação de F, é discutível até que ponto seria passível de crítica – se, além de ter falhado em fazer o que racionalmente deveria fazer, ele teria violado uma obrigação. Suponhamos que meu general me ordene prestar apoio a um ataque – o qual, não sabemos, seria um desastre. Por minha preguiça, intuição, ou pacifismo, meu regimento se atrasa, o que (depois descobrimos) nos poupa de um vexame, corte marcial ou morte. Posso ser passível de alguma forma de repreensão, mas é discutível se violei uma obrigação para com meu general, já que ele estaria justificado em agradecer-me.

Na realidade, uma das críticas que se faz à concepção de serviço é que mostrar que um sujeito tem uma *razão* especial para seguir outrem não é suficiente para mostrar que tem uma *obrigação* de fazê-lo (Hershovitz, 2010). Seria necessário mostrar que ele tem uma *relação* especial com quem emite a ordem. Se a relação entre comandante-subordinado exige obediência cega, então violei essa obrigação; mas pode haver outras relações de autoridade/subordinação que não sejam assim.

maneira correta (isso é, sem cometer nada que possa ser considerado um erro), também o é para conclusões práticas; nesse caso, não há falha no próprio ato de julgar. Enquanto instrumento, a autoridade acerta ao desempenhar adequadamente sua função: o sujeito age melhor ao aceitar seu comando do que ao ignorá-lo.

3. F não é uma autoridade legítima,  $\alpha$  não é devido, mas R está *epistemicamente* mais justificado em crer que F é uma autoridade legítima: um exemplo comum disso é quando a informação é limitada; nesse caso, a probabilidade *a priori* de R respeitar suas razões de primeira ordem é maior se obedecer ao comando – e, na ausência de informação relevante (que pode não estar acessível a R), ele não pode extrair a respectiva probabilidade *a posteriori*. Esse caso é semelhante ao de (2), e R está sendo “menos que racional”, mas com a diferença de que, agora, é mais difícil sustentar que F faça “diferença prática” – já que sua ilegitimidade implica que a preempção das razões de primeira ordem *não ocorre*.

4. F não é uma autoridade legítima,  $\alpha$  não é devido, e R está *epistemicamente* mais justificado em crer que F não é uma autoridade legítima: nesse caso, a Tese da Justificação Normal de Raz não vale, e a autoridade não é legítima. Ela é como um instrumento que não é útil – ainda que isso possa ocorrer apenas excepcionalmente (e.g., quando tenho informação que a autoridade não tem, de forma que a probabilidade *a posteriori* de que eu esteja certo, e ela errada, é maior)<sup>23</sup>.

Cabe discutir um ponto levantado por (3) e (4), ao qual se poderia objetar, inspirado pelos exemplos de Raz. A ideia é que parece difícil aceitar que uma autoridade possa ser “transitoriamente” ilegítima – i.e., que ora posso decidir que ela é legítima, ora posso decidir que não. Afinal, a peremptoriedade do comando parece implicar que ele deve sempre ser contado não como um prior, mas como um posterior - como uma razão conclusiva.

O estranho é que isso parece quebrar a analogia com instrumentos: ninguém se priva *a priori* da possibilidade de julgar quando eles são úteis ou não; posso rever minha submissão a um agente de investimentos, assim como posso deixar de usar um aplicativo, em ocasiões em que eles não se mostrem úteis – o que não implica que os deixe de usar completamente. Isso parece uma caracterização mais adequada da distinção raziana entre erros “claros” e erros “grandes”, vista na seção 2; erros claros são aqueles em que sabemos que a autoridade (ou

---

<sup>23</sup> Claro, ainda assim, pode ser justificável cumprir a ordem por outros motivos; e.g., por apego, medo ou admiração por F – alguma forma de “relação especial” com F.

qualquer outro instrumento) não são úteis. Afinal, se a legitimidade da autoridade deriva de seu caráter instrumental, por que quebraríamos a analogia nesse ponto?

Não há problema algum em acreditar que outrem tem maior probabilidade *a priori* de estar certo; o problema é acreditar que isso também vale sempre para a probabilidade *a posteriori*, no caso de um juízo determinado. Suponha que R saiba que F acerta com uma frequência 50% maior, e que F envie um e-mail dizendo que R deve comprar ações de certa companhia; no dia seguinte, pouco antes de confirmar a operação, R descobre que a companhia está quase quebrada. Não seria racional desconsiderar o e-mail neste caso?

Se R acreditar que, não importa quanta informação possua a mais, ainda assim o juízo de F será melhor do que o dele, R estará se comprometendo com um tipo estranho de dogmatismo (recusando evidência acessível – i.e., a informação “a mais”) e se considerando um *anti-expert* em relação ao caso. Essa é uma possibilidade lógica, e as outras pessoas poderiam, quiçá, constatá-lo, mas ele, não; pois, como aponta Sorensen (1988, 394-6), considerar a si mesmo um anti-expert implica crer num ponto-cego, do tipo “julgo que meus juízos não são confiáveis”.

Isso nos levaria de volta ao paradoxo de F e R; no caso, R estaria próximo de assumir que “devo fazer  $\alpha$  & não posso saber que devo fazer  $\alpha$ ”<sup>24</sup>. Esse problema vindica os que consideram que há algo de estranho na ideia de “submeter” sua capacidade de julgamento à autoridade. Se R deseje tal submissão, precisa de uma teoria afim à da “irracionalidade racional” para justificá-la.

## 5 Conclusões

Na primeira parte do artigo, apresentamos a concepção de serviço da autoridade de J. Raz, desenvolvida como uma resposta ao paradoxo com que começamos este artigo; Raz busca mostrar que seguir uma ordem pode ser justificado, desde que satisfeitas as teses da preempção, da dependência e da justificação normal: nesse caso, a ordem substitui as razões primárias, embora dependa (ainda que parcialmente) delas.

Como vimos em seguida, o paradoxo retorna quando a autoridade ou o súdito pronuncia um *ponto-cego* - um limite para aquilo em que estamos justificados em crer ou defender. A

---

<sup>24</sup> Isso é uma instância de (p& ~Kp), que, como já vimos, não pode ser conhecida – justamente por não ser passível de crença justificada.

primeira resposta ao paradoxo assume que o súdito “submete o próprio juízo” à autoridade, que se abstém de julgar por si mesmo; mas, na medida em que discorda da ordem, ele incorre em algo afim a uma inconsistência – embora se possa cogitar justificar a manutenção disso por meio de uma teoria da “irracionalidade racional”.

Alternativamente, assumindo a existência de falhas de racionalidade, pode-se apontar que um sujeito que segue uma ordem da qual discorda não está sendo irracional apenas *por seguir a ordem*, mas por manter *duas* atitudes incompatíveis entre si; sua falha pode ser antes epistêmica, se ele está deixando de acreditar naquilo em que tem mais razões para acreditar (i.e., no comando da autoridade legítima). Portanto, como em outras situações em que as evidências parecem conflitar com as crenças já mantidas por um agente, ele precisa definir, por si mesmo, qual é a atitude que ele está justificado em manter.

## **Capítulo 2: Interpretando a reivindicação de autoridade legítima: análise da refutação de Joseph Raz à incorporação de normas morais ao direito.**

**Resumo:** A partir de uma revisão crítica da literatura, analisa-se a incompatibilidade entre a possibilidade da incorporação de princípios morais ao direito e a sua natureza *autoritativa*, conforme arguido por positivistas exclusivistas como J. Raz. Alega-se que o argumento anti-incorporação é incompatível com premissas comumente aceites (inclusive por Raz) da teoria da interpretação jurídica – a menos que se pressuponha, entre as premissas, a versão forte da tese das fontes (que é o que Raz pretende provar).

### **1 Introdução**

Em *The model of Rules*, Ronald Dworkin critica a ‘tese das fontes’ do positivismo jurídico, aduzindo que o direito não se esgotaria no conjunto de regras definidos pelos testes factuais da “regra de reconhecimento”, como exposto em *O Conceito de Direito* por Hart; além dessas, o sistema jurídico apresentaria outra possibilidade de *standards* – os chamados ‘princípios’, cuja identidade e conteúdo seriam determinados não por uma regra de segunda ordem, mas por interpretação e deliberação moral (Dworkin, 1967). A essa objeção, Hart respondeu que sua teoria não implica que a identificação do direito ‘consiste apenas de fatos históricos’; na realidade, em alguns sistemas, o critério último de validade poderia incorporar (‘tese da incorporação’) princípios de justiça e valores morais:

He [Dworkin] treats my doctrine of the rule of recognition as requiring that the criteria which it provides for the identification of law must consist only of historical facts and so as an example of 'plain-fact positivism'. But though my main examples of the criteria provided by the rule of recognition are matters of what Dworkin has called 'pedigree', concerned only with the manner in which laws are adopted or created by legal institutions and not with their content, I expressly state both in this book (p. 72) and in my earlier article on 'Positivism and the Separation of Law and Morals' that in some systems of law, as in the United States, the ultimate criteria of legal validity might explicitly incorporate besides pedigree, principles of justice or substantive moral values, and these may form the content of legal constitutional restraints (Hart, 1994, 247).



Segundo essa linha de resposta, o “argumento dos princípios” seria um engano, na medida em que não se aplicaria a Hart. Alguns dos mais destacados autores positivistas adotaram outro estilo réplica ao argumento de Dworkin; a despeito do que a referência a conceitos morais, na prática jurídica, possa sugerir, o direito não os ‘incorpora’. Na realidade, segundo Joseph Raz, quando uma regra jurídica remete a um conceito desse tipo, estabelece uma lacuna no sistema, a ser completada pela autoridade que aplica a norma, de forma discricionária.

Let us assume, for example, that by law contracts are valid only if not immoral. Any particular contract can be judged to be *prima facie* valid if it conforms to the 'value-neutral' conditions for the validity of contract laid down by law. The proposition 'It is legally conclusive that this contract is valid' is neither true nor false until a court authoritatively determines its validity. This is a consequence of the fact that by the sources thesis courts have discretion when required to apply moral considerations (Raz, 1979, 75).

Portanto, há três concepções concorrentes nesse tema: o “interpretativismo” de Dworkin, o positivismo inclusivo de Hart (que aceita a tese da incorporação), e o positivismo exclusivo de Raz. Segundo os ‘positivistas exclusivos’, a impossibilidade de incorporação de conceitos morais não seria uma característica de apenas alguns sistemas jurídicos, mas uma ‘propriedade necessária’ do direito, decorrente de sua própria natureza *autoritativa*. Portanto, as teorias de Dworkin e Hart, nesse aspecto, seriam inconsistentes com a natureza do direito:

Dworkin's theory, one must conclude, is inconsistent with the authoritative nature of law. That is, it does not allow for the fact that the law necessarily claims authority and that it therefore must be capable of possessing legitimate authority. To do so it must occupy, as all authority does, a mediating role between the precepts of morality and their application by people in their behaviour. It is this mediating role of authority which is denied to the law by Dworkin's conception of it (Raz, 1996, 225-226).

A despeito da clareza do argumento e do seu apelo para uma conexão conceitual e intuitiva entre autoridade e direito, tudo indica que a maioria das pessoas não é (ou ao menos não se considera) formada por “positivistas exclusivos”. Se elas estiverem certas, deve haver algo errado com a teoria da autoridade de Raz, ou com a relação entre esta e a teoria do direito.

A seguir, explicitamos o argumento anti-incorporação de Raz e analisamos algumas das objeções apresentadas. Após, apresentaremos uma objeção ao argumento, usando premissas oriundas da própria teoria da interpretação de Raz. Ao concluir, elencamos uma possível réplica à objeção ora apresentada. Essa réplica, porém, já pressupõe a tese das fontes (que se quer provar).

## 2 O argumento de Raz

### *i. Todo sistema jurídico reivindica (claims) autoridade legítima.*

Essa é a tese da reivindicação; Raz a considera uma das propriedades fundamentais dos sistemas jurídicos, que os distingue de outros “sistemas normativos fechados” (Raz, 2002, 163); ela decorre da própria natureza da autoridade (Raz, 1986, 26-7), e é um dos elementos que a distingue da simples coerção. Ou bem o sistema jurídico ‘reivindica’ autoridade legítima explicitamente, ou é tratado como se a tivesse (Raz, 1995, 215). Essa ‘reivindicação’ pode ser constatada a partir dos seguintes elementos:

(1) the enforcement of law as exclusionary; (2) the use in the law of such terms as “right” and “duty”; (3) the official designation of legal institutions as “authorities”; (4) the claims of officials that subjects “owe” allegiance and “ought to obey the law”; and (5) the beliefs of officials that they have legitimate authority” (Himma, 2001, 283).

Pode parecer estranho atribuir atitudes proposicionais ou intencionais a um *sistema* jurídico (e não a pessoas); o sentido dessa atribuição é bem simples: autoridades jurídicas fazem tal reivindicação em nome da comunidade política: “Law makes claims only insofar as law-applying officials make those very same claims at the very same time and place” (Gardner, 2012, 131). Podemos tornar essa premissa mais aceitável aduzindo que há uma pressuposição ou pretensão de autoridade legítima imputável a uma instituição ou conjunto de pessoas; mesmo essa premissa ‘enfraquecida’ manteria a força do argumento de Raz<sup>25</sup>.

---

<sup>25</sup> Uma das principais objeções a Raz é que se poderia conceber um sistema jurídico sem que nenhum dos fatos correspondentes a essa reivindicação estivessem presentes (Dworkin, 2006, 231-2; Himma, 2001b, 289-90). De forma semelhante, pode-se alegar que a tese da reivindicação tem uma implicação epistêmica: se fosse uma verdade conceitual dos sistemas jurídicos, então seria impossível saber que existe um sistema jurídico sem saber se existe a correspondente reivindicação – embora esse seja precisamente o caso de um sistema jurídico do qual se conhece muito pouco, como o da Coreia do Norte (Himma, 2002). Não analisaremos tais críticas, no momento;

Tal reivindicação implica que sua teoria da autoridade não é axiologicamente vazia (*value-free*). Donde, mesmo que uma pessoa ou instituição seja “moralmente objetável”, ela só é uma autoridade se tratada por várias pessoas (elas devem acreditar, ou agir como se acreditassem) como moralmente defensável – do contrário, não poderia ser tratada como autoritativamente vinculante; a reivindicação em questão (assim como as diretrizes da autoridade) deve ser moralmente inteligível<sup>26</sup>.

*ii. Se o direito reivindica autoridade legítima, toda norma jurídica é autoritativa;*

Embora não haja localizado quem conteste essa tese, ela pode parecer um *non sequitur*, quando exposta sem uma explicação prévia. Na verdade, ela decorre da ideia de que o sistema jurídico pode ser identificado com o conjunto de normas correspondentes – i.e., que ele é um “sistema normativo fechado”, como defendido por Raz (2002, 160 e s.). Mas essa concepção do direito como um sistema fechado é precisamente a tese mais geral que Dworkin (2011, 401) rejeita – o que ele chama de “visão de dois sistemas”. Embora essa questão fuja ao nosso atual escopo, retornaremos a ela no fim do artigo, ao analisar as possíveis réplicas do positivismo exclusivista.

*iii. Para ser autoritativa, uma norma deve ser capaz de: a. guiar a conduta do agente, e b. ser identificada sem que a pessoa sujeita a ela precise deliberar sobre as razões que justificariam a ação determinada pela norma;*

Isso deriva da **tese da preempção**: “o fato de que uma autoridade requer a realização e uma ação é uma razão para sua realização, a qual não deve ser adicionada às outras razões relevantes quando se pensa sobre o que fazer, mas deveria substituir algumas delas” (Raz, 1994,

---

caso estejam corretas, é possível dividir a classe de sistemas jurídicos entre aqueles aos quais não se aplica a tese da reivindicação e aqueles aos quais se aplica – e é para estes que o argumento a seguir vale.

<sup>26</sup> Essa relação entre a pretensão de legitimidade e a capacidade de o direito exercer autoridade tem alguma base empírica, uma vez que há razões para crer que o sucesso dessa função preemptiva do direito depende de seus súditos tomarem sua autoridade como legítima (Priel, 2010)

Pode parecer arbitrário que, para que a reivindicação seja possível, devam estar presentes as características que tornem a reivindicação de autoridade verdadeira, mas não as que tornem a reivindicação de legitimidade verdadeira. De acordo com Raz, “If it [a legal system] lacks the moral attributes required to endow it with legitimate authority then it has none. But it must possess all the other features of authority, or else it would be odd to say that it claims authority” (Raz, 1995, 215). Contorna-se a objeção reformulando (ii): “o direito reivindica que sua autoridade é legítima” – o que quer dizer, simplesmente, que, para algo ser uma autoridade, deve reivindicar legitimidade.

214, trad. livre). Diretrizes *autoritativas* são, no vocabulário de Raz, “razões protegidas”: elas são, ao mesmo tempo, razões para realizar uma ação (razões de primeira ordem) e razões para excluir a deliberação sobre outras razões (razões excludentes, de segunda ordem); por isso, elas podem substituir as razões originais que o sujeito teria para agir. É isso que torna peremptória a autoridade.

*iv. Para ser identificável sem necessidade de levar em conta as considerações que a tornariam justificada, uma norma deve ser fundada numa atitude, crença ou intenção (estados mentais ou disposições) de alguém que reivindica autoridade legítima;*

Essa premissa parece seguir do que Raz supõe ser uma relação necessária entre interpretação e autoridade: “Making law is an intentional activity. One does not make law except through acts committed in the knowledge that they will make law. And one cannot make law intentionally if one is totally in the dark as to which law one is making” (Raz, 2009, 120).

Curiosamente, essa ideia parece ser refutada por um dos principais defensores de Raz, John Gardner, para quem pode existir criação de direito sem um agente determinado que o crie (é o caso do costume), nem mesmo a intenção de **criar** direito – i.e., existe criação de direito “sem querer” (é o que decisões judiciais, por vezes, fazem, ao estabelecer uma doutrina ou precedente). Nestes casos, não se trata de um caso de exercício de autoridade, embora nós possamos vir a tratá-los como *autoritativos* “num sentido derivado” (Gardner, 2012, 87). Contudo, se (ii) ou (iv) forem falsas (i.e., se há normas jurídicas que não são diretrizes *autoritativas*, ou se há diretrizes *autoritativas* que não são uma ‘função’ de atitudes intencionais de uma autoridade), a conclusão a que se pretende chegar infra – a refutação da tese da incorporação, na sentença (vi) – não se segue. Segue-se apenas a conclusão de que, para aquelas normas jurídicas que são, efetivamente, diretrizes *autoritativas* (i.e., um subgrupo do conjunto de normas jurídicas), aplica-se (vi).

*v. Normas que exigem que o sujeito interprete conceitos morais para cumpri-las demandam que ele delibere sobre as razões que justificariam a conduta prescrita;*

vi. Logo, por (iii) e (v), normas que exigem que o sujeito interprete conceitos morais para cumpri-las não são diretrizes autoritativas. Portanto, normas jurídicas não exigem que o sujeito interprete conceitos morais para cumpri-las.

### 3 A Objeção: Interpretação e Autoridade

Podemos prosseguir no raciocínio com algumas premissas derivadas da teoria de Raz sobre autoridade e interpretação, chegando a uma contradição:

vii. *Para ser capaz de guiar conduta/ser aplicada, uma norma deve ser compreendida;*

Essa condição parece decorrer, trivialmente, do caráter normativo do direito: se uma norma é um tipo de razão para agir, a relação entre ela e a conduta do agente não é de mera “causa e efeito”, mas sim algo semelhante ao que Mark Greenberg (2004) chama “rational relation”.

viii. *Para uma norma ser compreendida, deve-se interpretar a atitude em que ela se fundamenta;*

Considerando que o exercício de autoridade é um ato volitivo, Raz considera necessário que qualquer interpretação que busque conservar o significado da diretriz *autoritativa* deve levar em conta a intenção de quem a emitiu. É a *Tese da Intenção Autoritativa*: “to the extent that the law derives from deliberative law-making, its interpretation should reflect the intentions of its law-makers” (Raz, 2009, 275).

ix. *Diretrizes autoritativas devem ser baseadas (entre outros fatores) nas razões aplicáveis (i.e., razões dependentes) aos sujeitos obrigados por elas.*

A **tese da dependência** distingue o exercício da autoridade de outras formas de controle – como a simples coerção. Junto com a tese da preempção, forma a “concepção de serviço” da autoridade (Raz, 1995, 214).

x. *Uma autoridade é justificada quando o indivíduo a ela sujeito tem mais chances de atender às razões que se aplicam a ele aceitando as diretrizes dela e tentando segui-las do que se buscar seguir essas razões diretamente; portanto, para ser legítima, a autoridade deve levar em conta as considerações que tornariam a norma justificada.*

A **tese da justificção normal**, embora seja logicamente independente das duas teses da concepção de serviço(e vice-versa), é explicada por elas. Se o que distingue a autoridade é o ‘serviço’ que ela presta, ela pode ser justificada quando for mais racional para um sujeito ‘terceirizar’ a deliberação do que fazê-la por si mesmo.

xi. *Para interpretar a atitude em questão, é necessário levar em conta, ao menos em alguns casos, o que seria o caso se a norma que ela fundamenta fosse legítima – i.e., dentre as interpretações possíveis, selecionam-se as compatíveis com a afirmação de que a autoridade é legítima.*

Esta é a premissa mais discutível que introduzimos no presente argumento; não a encontramos na obra de Raz. Embora seja necessário fazer uma investigação distinta para sustentá-la, ela é coerente com sua teoria da interpretação.

A tese (xi) assume que a compreensão de algumas atitudes proposicionais, em parte, envolve juízos normativos<sup>27</sup> – principalmente quando as proposições embutidas em tais atitudes são elas mesmas normativas. O ponto é levantado *en passant* por Kevin Toh – que não o desenvolve por supor que não levaria a implicações relevantes sobre a tese das fontes; no entanto, como veremos, leva a uma objeção à refutação do incorporacionismo:

Of course, some philosophers have argued that attributions of psychological states, and of contents more generally, involve deployment of normative judgments (see e.g. Kripke 1982; Brandom 1994). If they were right, then ascertainments of social facts themselves would inevitably require appeals to normative considerations. But I believe that I am entitled to set aside this line of thinking for the following reason. Whatever the nature of psychological attributions, **once that task is done**, there is a question as to whether some further appeals to normative considerations are necessary to determine whether a particular law exists in a community or whether a law has a

---

<sup>27</sup> A plausibilidade dessa tese é reforçada pela sensibilidade das atribuições intencional a avaliações normativas na psicologia *folk* – as pessoas são mais suscetíveis de considerar intencionais os efeitos colaterais negativos de uma ação que os positivos (Knobe, 2003).

particular content. According to the common conception of legal positivism, the answer is negative. It is this answer (or a particular version of that answer which I will isolate in the next section) with which I want to take issue. For this purpose, in the spirit of philosophical division of labor, I will bracket the line of thinking about the normative nature of psychological attributions (Toh, 2008, 451; grifo nosso).

Nosso ponto aqui é precisamente a “tarefa prévia” de atribuir um certo estado mental a outrem. Via de regra, ao interpretar um determinado conjunto de sentenças e atitudes proposicionais inter-relacionadas de outrem, fazemo-lo dentro de um contexto, e pressupomos que elas fazem sentido, que são coerentes entre si, e que os conceitos utilizados são “públicos” (são, ou ao menos podem ser, compartilhados conosco). Essa “operação mental” pode ocorrer de forma automática e intuitiva, ou ser resultado de deliberação refletida; pode-se, por vezes, solicitar ao declarante que explicito o significado do enunciado – mas isso pode ser embalde, seja porque não nos livra de interpretar o significado de sua explicação, seja porque (a menos que se assuma uma “versão forte” da tese da autoridade da primeira pessoa<sup>28</sup>) o declarante pode não entender melhor do que o ouvinte o significado de sua asserção inicial. Frequentemente, porém, essa solicitação pode ser fisicamente impossível – e.g.: não seria possível ou útil que um juiz ou cidadão, para assegurar-se da compreensão de uma lei, consultasse o conjunto de parlamentares.

---

<sup>28</sup> Raz não parece pressupor a autoridade da primeira pessoa: “Law-making does not imply knowingly making the law one intended, but knowingly making a law that one could have known one is making. Therefore, law-makers can always know the meaning of the law they make. Moreover, the law they make must be the law **they are understood** to have made. It is simply the standard made by their law-making act when understood as such acts are normally understood in the context of their making. **Meaning, as we know, is a public and social phenomenon.** Words and actions have the meaning they are taken to have. Since they know that, they can, when making a law, take those actions which they know, or can establish, will be understood to be making that law which they want to make” [Raz, 2009, 121; grifo nosso]. Note-se que do fato de que o legislador *pode sempre* conhecer a lei que ele fez não segue que ele *sempre a conheça*; o domínio da falibilidade humana se estende ao auto-conhecimento.

Mas disso Raz parece tirar uma estranha conclusão: “It follows that it does not matter which way law-making actions will be interpreted. Provided there are established ways or conventions for interpreting them, they will ensure that the law-makers intentions become law, that the law is as the law-makers intended it to be” [ibid.]. A cópula ‘is’ aqui não indica identidade: da mesma forma como não podemos assegurar que legisladores *sempre* conhecem suas intenções, também não podemos garantir que *sempre* conhecem os meios estabelecidos para interpretá-las – mormente num sistema complexo como o direito. Se edito uma lei pronunciando “A&B”, mas os critérios de reconhecimento só permitem validar “A”, pode-se dizer que apenas “A” é direito; não é que as convenções de interpretação sejam ambíguas ou vagas, como Raz pressupõe, mas o simples fato de eu desconhecê-las ou ignorá-las me permite falhar, de uma forma simples, na determinação da lei que intencionara. Nesse caso, pode-se discutir se a lei é apenas em parte a vontade do legislador, ou se é apenas parte da vontade do legislador; de qualquer forma, não é idêntica ao objeto de sua intenção, e como tal, não é *transparente* a ele – embora esteja garantido o requisito que Raz impõe de que a interpretação deve preservar parcialmente a intenção original.

Em suma, se os enunciados da autoridade são normativos e usam a mesma linguagem que a moral<sup>29</sup>; se ela demanda autoridade legítima (ou se é tratada como tal); se o indivíduo a quem eles são dirigidos compartilha os mesmos conceitos que ela utiliza; se a tese da dependência (ou a da justificação normal) é verdadeira (i.e., os enunciados *autoritativos*, supõe-se, estão justificados em razões dependentes), então é plausível que ele utilize a deliberação moral para compreender a intenção da autoridade – e, dessa forma, compreender a norma que deve guiar sua conduta. Assim, se a deliberação moral lhe franquear acesso à psicologia moral da autoridade, o sujeito estará justificado em usá-la; e se sua conclusão for correta (i.e., se lhe permite compreender a intenção da autoridade), pode-se dizer que obteve conhecimento do significado da norma.

Na realidade, Raz admite essa possibilidade quando afirma que é possível que o sujeito compreenda e siga a norma ao estabelecer o que é moralmente requerido:

It may be worth pointing out that the view does not require that *every* person should be able to identify the law without recourse to moral considerations. [...] Nor does it deny that it may be possible to identify the law using moral reasoning. Indeed, it follows from my doctrine of authority that if one knows that the legislature did enact the morally required law it should be possible to identify the law by establishing which law was morally required. However, the thesis that it is possible to identify the law without recourse to moral argument is not satisfied if to identify the law one has either to rely on moral considerations or to rely on the word of an expert who relies on moral considerations (Raz, 2009, 4).

Dado que é possível ter conhecimento da identidade e do conteúdo de uma norma jurídica por deliberação moral dessa forma indireta em alguns casos, não há nenhuma razão *a priori* para pensar que não há, em algum sistema jurídico, **pelo menos uma** norma cuja identidade e conteúdo **somente** possam ser identificados por uma tal deliberação moral – i.e., um caso em que é preciso perguntar-se se algumas condutas são razoáveis, ou justas, ou cruéis, ou iníquas... a fim de ter acesso à intenção da autoridade, e assim ao significado da norma. Ao fazer isso, o intérprete/seguidor da norma estará seguindo a autoridade; e, principalmente, será

---

<sup>29</sup> Com a diferença de que costumam ser pronunciados de um ponto de vista “destacado” (Raz, 1979, 156; 2002, 177)



falível, na medida em que é possível errar ao fazer tal operação<sup>30</sup>. Enfim, se isso é possível para uma única norma, segue-se que é possível para várias delas, em diversos sistemas.

Esse tipo de “deliberação moral indireta” é paradigmático de sistemas de “autoridade divina”; está pressuposta em discussões tão distintas como o diálogo entre Sócrates e Eutífro sobre o conceito de impiedade, ou na relação entre a Lei Revelada (lei positiva divina) e a lei humana na tradição cristã, ou nos debates sobre a *fiqh* (teoria do direito islâmico), a *ijtihad* (a interpretação jurídica nas cortes islâmicas), e a *Shari’a* (Emon, 2009). Mesmo o preceito da *halakhah* (a tradição jurídica judaica) de que a interpretação da Torah prescinde de intervenção divina (“não está nos céus” – *o ba-shamayim hi*, do Deuteronômio, 30:12) deve ser compreendido no sentido de que essa é a interpretação da Sua Vontade<sup>31</sup>.

Raz considera que sua conclusão de que “interpretation in accord with legislative intention is demanded by any realistically conceivable theory of authority” não se aplica à “autoridade baseada em revelação” (Raz, 2009, 298). Contudo, se esse requisito da teoria da interpretação decorre de uma exigência do conceito de autoridade, por que deveria excluir a ideia de autoridade divina<sup>32</sup>? Afinal, nada indica que este conceito (mesmo que seja vazio, ou fictício) esteja necessariamente não contido dentro do de autoridade em geral; pelo contrário, historicamente, os filósofos os trataram de forma semelhante.

Se a distinção relevante é que a autoridade divina seria infalível ou onisciente (tendo, pois, acesso sobrenatural e privilegiado ao significado e à aplicação de suas normas), seria possível sustentar que para ela vale a Tese da Intenção Radical (i.e., o mandamento divino é completamente idêntico à intenção divina), enquanto para autoridades humanas vale a Tese da Intenção Autoritativa (Raz, 2009, 265 e s.). Ainda assim, *epistemicamente*, o intérprete estaria na mesma situação nos dois casos, na medida em que teria conhecimento<sup>33</sup> limitado do estado

---

<sup>30</sup> Se, ao cometer esse erro, o intérprete cria novo direito, é outra questão. O ponto é que, ao interpretar corretamente a diretriz *autoritativa*, ele **não** cria uma nova diretriz. Um dos pontos notáveis da teoria da interpretação de Raz é que, mesmo quando é impossível discriminar entre interpretações que conservam significado e interpretações criativas, isso não significa que, conceitualmente, a distinção não exista (v. Raz, 2009, 116-20).

<sup>31</sup> Essa é uma das interpretações usuais da parábola talmúdica do “forno das serpentes”, em que um grupo de rabinos discute com o próprio Deus sobre a impureza de um forno (v. Luban, 2004).

<sup>32</sup> Não estou sugerindo que toda ocorrência da palavra ‘autoridade’ refere o mesmo conceito. Nesse caso, porém, os teóricos da autoridade divina usualmente a tratam como algo afim à autoridade secular; mesmo que uma delas seja uma ficção, não se segue que o conceito seja inconsistente. Gardner (2012), ateu e um dos maiores defensores de Raz, usa, de forma análoga, a noção de autoridade divina para traçar semelhanças entre a Norma Fundamental de Kelsen e a Regra de Reconhecimento – além de criticar Dworkin.

<sup>33</sup> Conhecimento proposicional. Ao falar do intérprete, estou excluindo a figura do ‘profeta’, cujo conhecimento da revelação seria privilegiado, por inspiração divina. Note-se que não estou pressupondo que o intérprete esteja correto em sua premissa básica – i.e., que Deus existe e emitiu um comando.

mental da divindade, podendo conhecê-lo somente de uma maneira semelhante a como conhece os estados mentais de seres humanos.

Se isso é correto, o imperativo “honrar pai e mãe” é (ou ao menos pode ser) o mesmo se escrito em tábuas de pedra providas dos céus, proferido por sacerdotes, ou repetido por progenitores (i.e., o mesmo quanto a seu conteúdo semântico). E se, no primeiro caso, é uma diretriz *autoritativa* capaz de guiar conduta (pois, ao desonrar meus pais, não apenas deixo de seguir as razões que teria para honrá-los – que poderiam conflitar com outras razões primárias – mas falho em seguir uma prescrição divina<sup>34</sup>) e de excluir as razões primárias que eu teria para não praticar a ação prescrita (ao invés de uma outorga de discricionariedade para decidir, por mim mesmo, o que conta como honra em cada caso), também pode sê-lo nas demais situações; isso significa que o imperativo satisfaz a tese da preempção supra.

Essa versão mais ‘fraca’ da tese da preempção – i.e., a ideia de que ela exclui *apenas as razões conflitantes com a norma* – é coerente com outras passagens da obra de Raz:

A binding authoritative directive is [...] an exclusionary reason, that is a reason for not following (i.e., not acting for) reasons that conflict with the rule. That is how authoritative directives pre-empt. [...] These exclusionary reasons do not, of course, exclude relying on reasons for behaving in the same way as the directives requires. [...] Hence the pre-emption excludes only reasons that conflict with the authority’s directive (Raz, 2009, 144).

*xii. por (xi), para interpretar a atitude que fundamenta a norma, pode ser necessário levar em conta as considerações que tornariam a norma justificada;*

*xiii. por (xii), para uma norma guiar conduta, pode ser necessário levar em conta as considerações que a tornariam justificada;*

*xiv. **Contradição:** por (iii), para ser autoritativa, uma norma deve ser capaz de guiar conduta e de ser identificada sem que a pessoa sujeita a ela precise deliberar sobre as razões que justificariam a ação determinada pela norma; por (xiii), para uma norma guiar conduta, pode ser necessário levar em conta as considerações que a tornariam justificada.*

---

<sup>34</sup> De forma semelhante, Dworkin (2006, 207) imagina um comerciante que, tendo operado a vida toda pelo princípio *caveat emptor*, fosse convertido a uma religião que pregasse a boa-fé nas relações comerciais – ele estaria justificado em pensar que violara uma prescrição religiosa.

Para evitar a conclusão em (xiv), bastaria abdicar da segunda sentença (b) de (iii); isso manteria o restante da teoria da autoridade intacta. De fato, a teoria da autoridade em nada perderia sem essa “proscrição da deliberação moral” – exceto que, nesse caso, seria impossível concluir por (vi), a refutação da tese da incorporação.

#### 4 Conclusão: Por que o argumento anti-incorporação falha?

Podemos sintetizar, agora, o argumento de Raz: se (a) o direito reivindica autoridade legítima, então, (b) ele deve ser capaz de guiar a conduta do sujeito e de ser identificado sem recurso à deliberação moral; (c) conceitos morais dependem de deliberação moral para sua aplicação; logo, (d) normas que incorporem conceitos morais exigem deliberação moral; (e) por (b & c), normas jurídicas não os incorporam. Isso fundamenta a adoção da concepção alternativa de direito, a tese das fontes, a qual pode ser lida como: (f) o direito tem por fundamento, exclusivamente, um conjunto de fatos (atitudes, crenças ou intenções – estados mentais ou disposições) relativos a uma autoridade.

Não faltam críticas a (f; mas a que ora formulamos é a seguinte: (g) para ser aplicada, uma norma deve ser compreendida, o que implica conhecer e interpretar a atitude proposicional em que se fundamenta; (h) essa interpretação leva em conta (ao menos em alguns casos) o que seria o caso se a respectiva autoridade fosse legítima – i.e., a própria compreensão dessa atitude pode demandar juízos normativos; (i) donde, ao menos em alguns casos, pode-se conhecer uma norma jurídica mediante deliberação moral; (j) então, ao menos em alguns casos (como quando não se dispõe de informação direta sobre a intenção da autoridade), é necessário deliberação moral para segui-la – o que contraria (b).

O leitor talvez nos acuse de injustiça para com Raz; afinal, como mencionado supra, ele poderia “morder a bala” e afirmar que, quando a interpretação depende de deliberação moral, mesmo que para descobrir a intenção da autoridade, estamos diante de uma **lacuna jurídica** – i.e., como não há critérios factuais com base nos quais decidir se a proposição jurídica é verdadeira ou falsa, o intérprete tem discricionariedade (Raz, 1979, 75). Dito isso, cabem quatro considerações.

Primeiro, o que atacamos supra foi apenas a refutação da tese da incorporação, a partir do argumento da autoridade; nosso raciocínio é agnóstico quanto à tricotomiatese das fontes, interpretativismo e incorporacionismo. Não estamos fazendo uma defesa da tese da incorporação.

Segundo, é preciso evitar usar uma dessas teorias como *premissa* para concluir pela refutação das demais, sob pena de petição de princípio. Isso implica que não se pode, pois, argumentar em prol da tese das fontes a partir da concepção *raziana* de lacuna, **pois esta pressupõe a própria tese das fontes**; da mesma forma, não podemos fundar o interpretativismo de Dworkin na teoria do direito como integridade pressupondo sua teoria da interpretação<sup>35</sup>.

Terceiro, deve-se evitar confundir a questão “o que o direito determina” com “como o caso deveria ser decidido, de acordo com o direito” – uma distinção frequentemente realçada por Raz (1995, 328). A primeira corresponderia ao debate sobre o conceito de direito, e a ela se aplicaria a tese das fontes; a segunda refere-se ao raciocínio jurídico, e Raz (1995, 340) deixa explícito que não vê qualquer problema em que este envolva argumentação moral. Não obstante, é precisamente essa distinção que é subestimada pelo argumento anti-incorporação de Raz. Pois, para chegar à conclusão (ontológica) de que conceitos morais não podem ser incorporados ao direito, o argumento da autoridade usa a premissa (prática e epistêmica) de que a teoria da autoridade proscreve a deliberação moral.

Se a teoria do direito é uma teoria ontológica sobre o significado e a validade de normas jurídicas (e de proposições derivadas da aplicação dessas normas), sem impor restrições epistêmicas ou metodológicas sobre como os agentes devem fazer inferências a partir dessas normas (e, portanto, sobre como as *aplicam*), então ela não acarreta qualquer restrição *a priori* quanto ao uso da deliberação moral – e, conseqüentemente, não pode servir para justificar o argumento anti-incorporação supra. Podemos imaginar, por exemplo, que os conceitos morais são passíveis de incorporação tanto quanto noções empíricas ou matemáticas, na medida em que a *aplicação* de uma norma jurídica depende da correção da proposição incorporada.

Nesse caso, a incorporação pode parecer assaz contingente e trivial; afinal, quando o direito demanda um raciocínio aritmético ou uma investigação empírica, não dizemos que a matemática e as ciências foram incorporadas ao direito. Contudo, dizemos, sim, que “X deve pagar mil reais”, “um testamento, com apenas duas testemunhas, é inválido” ou “a mortandade de peixes constitui crime ambiental” são proposições jurídicas; e, se, na aplicação do direito,

---

<sup>35</sup> Dworkin (2011, 403) argumenta que o debate sobre o conceito de direito é irremediavelmente circular – mas que sua teoria seria superior porque, no *interpretativismo*, esse pode ser um círculo virtuoso: “We construct a conception of law— an account of the grounds needed to support a claim of right enforceable on demand in that way— by finding a justification of those practices in a larger integrated network of political value. We construct a theory of law, that is, in the same way that we construct a theory of other political values— of equality, liberty, and democracy” (Dworkin, 2011, 405).

uma parte comete um erro de fato, dizemos que *o direito foi violado*. E tem de ser assim, se temos a pretensão de aplicar o direito à realidade (se existe tal coisa como o “suporte fático” de um fato jurídico), mormente para uma teoria que frisa que “o direito é uma questão de fato”. Mas, se o argumento anti-incorporação fosse correto, mesmo esse tipo de incorporação trivial seria vedado; para Raz (1979, 75), não deveríamos dizer “esse contrato é inválido, porque é injusto, e há uma norma que veda contratos injustos”, e sim “esse contrato é inválido porque um juiz decidiu que o é, autorizado por uma lei que veda contratos injustos”.

Por fim, mesmo que, no nível teórico, negássemos (em nosso contra-exemplo supra) haver incorporação de normas morais ao direito, essa negação não estaria acessível ao intérprete. Assim como a autoridade não pode afirmar que é ilegítima, nem que seus comandos não são justificados por razões dependentes (em virtude da tese da reivindicação e da tese da dependência<sup>36</sup>), o intérprete, em nosso contraexemplo supra, não poderia afirmar que está criando novo direito – ao menos não quando tais fatos ocorressem<sup>37</sup> (salvo hipocrisia, claro).

Afinal, se o intérprete reconhecesse que a norma é indeterminada, e não apenas incerta, e que a investigação moral lhe leva a *criar*, ao invés de *descobrir*, o direito, ele não poderia apresentar suas decisões como vinculantes *por causa* da autoridade original. Portanto, à tese anti-incorporação se aplicaria aqui uma condenação semelhante à que Bernard Williams (2005, 165) imputou ao utilitarismo: se verdadeira, é melhor não crer nela; se falsa, certamente é melhor não crer nela.

---

<sup>36</sup> A reivindicação de legitimidade, nesse ponto, tem um efeito afim ao da “pretensão de correção” de Robert Alexy (v. Gardner, 2012, 125).

<sup>37</sup> Obviamente, não estou referindo os casos em que a lei outorga discricionariedade ao juiz de forma expressa.

### **Capítulo 3: O que Tartaruga dirá a Aquiles, ou “levando a sério a interpretação tradicional do argumento da batalha naval”**

**Abstract:** This dialogue between Achilles and the Tortoise – in the spirit of those of Carroll and Hofstadter – argues against the idea, identified with the “traditional” interpretation of Aristotle's “sea battle argument”, that future contingents are an exception to the Principle of Bivalence. It presents examples of correct everyday predictions, without which one wouldn't be able to decide and act; however, doing this is incompatible with the belief that the content of these predictions lacks a truth-value. In the end, it's suggested that the real problem of future contingents might be a limitation on what an agent can believe before a decision.

**Resumo:** Este diálogo entre Aquiles e Tartaruga – no espírito dos de Carroll e Hofstadter – argumenta contra a ideia, identificada com a interpretação "tradicional" do "argumento batalha naval" de Aristóteles, de que futuros contingentes são uma exceção ao princípio da bivalência. Apresentam-se exemplos de previsões cotidianas corretas, sem as quais não se seria capaz de decidir e agir; no entanto, isso é incompatível com a crença de que o conteúdo destas previsões não tem valor de verdade. No final, é sugerido que o verdadeiro problema de futuros contingentes pode ser uma limitação a respeito daquilo em que um agente pode acreditar antes de uma decisão.

Aquiles e Tartaruga estão discutindo e jogando Batalha Naval.

**Aquiles:** ... então Aristóteles conclui que o infinito, nesse caso, só existiria em potência. Logo, como já sabia, alcançarei você. QED.

**Tartaruga:** Você, sem dúvida, está confiante. Mas note que isso é um non sequitur: você apenas teria provado que pode me alcançar, mas não que sabe que vai me alcançar, necessariamente. A menos que você haja sido convertido por Zeno ao determinismo. B5?

**A.:** Água! Bem, provar isso logicamente seria impossível. Afinal, como provará um professor gaúcho (Barbosa Filho, 1999, 15) em alguns anos, agir é tornar uma proposição verdadeira, é tornar determinado o que não o é. Logo, não podemos conhecer futuros contingentes – sobre os quais ainda nos cabe agir. B4?

**T.:** Água! Poderia explicar um pouco melhor? A6.

**A.:** Como Aristóteles (1963; *De Int.*,9) dirá em alguns anos, suponha que façamos a previsão de que amanhã não haverá uma batalha naval; suponha que é verdade que ela não ocorrerá. Isso implica que também era verdade ontem que ela não ocorreria, assim como no dia anterior. Mas se essa proposição sempre foi verdadeira, significa que a batalha naval jamais poderia ocorrer; trata-se de uma necessidade lógica e, a menos que voltemos àquela desagradável discussão sobre *modusponens* (Carroll, 1895), a lógica é inescapável. Logo, se nossa previsão estiver correta, a batalha é impossível; se estiver errada, a batalha é necessária. Mas isso é absurdo; significa que não teríamos qualquer tipo de livre-arbítrio (pois tudo que ocorrerá está fadado a ocorrer), e a vida seria inócua: “a tale told by an idiot, signifying nothing”.

**T.:** Sabe, você me lembra muito a história de Arjuna, um general indiano que, diante da perspectiva de desolação que a guerra traria, considerou tornar-se pacifista na véspera da grande batalha; seu conselheiro Krishna (na verdade, o avatar de uma divindade) o reprovou, alegando que tudo já estaria determinado, e *que*, na prática, os guerreiros que morreriam já estavam mortos, donde o melhor para Arjuna seria lutar e colher a glória que lhe cabia. Mas Arjuna ainda se perguntou: então, por que lutar (*Bhagavad-Gita*, III, 1-2 – Prasad, 1996)?

**A.:** Exatamente. Odeio guerreiros que se abstém de lutar, fugindo de sua missão. Mas, voltando ao assunto, a solução de Aristóteles (ao menos, segundo a chamada interpretação tradicional) é abandonar a bivalência: conquanto não seja verdade que amanhã haverá uma batalha, nem que não a haverá, a disjunção é verdadeira: “ou haverá uma batalha, ou não haverá”. Isso preserva nosso livre-arbítrio (Schmidt, 2009, 8). Afinal, se o futuro ainda é indeterminado, não há nenhum fato que corresponda a essas sentenças e possa torná-las verdadeiras; como poderíamos atribuir-lhes um valor de verdade?

**T.:** Não estou plenamente convencido de que não há uma versão enfraquecida da falácia modal nessa interpretação, como acusa G. E. M. Anscombe (apud. Sorensen, 2003, 122-3); afinal, o que esse uso de “determinado” ou “necessário” acrescenta a uma proposição verdadeira? Além disso, junto com o princípio da bivalência, parece que teríamos de sacrificar a ideia de que a disjunção é uma função do valor de verdade dos disjuntos. Isso não pode ser feito com o coração leve, pois é a partir dessa relação veri-funcional que podemos, entre outras coisas, fazer previsões certas mas imprecisas, do tipo “um dia morrerei” - equivalente a “morrerei amanhã, ou no dia seguinte, ou no dia seguinte....”.

**A.:** Talvez você tenha razão. Duas ressalvas, porém: como uma crítica ao Estagirita, isso seria anacrônico, já que não estamos tratando de uma teoria *Tarskiana* da verdade (Barbosa Filho, 2004). Se você deseja fazê-lo, deveria escolher outro alvo, como a lógica multi-valorativa de Łukasiewicz (1930). Mas alguém que nos é posterior cronologicamente, mas Prior no nome (Prior, 1967), desenvolverá uma bela teoria para lidar com essa crítica funcionalista.

**T.:** Sem dúvida, seria divertido usar lógica não-aristotélica para salvar Aristóteles<sup>38</sup>. E adoro a ideia de infinitamente muitos valores de verdade – poderíamos discuti-los para sempre....

**A.:** Por favor, não! Retiro o que acabei de dizer.

**T.:**... mas não sou exatamente um especialista em lógica modal, ainda mais se exige lidar com a notação polonesa; às vezes, preciso ir devagar até com algo simples como *modusponens*, como já pudemos constatar em outra ocasião. Minha dúvida é muito menos técnica que isso. Nem tenho certeza se é um argumento geral a favor da bivalência, como veremos - ou se isso se aplica somente àqueles que a rejeitam a fim de manter uma teoria da verdade como correspondência.

Se você me permitir, vamos avançar para um novo caminho, para que possamos *convergir*: você afirmou que iria me alcançar e se comprometeu com a verdade dessa proposição. Agora, pode sentir-se tentado a dizer que estava enganado; isso seria um erro normal, como todos podem cometer ao tentar fazer uma previsão ordinária. Mas não podemos abster-nos de fazer previsões – algumas erradas, outras corretas: seria impossível, e.g., planejar e governar (tomar decisões sobre taxa de juros, agricultura, guerra, etc.) se não pudéssemos acreditar que as coisas serão de uma determinada maneira – ainda que essa crença possa ser falsa, ou imprecisa. Todas as nossas ações, em especial as que dizem respeito a outrem, visam a um estado de coisas futuro que queremos produzir. E não poderíamos visar a coisa nenhuma se não pudéssemos raciocinar “se eu fizer isso, ele fará aquilo. Logo, farei x; ele fará y”. Do contrário, seria um desemprego geral de estrategistas: “that would be complete anarchy in the conduct of men, with experience being plainly in contradiction. Such an assumption would make the existence of economy just as impossible the economic theory” (Morgenstern, 1935, 175)

---

<sup>38</sup> Na verdade, o seguinte argumento da Tartaruga a partir de atitudes proposicionais não é muito eficaz contra as teorias anti-realistas em geral (para as quais uma dissociação entre crença e verdade não implica nenhum problema), e em especial contra as teorias justificacionistas, como as de Dummett. (1991) - uma vez que essas teorias colapsam a classe do que pode ser justificadamente acreditado com o que pode ser verdade.



Donde poderíamos acrescentar: se o conhecimento de uma proposição é incompatível com o agir sobre ela (pois é a ação que a torna verdadeira), também é impossível agir sem acreditar que várias outras sentenças sobre futuros contingentes são verdadeiras. Chame-o de exigência doxástica: como podemos abdicar do conhecimento (e verdade) de  $p$ , e ainda acreditar em  $p$ <sup>39</sup>?

Pior ainda: há uma gama de decisões (se posso chamá-las assim) que animais sencientes como nós sequer precisam tomar de forma consciente, porque sabemos que um estado de coisas ocorrerá, embora se trate de um futuro contingente sujeito à decisão de alguém. Eu sei, neste momento, que você não ficará nu, que não cometerá suicídio, que não vai embora sem despedir-se. Sei disso porque sei que você é um animal suficientemente racional, com certos interesses e instintos; sei que você tem várias razões para não o fazer, e nenhuma para fazê-lo (exceto, talvez, para contradizer-me). Sei disso mesmo que a hipótese sequer haja passado pela sua cabeça em momento algum. E posso aplicar a mesma previsão a qualquer animal inteligente normal – com exceção, talvez, de alguns existencialistas.

Mas, se a interpretação tradicional fosse a teoria correta sobre futuros contingentes, mesmo que você decidisse tirar a roupa, não poderíamos acreditar que seria verdadeiro que você o faria – afinal, você ainda poderia mudar de ideia até o momento da ação. Logo, nem mesmo a sua decisão pode suprir essa lacuna do valor de verdade, até a ação ocorrer.

**A.:** Curioso, pois Aristóteles aparentemente não ignora essa relação entre crenças e estados futuros. Afinal, se toda ação visa a um bem (EN, I), então o agente deve acreditar que o bem seguirá depois da prática da ação; i.e., as crenças e desejos necessitam a ação (EN, VII, 3). Ele não parece ver incompatibilidade entre a teoria da ação e da responsabilidade moral com a noção de determinismo psicológico relacionado à formação do caráter (EN, III, 5) arraigada em nossa cultura grega – a ideia de que nossas ações são determinadas por quem nós somos, por como nosso caráter é constituído. Mas, se o determinismo psicológico não é incompatível

---

<sup>39</sup> Vamos mostrar que isso é uma exigência séria, usando uma versão da “prova de incognoscibilidade de Fitch” (apud. Sorensen, 2014). Seja  $p$  uma variável para qualquer proposição e  $K$  para o “operador do conhecimento”. Assim:

1.  $K(p \ \& \ \sim Kp)$  (traduzido para "Eu sei '  $p$  ', mas não sei '  $p$  '")
2.  $Kp \ \& \ K \sim Kp$  (o conhecimento é distribuído pela conjunção)
3.  $\sim Kp$  (conhecimento implica verdade)
4.  $Kp$  e  $\sim Kp$  (*reductio ad absurdum*).

É claro que existem infinitas proposições para as quais  $(p \ \& \ \sim Kp)$  é verdadeiro - isto é, todas as verdades que ignoro, como “o número de estrelas é ímpar” ou “o número de estrelas é par”. Então, como (1) é uma contradição, apesar da verdade de  $(p \ \& \ \sim Kp)$ , o problema (assumindo que conhecimento é crença verdadeira justificada) com esta sentença é que eu não posso *acreditar justificadamente* em “ $p$  e acredito que não sei  $p$ ”. Mas isso é precisamente o que alguém tem a dizer se quiser negar a exigência doxástica.

com a noção de responsabilidade moral (Dworkin, 2011, 236-7), porque a “necessidade veritativa” (para usar a expressão de Barbosa Filho, 2004, 234) o seria? Como poderá se perguntar Hardie (1968, 278), qual das duas versões do Aristóteles é a real: o libertário ou o compatibilista – ainda que meramente “psicologicamente compatibilista”?

**T.:** Ambas, quiçá. Talvez ele nunca se convença plenamente de nenhuma dessas teorias. Talvez escreva um excelente diálogo sobre o tema – que se perderá, para sempre. Alguns intérpretes podem discordar de mim, mas creio que nunca saberemos; nem por isso penso que não há verdade a respeito.

**A.:** Poderíamos dizer, por outro lado, que as crenças sobre futuros contingentes embutem enunciados meramente prováveis; seria razoável avaliar seus valores de verdade como números reais no intervalo[0, 1], aplicando os princípios correspondentes do cálculo de probabilidades (Łukasiewicz, 1970, 47-48). Isso atenderia à sua exigência de que uma teoria da verdade deve ser compatível com a crença em futuros contingentes.

**T.:** De fato, atenderia. Mas primeiro, note que não vindicaria a “interpretação tradicional”, uma vez que esta deriva, a partir de uma visão do senso comum de correspondência, a tese de que futuros contingentes não têm valor de verdade – em vez de muitos deles, como você propõe. Como eu disse antes, você está tentando resgatar o Estagirita com lógicas não-aristotélicas.

Além disso, você ainda me deve uma explicação de por que eu penso que:

- (i) minhas crenças sobre o futuro são atitudes comprometidas com a verdade das respectivas proposições, as quais se referem a fatos futuros;
- (ii) a proposição P, referente a um fato F, é a mesma quando dita antes (i.e., quando é uma previsão) ou depois da ocorrência de F (uma descrição)<sup>40</sup>;
- (iii) há uma diferença entre dizer que “haverá uma batalha naval” e “é provável que haja uma batalha naval” (o que explica (i));
- (iv) a segunda sentença (“é provável...”) *justifica* a crença na primeira (“haverá...”), e a verdade da segunda costuma ser explicada a partir da verdade da primeira – eu diria até que a relação entre elas segue algo afim a uma função – ou algo como o que Sorensen (2006, 607) defende: se uma proposição p carece

---

<sup>40</sup>Talvez isso não implique que o valor de verdade dos dois enunciados seja o mesmo. Para MacFarlane (2003), o valor de verdade dessas afirmações seria relativo ao contexto de avaliação.

de valor de verdade, a proposição modesta “talvez p” também carece. Se, para dar um exemplo grosseiro, há nove vezes mais mundos possíveis em que P é verdadeiro (ou se a frequência desse acontecimento é nove vezes maior que seu contrário), podemos dizer que é provável que P.

Seria curioso eu estar errado no primeiro ponto, sobre o conteúdo de minha crença; aliás, pergunto-me o que seria acreditar que estou errado, nesse caso – não no sentido de que o conteúdo de minhas crenças é falso, mas no sentido de que, na realidade, não sei qual o conteúdo dessas crenças! E isso não ocorreria apenas num caso excepcional, como quando alguém professa uma crença absurda (“acho que ganharei uma bola quadrada”), mas sim de modo rotineiro e sistemático.

**A.:** No entanto, você acredita que é logicamente possível acreditar no impossível<sup>41</sup>; é (aparentemente) logicamente possível que você esteja errado, e que proposições podem ter um *continuum* de valores de verdade – que, por exemplo, “não haverá batalha naval amanhã” é verificado por 95% dos valores das variáveis (neste caso, possíveis estados de coisas para amanhã). Na verdade, há motivos para acreditar que nossos procedimentos inferenciais básicos são *bayesianos* (Oaksford e Chater, 2009); e estatísticos e gestores de risco falam assim: “há apenas 5% de chance de que nossa perda ultrapasse nosso *Value-at-Risk*”, ou “com um nível de confiança de 95%, nosso resultado estará dentro da nossa margem de erro de 2%”. Talvez você devesse fazer o mesmo.

**T.:** Mas não; falo português, como todo mundo. Além disso, ainda poderíamos perguntar a essas pessoas se as previsões delas são verdadeiras, se acreditam nelas. Afinal, se disserem “sim”, a bivalência retorna pela porta dos fundos, uma vez que “é provável que haverá uma batalha amanhã” é verdadeira agora, ou falso agora - embora o valor de “haverá uma batalha...” possa mudar com o passar do tempo e com a atualização de sua probabilidade posterior. Se é isso que quero dizer com “haverá uma batalha naval amanhã”, então podemos concluir que o que eu digo tem um valor de verdade bem definido agora.

---

<sup>41</sup> Sorensen (2001, 124-5) apresenta um “argumento transcendental” (supostamente) provando que podemos acreditar em inconsistências - mas isso implica que não saiba que essas crenças são inconsistentes. No entanto, tais crenças “impossíveis” são excepcionais (embora inevitáveis: a razão exige que as tenhamos), tais como as crenças sobre a imprecisão (“não há nenhum último 'minuto do noon’”). Poderíamos dizer o mesmo sobre futuros contingentes?

**A.:** Ok, mas o nosso valor de meta-verdade "verdade agora" pode estar em um *continuum* também. Pode-se dizer "há 99% de chances de haver 95% de chances de que...". Meta-analistas fazem isso o tempo todo.

**T.:** Sem problemas; ficarei satisfeito com a meta-meta-afirmação "é verdade que é provável que seja provável que... ". Se chegarmos ao final disso, é a noção de contingência que pode desaparecer, não a verdade.

**A.:** Espere. Isso é divertido, mas, se continuarmos nisso, vamos seguir *para sempre – novamente!* Nós já concordamos em evitar isso. Assim, permita-me recapitular. Se entendi bem o seu ponto, nós sabemos que não ficaremos nus, nem cometeremos suicídio...

**T.:** Não! Eu disse que sabia que  *você*  não ficaria nu, nem se suicidaria. Não falei nada sobre mim. Eu não sei nada disso; e não posso acreditar que  *você*  o saiba.

**A.:** Não entendo. Certamente, se  *você*  pensa isso sobre mim, também posso pensar sobre  *você* . E  *você*  disse "qualquer ser humano normal"... E  *você*  não teria motivos para fazer nada disso, para se matar, ou...

**T.:** ... ficar nu? Claro. Primeiro, não sou um ser humano, nem me considero normal – da minha perspectiva, sou um réptil bastante especial. E já estou nu: não consigo imaginar o que seria um quelônio de roupas – algo tão inconcebível quanto um leão que falasse. Quanto ao suicídio... bem, algum filósofo com talento para a escrita dirá, um dia, que essa é a verdadeira questão filosófica; aliás, com exceção de filósofos e doentes terminais, o suicídio de uma pessoa costuma ser uma surpresa – afinal, se ele fosse previsto, dificilmente os amigos do suicida o permitiriam. Além disso, seria uma maneira de provar que  *você*  não pode me alcançar, e um desfecho dramático para este diálogo.

**A.:** Mas  *você*  não está pensando realmente nisso. Eu sei que não vai se matar; até porque seria imoral, e  *você*  é uma tartaruga virtuosa.

**T.:** Talvez  *você*  tenha razão; mas, se a modéstia é uma virtude, também não posso afirmar que sou virtuoso (Sorensen, 1988, 9). Meu ponto é que eu não poderia acreditar nisso, até o momento da decisão; e depois de decidir, não poderia deixar de acreditar, sem rever a mesma decisão. Crer em "vou fazer x, mas não o decidi" seria como crer em "chove, mas não acredito que chove". Digamos que tenho uma visão a meu respeito que, embora privilegiada, tem alguns "pontos-cegos". Mas eu não tenho a pretensão de que isso valha para  *você* . Pode acreditar no que  *você*  decidir acreditar.

**A.:** Isso me faz pensar sobre nossa conversa em Paris (Hofstadter & Dennett, 1981, 430), quando você me fez imaginar como seria ler um livro que contivesse a descrição completa de meu cérebro, a fim de prever como eu agiria....

Se entendi bem, o problema que você aponta não é tanto lógico quanto *doxástico*, e trata menos da teoria da ação que da teoria da decisão. Diz respeito não ao estudo do valor de verdade das proposições futuras, mas àquilo em que precisamos acreditar para poder agir (nossas previsões) e àquilo em que não podemos acreditar (qual seja, o conteúdo da própria decisão, antes que ela ocorra). O agente precisa ser capaz de acreditar que os outros agirão de uma determinada maneira (ainda que seja possível que eles não ajam dessa forma), mas ele não pode crer que ele vai agir de uma determinada forma sem decidir isso. Claro, ele pode se sentir tentado a colapsar esse ponto-cego com indeterminação lógica (como faz McKay, 1960).

Fica uma dúvida: e quando a ação de outrem está relacionada à minha decisão? Eu tentarei prever o que o outro decidirá, e tomarei minha decisão com base nisso; mas ele estará fazendo o mesmo. Como eu estava tentando adivinhar onde você colocaria seus navios no tabuleiro – sabendo que você sabia que eu tentava adivinhá-lo... aliás, desculpe, qual foi sua última jogada, mesmo?

**T.:** A6. É por isso que inventarão a teoria dos jogos (Morgenstern & Von Neumann, 1953). Em alguns casos, isso não implicará nada paradoxal: assumindo que você é racional, e que eu também sou, sabemos que não valeria a pena dedicar meu tempo e energia a descobrir o conjunto de estratégias ótimas de distribuição de navios, razão pela qual os distribuí aleatoriamente; na realidade, é provável que seja esta a estratégia ótima de distribuição. Presumo que você haja feito o mesmo – até porque, se pensasse o contrário, seria melhor rever minha estratégia... será que eu sabia disso? Depende, em parte, se isso é garantia ou justificativa suficiente para minha conclusão.

Acho que esse é o verdadeiro problema por trás do argumento da batalha naval: imagine que, em um mundo possível (vamos chamá-lo “Grécia -1”), dois generais (chamados de Sr. T e Sr. A) estão se perguntando se eles vão se envolver em uma batalha no dia seguinte, e ambos ignoram que a batalha será impossível, por causa de algum desastre, como um tsunami; não há indeterminação ontológica<sup>42</sup> neste caso (não se pode lutar uma batalha enquanto a terra está caindo aos pedaços), apenas indeterminação epistêmica e doxástica – ou seja, os generais não

---

<sup>42</sup>Se você acha que desastres são ontologicamente indeterminados até serem previstos, imagine que algum cientista ou profeta o previu.

sabem que não haverá batalha. Isso sugere que qualquer alegada indeterminação, neste cenário, vem dos tomadores de decisão, não do ambiente.

Em outro mundo possível (“Grécia-2”) acessível a partir desse, mas sem desastres, o Sr. A pensa “De jeito nenhum! Não haverá batalha, tenho certeza, já que estou desertando para o lado do Sr. T”. Mas nesse mundo, também não há mais indeterminação, pelo menos para A; ele não pode deixar de acreditar que não haverá batalha, uma vez que decidiu sobre isso (e “quando um não quer, dois não brigam”). Engraçado, ele não poderia nem acreditar em sua deserção, até o ter decidido; na realidade, ele ainda pode mudar de ideia (o que significa que ainda é possível que a batalha aconteça), mas ele não pode acreditar que fará isso, enquanto estiver decidido a desertar.

No entanto, até que T esteja ciente desse fato, ele ainda pode acreditar na indeterminação da batalha. E essa é a questão real: ambos os generais consideravam a batalha indeterminada porque nenhum deles conhecia sua própria decisão (porque nenhum deles a havia tomado ainda), nem a do oponente (porque, novamente, nenhum deles havia feito isso ainda); é preciso saber a que se está respondendo para dar a melhor resposta. Imagine que, em “Grécia-3”, os dois generais tivessem conhecimento comum de seu compromisso com a batalha, ou mesmo, em “Grécia-4”, com a paz; então não haveria mais dúvidas.

**A.:** É difícil imaginar como, a partir da ignorância comum, esse conhecimento comum poderia ser alcançado em uma série finita e discreta de etapas; Gen. T envia um sinal 1 de que ele quer lutar, então Gen. A recebe e responde com um sinal 2 de que ele quer lutar também, que é, depois, recebido por T – que responde, com o sinal 3, que ele recebeu 2 e ainda vai lutar, etc. Isso pode parecer implausível à primeira vista – mas então me lembra todas as vezes em que tive problemas para agendar um almoço com amigos ou para marcar um encontro. Acho que isso é o que eles chamarão de "Problema dos Generais Bizantinos" e de "Jogo do E-mail" (Binmore, 2007, 170-171). É pior do que tentar correr com você.

**T.:** Certo, e apesar disso, geralmente nos encontramos no momento e no lugar corretos – e ficará mais fácil quando alguma tecnologia nos oferecer a possibilidade de enviar e receber sinais confiáveis simultaneamente. Não poderíamos nos encontrar se houvesse uma dúvida real – se nenhum de nós acreditasse que a reunião aconteceria; podemos mudar de ideia, é claro, mas com isso também mudaremos nossas crenças. E podemos estar enganados; mas isso é a ignorância ordinária, e nós não precisamos de indeterminação para explicar isso.

O enigma ontológico é: como a indeterminação ontológica poderia ser afetada pela tomada de decisão unilateral? Como isso poderia ser afetado, na verdade, apenas por uma tomada de decisão, não importa de quantos lados? A batalha ainda é possível, mas A não pode ver isso como indeterminado; e isso está determinado para T, na Grécia-2, apenas por causa de sua ignorância.

Mesmo que haja uma explicação para essa indeterminação ontológica, ela não tem o apelo e a relevância prática que parecia ter quando consideramos pela primeira vez o argumento da batalha naval; agora, o que realmente deve ser explicado é a relação entre as crenças de um agente (incluindo suas crenças sobre si mesmo) e suas próprias decisões.

**A.:** ...A6, certo? Água. Essa discussão me lembrou o filho de um conhecido, que para vingar a morte do pai, matou a própria mãe, porque o Oráculo assim vaticinara (Ésquilo, 1986). Não lembro bem quando ocorreu. Mas, no momento do golpe, o rapaz hesita e pergunta a um cúmplice: “devo matar minha mãe?”. Então o cúmplice lhe recorda sua obrigação para com Apolo e seu juramento.

Ele não deixou de acreditar que o que Oráculo enuncia é verdadeiro – mesmo assim, ele teve dúvidas no momento de decidir. Do contrário, como poderíamos chamar isso de decisão (Williams, 1993, 138)? Isso vindica o objetivo principal de Barbosa Filho (2004): a incompatibilidade entre saber e fazer – ao menos antes da decisão. E3.

**T.:** Posso me solidarizar com esse rapaz. Mesmo que eu fosse apenas um personagem de uma história, não poderia acreditar a sério que posso decidir fazer p sem acreditar que farei p, e (via de regra) vice-versa; nem mesmo o autor da história poderia me fazer crer nisso. Mas nesse caso, alguém poderia dizer que eu não poderia acreditar em coisa alguma; que não haveria crenças, pois não haveria 'eu' algum para acreditar. Mas, se fosse esse o caso, também não poderia crer nisso: seria como acreditar na sentença “não existo” – que não é nenhum absurdo lógico, até porque ela será verdadeira, em algum momento, para todo mortal.

Aliás, água!

## **Capítulo 4: A Escolha de Greenspan e o Elogio da Passividade – considerações sobre decisões da autoridade em contextos de grande incerteza**

Financial markets are the machines in which much of human welfare is decided; yet we know more about how our car engines work than about how our global financial system functions. We lurch from crisis to crisis. In a networked world, mayhem in one market spreads instantaneously to all others—and we have only the vaguest of notions how this happens, or how to regulate it. So limited is our knowledge that we resort, not to science, but to shamans. We place control of the world’s largest economy in the hands of a few elderly men, the central bankers. We do not understand what they do or how, but we have blind faith that they can somehow induce the economic spirits to bring us financial sunshine and rain, and save us from financial frost and pestilence. (Mandelbrot & Hudson, 2008, 184)

**Palavras-chave:** Autoridade, Incerteza, Teoria da decisão racional

### **Introdução**

Na primeira seção, apresentamos o argumento do “elogio da passividade” de Huemer (2012), aplicável a situações em que ignoramos as consequências da ação – em especial, no domínio das ciências sociais, que careceriam de poder preditivo. Afinal, como a ação tende a um fim, a decisão de agir é irracional se não podemos crer, justificadamente, que podemos obter os fins visados – ou auto-derrotável, se agir de acordo com a decisão nos impede de chegar a esse fim.

Em política, essa objeção serve para libertários e democratas atacarem a autoridade baseada em *expertise*; na terceira seção, mostramos como esse argumento modela a explicação de A. Greenspan, ex-presidente do FED, sobre a omissão quanto a adotar medidas para evitar a bolha imobiliária e prevenir a crise financeira de 2007: dada a incerteza, autoridades regulatórias e macroeconômicas não estavam justificadas em intervir – mormente porque os agentes econômicos conheceriam melhor seus próprios riscos. Vemos como, malgrado a canônica separação entre fato e valor, o argumento abala a suposta “divisão de trabalho” entre teorias morais e teorias da decisão – pois aquelas serão inaplicáveis se não esclarecerem como



o agente pode/deve formar juízos racionais sobre as consequências (prováveis/futuras) das ações prescritas.

Todavia, no limite, o elogio da passividade levaria à impossibilidade da ação racional *simpliciter*, porquanto impossível agir sem ter expectativas sobre o futuro; nem poderia servir para favorecer teorias particulares, como o libertarismo – pois não se pode mostrar que levaria ao estado de coisas desejado. O problema, arguimos, está na escolha de um princípio que permita estender a situações de incerteza e ignorância a teoria da decisão racional ao estilo de Savage (1951) – onde probabilidades subjetivas funcionam como a “lógica da crença” (quantificando a incerteza subjetiva), e preferências são ordenadas segundo uma função-utilidade. As alternativas de complemento são o princípio de razão insuficiente de Laplace (i.e., inexistindo informação, eventos complementares são tidos por equiprováveis) – que nos leva ao que chamamos de *Bayesianismo* (Binmore, 2009) – e a regra de *maximin* de Wald. Em filosofia política, a divergência entre essas duas opções corresponde à disputa entre as concepções de contrato social de Rawls (com o princípio da diferença) e de Harsanyi (utilitarismo).

Aduzimos que o Bayesianismo não responde bem (normativamente) em “mundos amplos” (*large worlds*), pois sua epistemologia tem “pontos-cegos” no que tange a catástrofes (Cooke, 1986), a eventos da primeira pessoa (Binmore, 2009) e a decisões sociais. Nessas condições, devemos evitar os piores resultados para os que estão em pior situação, favorecendo o princípio da diferença (Shrader-Frechette, 2014). Arguimos, na terceira seção, que isso fornece uma objeção contra o argumento de Greenspan: dada a possibilidade de bolhas, autoridades estão justificadas em incentivar os agentes a incorrerem em menos riscos; isso demonstra que restrições regulatórias e supervisão intensiva podem ser justificadas. Por fim, avaliamos se isso justificaria medidas monetárias, “mirando” ativos específicos; nesse caso, porém, apresentamos uma objeção distinta de Greenspan: que, dadas as implicações políticas dessa atuação, uma autoridade pode duvidar que *ela* seja legítima para uma decisão, ainda que tal decisão seja justificada.

## 1 Ceticismo e Passividade

No *Elogio da Passividade*, Huemer (2012, 12) evoca a morte de G. Washington: os médicos, desconhecendo como tratar a doença (em razão do atraso científico da época), apenas aumentaram sua agonia com tratamentos ineficazes (como sangria). O exemplo fornece uma

lição para a teoria da ação: ao agir sob incerteza – por desconhecer qual a melhor coisa a fazer, ou quais as consequências da ação – é improvável atingir os fins desejados, e é possível atingir fins indesejados. Isso compromete a própria racionalidade da ação<sup>43</sup> – afinal, como Aristóteles (1966) apregoa ao começar a *Ética a Nicômaco*, “a ação tende a um fim” (EN, I, 1); como poderíamos, pois, agir racionalmente sem acreditar que tal fim será obtido<sup>44</sup>?

Essa *objeção cética* serve de base à *conclusão pela passividade* de Huemer (2012, 23-24) de que, frequentemente, em questões sociais, é melhor *não agir* do que *agir*; portanto, não devemos votar, tentar resolver problemas sociais complexos, nem “militar” e lutar por aquilo

---

<sup>43</sup> Cabe notar que estamos evitando usar o termo *intenção*, embora estejamos entrando em seu “domínio”; isso se explica em razão da alta controvérsia filosófica em torno desse conceito, que é psicológica e normativamente “carregado” (Knobe, 2003). No entanto, não sentimos falta dele: pode-se abordar o tema da escolha e decisão racional sem mencioná-lo – como o faz boa parte da literatura em economia e em filosofia política.

<sup>44</sup> De fato, há filósofos que, confrontados com esta questão, e diante do caráter paradoxal do argumento que estamos apresentando, simplesmente abdicam da ideia de que deve existir uma conexão entre intenção e cognição – a exemplo de Holton (2011, x): “we need a capacity to make choices that is independent of our judgements. Even if we can make judgements about what is best, they are often corrupted by ensuing temptation”; outros, mesmo mantendo essa conexão (o “cognitivismo forte”), adotam uma postura pragmática e abdicam da conexão entre juízo e *evidência* – como em Marušić (2013) e em Marušić & Schwenkler (2017). Mesmo nessas teorias, porém, essa abdição ocorre não em situações ordinárias, mas como *exceções*, decorrentes de nossa racionalidade limitada; nosso argumento ao longo do texto apela para limites afins como uma restrição ao escopo do *Bayesianismo* – a teoria normativa padrão para decisões racionais.

Podemos conceber que um agente tome uma decisão sem crer (justificadamente) que o fim tende a ser atingido, mas é difícil dizer que essa é uma *boa* decisão, preocupada com a consistência entre crenças (representações de estados de coisas) e valores (objetivos, ou preferências sobre estados de coisas); e, ao fim e ao cabo, nossa investigação aqui é sobre como devemos decidir (teoria *normativa*). É claro que se pode destacar (como o faremos) que seres humanos não são agentes racionais ideais, e que há outras restrições (*constraints*) à decisão além da consistência entre meios e fins: talvez eu possa fazer a você uma promessa sincera mesmo sem ter razões para crer que a cumprirei, na esperança de que a própria promessa crie uma relação especial que me motive a cumpri-la; mas, neste caso, o fim de minha ação inclui estabelecer essa relação especial, e minha decisão pode ser vista como afim a uma aposta, em que o ganho esperado “compensa” o fato de que o evento não tem uma probabilidade tão alta (ou seja, nesse caso, ela não viola o Bayesianismo). No entanto, mesmo casos assim só podem ser marginais: a prática social da promessa só existe porque, nos seus casos paradigmáticos, os agentes atribuem uma alta probabilidade ao seu cumprimento; *confiar* que a promessa será cumprida (na realidade, confiança, de forma geral) é um “estado que é essencialmente colateral” (Elster, 2016, 44 e s.).

Uma postura com o mesmo “espírito” dessa abdição, mas que mantém a conexão *bayesianista* entre intenção, juízo e evidência, é adotar uma teoria *evidencial* da decisão, como em Jeffrey (1965) – na qual a disposição do sujeito em relação a uma ação *conta como evidência* (no sentido de que aumenta a probabilidade subjetiva) sobre um estado de coisas futuro (como, e.g., de que forma eu ou outros agentes decidirão); nessa linha, a melhor decisão é aquela com a maior *expectativa* (i.e., utilidade esperada), condicionada ao fato de que tal decisão foi tomada, sem exigir-se uma conexão causal entre a decisão e o resultado. Enquanto na teoria causal o agente escolhe a opção que “tende a causar” a maior utilidade esperada, na evidencial ele decide de maneira a se inserir na classe com o melhor resultado.

A principal crítica a essa teoria é que admite correlações espúrias (Soares & Levinstein, 2017): é circular pensar que minha decisão de fazer *x* conta como uma razão para fazer *x*; e pensar que é uma razão para considerar que “*x* é bom”, ou que outros agentes farão *x* (como na discussão do “dilema do prisioneiro gêmeo”) parece *wishful thinking* – é o raciocínio da raposa de Ésope, que conclui que as uvas inatingíveis estão amargas (Binmore, 2009). No entanto, se consideramos que decisões são fatos que ocorrem *dentro do mundo*, então segue-se que esse fato pode ser um *input* relevante na decisão do agente (o problema da decisão auto-referente que abordaremos *en passant* diversas vezes ao longo do texto). Restringir a teoria evidencial a correlações relevantes é o objetivo a teoria “funcional” (*functional decision theory*), pela qual “the normative principle for action is to treat one’s decision as the output of a fixed mathematical function that answers the question, ‘Which output of this very function would yield the best outcome?’” (Yudkowski & Soares, 2017, 2).

em que cremos<sup>45</sup>; isto porque, em questões sociais, frequentemente estamos em situação de ignorância – e as consequências das políticas adotadas podem ser o *contrário* do que intencionávamos<sup>46</sup>. Um exemplo é o programa americano *Scared Straight*, em que adolescentes infratores foram levados a conhecer penitenciárias a fim de os assustar e reabilitar; embora inicialmente considerado um sucesso, ele foi posteriormente contestado, uma vez que a diminuição da taxa de cometimento de infrações foi *maior* entre indivíduos que não participaram do programa do que entre os que participaram – i.e., um jovem tinha mais chance de infringir a lei no futuro se participasse do programa do que se não participasse (Petrosino et al., 2013)<sup>47</sup>.

O *elogio* parte de uma teoria “inocente” da decisão, em que não há comparação entre a incerteza dos estados possíveis e os valores em jogo; na teoria da decisão *standard*, como na teoria da utilidade esperada subjetiva de Savage (1951), associamos a estados de coisas uma distribuição de probabilidades subjetivas (medindo o grau de incerteza que temos em relação a esses cenários) e utilidades (medindo o quanto os valorizamos, uns em relação aos outros). Isso resolveria o problema: não preciso conhecer o futuro para agir racionalmente, apenas qual a “expectativa” da ação (uma função da probabilidade de atingir os fins e do valor atribuído a eles); mas não há nada de simples nessa resposta: o problema agora é como atribuir essas probabilidades (uma teoria do juízo) e como definir essa valoração, mormente quando inclui as

---

<sup>45</sup> Estou focando apenas uma parte do argumento do artigo, restrita às questões sociais sobre as quais não se poderia ter conhecimento. O principal alvo de Huemer não são, na realidade, os *experts*, mas justamente os cidadãos comuns, que, via de regra, são presa de viés confirmatório e ignoram a opinião especializada.

<sup>46</sup> O *Elogio da Passividade* se aproxima do que A. Hirschmann (2004) chama de “retórica reacionária” – argumentos no sentido de que reformas sociais tendem a *serperversas* (prejudicam valores já conquistados), fúteis (não têm efeito prático detectável) ou perigosas (produzem riscos não previstos). Hirschmann também reconhece que existe uma “retórica progressista” falaciosa: o argumento do “perigo iminente” (i.e., o risco de inatividade), de “ter a história do seu lado” (i.e., determinismo progressista) e a inconsequente necessidade de radicalização: “old order must be smashed and a new one rebuilt regardless of any counterproductive consequences that might ensue” (Hirschmann, 2004, 162).

Como um otimista e progressista, ele tende mais para este segundo lado, invocando argumentos como o da *mão escondida benevolente*: há casos em que planejadores sociais otimistas são “salvos” de sua ingenuidade pela intervenção não-antecipada de outrem, colhendo benefícios sociais inesperadamente altos. Contudo, Hirschmann negligencia que, para cada “mão benévola”, há pelo menos uma “mão malévola” punindo tomadores de decisão injustificadamente otimistas que se deixam seduzir por essa versão exagerada “falácia do planejamento” (Flyvbjerg & Sunstein, 2016).

Cabe deixar claro, porém, que nos distanciamos da “retórica reacionária”; afinal, se temos motivos para evitar otimismo com a “mão benévola”, também o temos em relação à “mão invisível” (a “ordem espontânea” de sistemas complexos). Nossa ênfase em evitar riscos deriva de um otimismo sutil, mas profundo: se pudermos evitar eventos negativos de grandes proporções, então, contrariando o aforisma keynesiano, a tendência é que, no longo prazo, estejamos muito melhor.

<sup>47</sup> Poucas áreas são mais sensíveis a esse argumento que o direito penal – que, via de regra, tem por objetivo a prevenção geral das condutas proibidas e dos riscos sociais correspondentes (Tadros, 2015); ainda assim, a evidência (quando existe) sobre a eficácia das medidas é particularmente controversa (Bun; Sarafidis; Kelaheer, 2016).

preferências de diferentes indivíduos. Outrossim, um agente racional precisaria ter uma ideia do *custo de oportunidade* dessa alternativa em relação a outras.

A seguir, exploraremos a justificação dessa objeção cética no contexto da teoria da decisão sob racionalidade limitada; arguiremos que, embora seja preciso formular princípios especiais para lidar com a incerteza, há outros candidatos que não a passividade. Confrontamos o *elogio da passividade* com a filosofia política normativa: embora a objeção cética ofereça um desafio para a teoria da autoridade, isso não implica a ilegitimidade desta em situações de incerteza; pelo contrário: aduzimos, a partir da análise do debate Harsanyi-Rawls, que a atuação da autoridade pode ser legítima para mitigar a incerteza e riscos de catástrofes. Na terceira seção, aplicamos o argumento ao estudar as alegações de Greenspan (2013) sobre a responsabilidade de autoridades regulatórias em relação à possibilidade de crises financeiras.

### 1.1 Incerteza e risco

É comum adotar a distinção de Knight (1921, 46) entre *risco* (eventos negativos com probabilidade mensurável de forma precisa – i.e., o “risco conhecido”) e *incerteza*<sup>48</sup> (o “risco desconhecido”). Incerteza é um termo vago e ambíguo; podemos vê-lo como oposto ao conceito de informação. Em geral, utiliza-se “incerteza” num sentido amplo para designar as situações de decisão em que não há conhecimento completo e perfeito; já na distinção de Knight, ela assume um sentido mais estrito, no qual não se tem uma definição adequada das probabilidades

---

<sup>48</sup> Também a expressão “aversão a risco” é ambígua:

- em teoria dos jogos, equivale à *concavidade* da função-utilidade (Binmore, 2009, 50); i.e., se sou indiferente entre 1 brigadeiro e 50% de chances de ganhar dois brigadeiros, então minha aversão a risco é zero. Doravante, usaremos a expressão *neste* sentido.

- em finanças, refere-se à volatilidade – i.e., a (raiz quadrada da) *variância* relativa ao retorno esperado de um portfólio (Mandelbrot & Hudson, 2008)); se um investimento tem retorno esperado de 1%, com volatilidade de 3%, então, *grosso modo*, se pode esperar receber entre \$98 e \$104. Trata-se de um fator importante, dado que o investidor tem um limite inferior para as perdas que pode suportar (i.e., seu capital) – o que explica a assimetria entre a aversão a risco enquanto perda (limitada ao capital) e enquanto ganhos.

Essa segunda noção oferece uma forma conservadora de modelar a “grave incerteza”, ainda que aplicando o princípio de Laplace – como *variância* desconhecida (e potencialmente ilimitada). Buchak (2013, 44) aborda essa possibilidade de modo particularmente interessante, aduzindo que a aversão a risco deveria ser um *terceiro* componente subjetivo da decisão, não redutível aos outros dois (probabilidade subjetiva e preferência) – donde a decisão mais racional é a com maior utilidade esperada subjetiva *ponderada pelo risco*; essa abordagem é coerente com a prática financeira.

Essa explicação da aversão hiperbólica à incerteza desempenha um papel importante na análise do argumento da posição original a seguir; no entanto, ela não é suficiente para justificar a adoção do *maximin* e do princípio da diferença (exceto como *proxy*), pois o princípio de utilidade esperada é facilmente adaptável para produzir uma análise de custo-benefício mesmo em decisões voláteis. Em tese, um investidor com a aversão a risco *finita* pode racionalmente escolher um portfólio com volatilidade superior a seu capital, se o retorno esperado for suficientemente alto (e.g., se maior do que o que espera obter antes de se aposentar, ponderado por sua taxa de desconto do futuro).

dos estados de coisas possíveis – o sentido usado pela teoria da decisão sob ignorância. Contudo, mesmo nesse caso, pode-se aumentar a incerteza, se o agente desconhece quais os estados de coisas possíveis, ou mesmo sua própria função utilidade. Por isso, não há uma taxonomia canônica dos diferentes graus de incerteza; mas pode ser interessante distinguir, conforme Lo & Mueller (2010):

- a) certeza completa: o passado é conhecido, e o futuro é determinado;
- b) risco estrito, sem incerteza: aleatoriedade governada por uma distribuição de probabilidades conhecida);
- c) incerteza redutível a risco: embora a distribuição seja desconhecida, ela permanece estável ao longo do tempo e, com dados suficientes, pode ser inferida;
- d) incerteza parcialmente redutível: os parâmetros da distribuição mudam ao longo do tempo, de forma que não podemos representá-los adequadamente num modelo (e.g., processos caóticos; jogos de informação incompleta redutíveis à informação completa);
- e) incerteza *irredutível*: há parâmetros desconhecidos, fora de nosso alcance.

A partir da noção *knightiana* de incerteza (não completamente redutível a risco), desenvolveu-se a teoria da decisão sob ignorância *completa* que trata de critérios para lidar com a falta dessa informação (Binmore, 2009, 154), como se se tratasse de um *jogo contra a Natureza*, “an unreasoning entity whose choice affects your pay-off, but which has no awareness of, or interest in, the outcome of the game” (Straffin, 2010, 56). Um desses critérios é a aplicação do “princípio de razão insuficiente” de Laplace – pelo qual, na ausência de outras informações, atribuímos aos estados de coisas possíveis a mesma probabilidade; trata-se de uma extensão natural da teoria da utilidade esperada, a qual podemos chamar de *bayesianismo* (Binmore, 2009, 154-55). Outro critério é o conservador *maximin*, pelo qual o agente reduz sua exposição ao risco, como se a “Natureza” estivesse contra ele (Binmore, 2005, 33) – selecionando a ação que maximiza o resultado positivo do pior cenário possível. Estes dois critérios<sup>49</sup> estão na base de duas teorias da justiça concorrentes: respectivamente, o utilitarismo de Harsanyi e o liberalismo de Rawls – como veremos na seção 2. Uma maneira de ilustrar a

---

<sup>49</sup>Embora estes sejam os dois principais critérios discutidos como extensão da teoria da decisão à ignorância completa (Giocoli, 2011), a literatura (Straffin, 2010, Cap. 10) também apresenta a revisão de dois critérios “intermediários” entre estes: o de Hurwicz (Arrow & Hurwicz, 1977) e a “minimização de arrependimento” proposta pelo próprio Savage (1951) – particularmente atraente quando se considera, a partir do “princípio reflexivo”, que o principal *desiderata* da decisão seria minimizar arrependimento posterior (Arntzenius, 2008). As discussões canônicas, com uma comparação formal entre os quatro critérios se encontra em Milnor (1954) e em Luce & Raiffa (1957). Estes quatro critérios são os únicos que satisfazem os *desiderata* axiomáticos de *simetria*, *dominância forte* e *linearidade*; contudo, a teoria da decisão sob racionalidade limitada apresenta outros critérios que não satisfazem plenamente esses axiomas.

divergência entre os dois critérios é o paradoxo de Ellsberg (1961), cuja formulação mais simples pode ser identificada em por Gintis:

Consider two urns. Urn A has 51 red balls and 49 white balls. Urn B also has 100 red and white balls, but the fraction of red balls is unknown. One ball is chosen from each urn but remains hidden from sight. Subjects are asked to choose in two situations. First, a subject can choose the ball from urn A or urn B, and if the ball is red, the subject wins \$10. In the second situation, the subject can choose the ball from urn A or urn B, and if the ball is white, the subject wins \$10. Many subjects choose the ball from urn A in both cases. This violates the expected utility principle no matter what probability the subject places on the probability  $p$  that the ball from urn B is white. For in the first situation, the payoff from choosing urn A is  $[0.51u(10) + 0.49u(0)]$  and the payoff from choosing urn B is  $[(1-p)u(10) + pu(0)]$ , so strictly preferring urn A means  $p > 0.49$ . In the second situation, the payoff from choosing urn A is  $[0.49u(10) + 0.51u(0)]$  and the payoff from choosing urn B is  $[pu(10) + (1-p)u(0)]$ , so strictly preferring urn A means  $p < 0.49$ . This shows that the expected utility principle does not hold.

[...]

The usual explanation of this behavior is that the subject knows the probabilities associated with the first urn, while the probabilities associated with the second urn are unknown, and hence there appears to be an added degree of risk associated with choosing from the second urn rather than the first. If decision makers are risk-averse and if they perceive that the second urn is considerably riskier than the first, they will prefer the first urn. (Gintis, 2009, 17-18)

O que chama atenção no paradoxo de Ellsberg é que, mesmo depois de notar a inconsistência, frequentemente, os sujeitos continuam apegados à decisão de ficar com a urna “segura”. Primeiro, eles não o veem como uma contradição: não consideram a informação de que, na urna B, a probabilidade de retirar uma bola de cor vermelha é *correlata à de retirar uma branca*; i.e., eles fazem as escolhas nas duas situações como se se tratassem de urnas completamente distintas. Afinal, se, na segunda situação, ao invés da urna B tivéssemos outra urna qualquer C (com um percentual desconhecido  $q$  de bolas vermelhas), não haveria propriamente contradição – mas sim um tratamento (aparentemente) arbitrário na distribuição de probabilidades subjetivas a dois conjuntos de informação idênticos (urna B e urna C).

Contudo, os sujeitos *não* tomam a decisão por meio da definição de uma distribuição de probabilidades, mas preferindo a urna sobre a qual têm maior informação. Binmore (2009, 89) equipara essa preferência pela urna A a “preferring to settle issues by tossing a coin known

to be fair, rather than tossing a coin about which nothing is known at all” – algo comum e aceitável. Todavia, de acordo com experimentos conduzidos por Binmore e Voorhoeve (2012), a aversão à ambiguidade diminui quando os sujeitos se asseguram de que os resultados são definidos por um processo aleatório – casos em que faria sentido adotar o princípio de Laplace.

### 1.1.1 Causas da incerteza: reflexividade, interações sociais e eficiência informacional

Uma das principais fontes de incerteza (no domínio social) é a existência de *outros* tomadores de decisão, tentando prever as decisões uns dos outros e reagir a elas. Isso insere um aspecto “auto-referencial” ou reflexivo na ciência social (Sandri, 2009), ensejando o que Harari (2015, 57) chama de “paradoxo do conhecimento histórico”, graças ao qual o conhecimento social rapidamente se torna obsoleto<sup>50</sup>.

Em microeconomia, algo afim se dá na discussão sobre *bolhas* financeiras – que deveriam ser impossíveis num mercado eficiente, em que no preço do ativo se inclui o valor de toda informação disponível, inclusive a informação sobre seu valor futuro; investidores racionais não comprariam, pois, o ativo inflacionado – já que aplicariam o raciocínio da indução retroativa (*backward induction*) e antecipariam sua desvalorização. Mas a dinâmica da bolha é a do “concurso de beleza” de Keynes, em que os participantes buscam adivinhar qual será o candidato mais votado, e sua estrutura é a de uma *race to the bottom*<sup>51</sup> (Talwalkar, 2014).

---

<sup>50</sup> Essa é uma possível explicação por que processos sociais não são, ao contrário de fenômenos da física clássica, *ergódicos* – processos em que as propriedades estatísticas de um “conjunto” de observações num determinado período equivale às propriedades estatísticas do sistema ao longo do tempo. A implicação é o lugar-comum de que “não há leis gerais” sobre fenômenos sociais (ao menos sobre os que são “não-ergódicos”). Segundo Peters (2011), desde Bernoulli, em 1738, a teoria da utilidade esperada (e com ela, boa parte dos economistas) pressupõe a ergodicidade – o que permitiria a economistas (a) modelar estados de coisas por uma distribuição normal e, usando o teorema do limite central, aplicar a *média* como medida de *tendência*, e (b) tratar a economia como uma ciência exata.

<sup>51</sup> Numa *race to the bottom*, os jogadores devem enunciar um número inteiro dentro dum intervalo finito; o vencedor é aquele que enunciar o número mais próximo do antecedente à média dos demais “lances” (arredondada para baixo) – o “resultado” do jogo. Por indução retroativa (“Se a média do grupo, com exceção de  $x$ , equivale ao número  $n$ , então  $x$  deve responder ‘ $n-1$ ’”), verifica-se que a solução (o equilíbrio perfeito sub-jogo) é todos convergirem para o menor número do intervalo; no entanto, em exemplos reais de *race to the bottom*, os resultados nunca coincidem com tal equilíbrio – *mormente* se a recompensa prometida for uma função do resultado do jogo. Imagine-se, e.g., que o intervalo é  $[0, 100]$  e que a recompensa corresponderá ao triplo do resultado, em dinheiro, dividido entre os vencedores; nesse caso, a indução retroativa implica a resposta “0” – o que implica remunerar o agente racional com nenhuma recompensa!

O ponto não é apenas o lugar-comum de que os indivíduos reais não se comportam como agentes racionais (conquanto tangenciemos o tema na seção sobre *racionalidade limitada*), nem que, em jogos iterados, pode ser racional não seguir o equilíbrio perfeito sub-jogo (dada o conhecimento comum da possibilidade de não o seguir) – mas também que um mercado competitivo pode selecionar indivíduos que assim se comportem.

Essa é a versão *forte* da “hipótese dos mercados eficientes” (HME), em que os investidores esgotam rapidamente todas as oportunidades possíveis de arbitragem – donde devemos tratar a oscilação dos preços como aleatória. Isso implicaria a impossibilidade de bolhas – donde Quiggin (2010) considera HME uma “ideia zumbi”, dada a grande quantidade de contra-exemplos. A ideia é que, mesmo que o mercado tenha indivíduos irracionais, seus erros aleatórios “compensam-se” mutuamente, de modo que a tendência do mercado é determinada por uma minoria de indivíduos racionais bem-informados (Sunstein, 2006, 106).

Como veremos na seção 1.3 a seguir, essa concepção de irracionalidade como “ruído” (*noise*) aleatório é falsa: há erros sistemáticos (vieses) que podem ser amplificados em algumas situações, como na ocorrência de *bolhas*; ainda que haja indivíduos racionais (ou que pensam que o são) reconhecendo que se trata de uma bolha, alguns deles, ao invés de apostar no seu estouro (por meio de derivativos), supõem que serão capazes de “comprar na baixa e vender na alta”, confiando na execução de ordens de *stop-loss* para se retirar *antes* da queda – na ignorância de que, frequentemente, o valor do ativo pode cair abruptamente e atingir um valor inferior ao da ordem antes da execução desta (Mandelbrot & Hudson, 2008, Cap. 1).

Em sua versão fraca – não contestada por Quiggin (2010, 37-8) – HME aduz que os preços refletem toda informação *pública passada disponível* e implica não a impossibilidade, mas a imprevisibilidade por um agente confiável (*reliable*). O problema é que essa versão é resistente a contra-exemplos; afinal, se alguém confiável previsse uma bolha, uma autoridade responsável a impediria. Sorensen (1988, 110), aduz que a HME é um tipo de *blindspot* – uma tese que, ainda que verdadeira, não pode estar contida em algumas atitudes proposicionais dos investidores: ela só é verdadeira na medida em que alguns agem sem acreditar nela, como investidores que gastam recursos buscando nova informação que posteriormente, a partir de seus investimentos, será agregada ao mercado (no mesmo sentido: Lomasky, 2011, 150)<sup>52</sup>.

---

<sup>52</sup> Embora HME seja invocada em argumentos contra interferências no mercado, é importante notar que nosso argumento a partir do princípio de *maximin* em favor da estabilidade **não** depende da falsidade da versão fraca de HME: mesmo que uma autoridade acredite na impossibilidade de prever o estouro de uma bolha num mercado eficiente, isso não é motivo para não o prevenir, mesmo que implique reduzir a eficiência do mercado – como o fez, recentemente, a SEC ao proibir o abuso da inclusão da expressão “*blockchain*” no nome de companhias. Mesmo nas condições em que um mercado é a melhor forma de agregar a informação dispersa na sociedade (como no argumento de Hayek, 1996, opondo o mercado de preços ao planejador central), ele pode não ser uma forma satisfatória de minimizar certos riscos. A ideia é que se devem distinguir dois níveis de análise, como proposto por Paul Samuelson: embora HME seja aplicável no âmbito *micro*, pode haver outros agentes com a tarefa de estimar a ocorrência de falhas e de intervir no âmbito *macro* (Quiggin, 2010, 49). Pode-se criticar esta posição como assaz suscetível a um “viés pessimista” e à tendência de superestimar eventos de baixa probabilidade; mas nosso argumento, na seção 2 a seguir, é que *ignorar* esse viés também é uma má política – ele desempenha um papel na construção da confiança necessária para a estabilidade e a justiça de nossas instituições.



Um problema análogo se aplica à macroeconomia, conforme Kydland & Prescott (1977, 473):

Even if there is an agreed upon, fixed social objective function and policymakers know the timing and magnitude of the effects of their actions, discretionary policy, namely, the selection of that decision which is best, given the current situation and a correct evaluation of the end-of-period position, does not result in the social objective function being maximized. The reason for this apparent paradox is that economic planning is not a game against nature but, rather, a game against rational economic agents.

A ideia não é *apenas* que existe um conflito entre políticas discricionárias (ótimas no curto prazo) e o resultado no longo prazo (como sugere Combe, 2017, 249)<sup>53</sup>, e sim que, dado que os agentes ajustam suas expectativas racionais às decisões da autoridade, então qualquer política consistente converge para um resultado sub-ótimo. Na base do argumento está o modelo de expectativas racionais, em que os agentes têm conhecimento comum de todas as variáveis relevantes; mas essa assunção é paradoxal, pois elimina a incerteza em relação à decisão dos demais. Como no debate sobre bolhas à luz da HME, a premissa do conhecimento comum gera um paradoxo da *indução retroativa*, a exemplado *jogo do centípede* e do dilema do prisioneiro reiterado um número determinado de vezes (Gintis, 2009; cf. Sorensen, 1999, 279) – onde mesmo agentes racionais podem desviar da estratégia dominante do que seria o “equilíbrio perfeito sub-jogo”, dada a probabilidade positiva de que os demais (fazendo o mesmo raciocínio) também desviem. Em filosofia, o exemplo clássico de paradoxo da indução retroativa é o *teste surpresa*, ora descrito:

A teacher announces that there will be a surprise test next week. A student objects that this is impossible: ‘The class meets on Monday, Wednesday, and Friday. If the test is given on Friday, then on Thursday I would be able to predict that the test is on Friday. It would not be a surprise. Can the test be given on Wednesday? No, because on Tuesday I would know that the test will not be on Friday (thanks to the previous reasoning) and know that the test was not on Monday (thanks to memory). Therefore, on Tuesday I could foresee that the test will be on Wednesday. A test on Wednesday would not be a surprise. Could the surprise test be on Monday? On Sunday, the previous two eliminations would be available to me. Consequently, I would know that the test must be on Monday. So a Monday test would also fail to be a surprise. Therefore, it is impossible for there to be a surprise test’” (Sorensen, 2014).

---

<sup>53</sup> Donde o título do artigo: é preferível adotar regras (compromissos críveis, “atando” as mãos da autoridade) a adotar políticas discricionárias.

Considerando o argumento do aluno, o professor poderá aplicar o teste na quarta-feira seguinte, surpreendendo aquele – portanto, a previsão “haverá um teste surpresa” é verdadeira (logo, não é contraditória). Contudo, essa proposição não pode ser objeto de *conhecimento comum* entre o aluno e o professor (Gintis, 2010, 171); aplicando o mesmo raciocínio à economia, pode haver *blindspots* para o conhecimento comum das expectativas racionais – o problema é que geralmente não é sábio contar com isso.

### 1.2. Eventos catastróficos e falhas

Mesmo defensores da eficiência informacional do mercado (a ideia de que os preços captam toda a informação relevante sobre os respectivos custos de oportunidade) tendem a admitir que ela só se aplica a mercados “grossos” (*thick*), com “homogenous products, large numbers of buyers and sellers who regularly engage in repeated transactions, transparent pricing and, ideally, forward markets for purchase or delivery at future dates” (Quiggin, 2018 - Cap. 11, p. 3). Essa restrição, portanto, excluiria mercados:

- a) imobiliários, com preços rígidos e ativos insuscetíveis de venda a descoberto (Millan, 1977);
- b) de previsões sobre ataques terroristas – além de ser moralmente repugnante (Sandel, 2012, 149), é de difícil precificação e suscetível a manipulação (Posner, 2006, 175);
- c) de cuidados de saúde – sujeito a falhas porque os consumidores estão expostos a riscos que não são totalmente seguráveis, e porque a avaliação de riscos em tal mercado demanda informação especializada, difícil de obter:

the market failure to insure against uncertainties has created many social institutions in which the usual assumptions of the market are somewhat contradicted. [...] The logic and limitations of ideal competitive behaviour under uncertainty force us to recognize the incomplete description of reality supplied by the impersonal price system (Arrow, 1963, 149).

Os exemplos (b) e (c) sugerem uma dificuldade para a agregação de informações sobre eventos graves. No limite, Cooke (1985) aponta que, além de eventos da primeira pessoa (qual a probabilidade do evento “decidirei comer brigadeiro?”), eventos *catastróficos*<sup>54</sup> seriam um desafio ao *bayesianismo*, pois não teriam valor mensurável.

---

<sup>54</sup> Cooke (1985) usa como exemplo de evento catastrófico desastres que extinguissem a vida – o que implica um estado de coisas sem valor mensurável; isso é diferente de um estado de coisas com “utilidade zero”; já Yudkowsky (2008) trata de eventos de impacto global (como a morte de milhões de pessoas). Isso não corresponde ao sentido *relativo* de catástrofe que ora utilizamos, que inclui eventos de impacto dentro de uma atividade. Contudo, também não corresponde ao sentido mais usual de catástrofe usado por Shrader-Frechette (2014), que inclui incidentes que impliquem a morte ou invalidez de um grupo grande de pessoas.

Apesar disso, Posner (2006, 141) avalia o custo de extinção da espécie humana em U\$600 trilhões, considerando o valor de uma vida estatística em cerca de U\$500 mil e aplicando uma taxa de desconto correspondente à taxa de juros anual. O cálculo do valor de uma “vida estatística” reflete o ponto desta sub-seção: trata-se de uma extrapolação a partir do que as pessoas se dispõem a pagar/receber para mitigar, assumir ou se assegurar contra risco de morte (Sunstein, 2013). O problema não é que isso seja repugnante (como argúi Sandel, 2012, 131 e s.), mas que, via de regra, *não faria sentido* dizer que, para um determinado agente, sua vida vale um determinado valor (quantificável numa unidade monetária ou *commodity* qualquer); logo, a propensão do agente em apostar não fornece uma medida consistente da avaliação que ele faz desse evento<sup>55</sup> - ele é (ao menos análogo a) uma catástrofe. Isso pode ser ilustrado pelo “paradoxo da Roleta Russa”; pessoas normais pagariam mais para, num jogo de roleta russa, reduzir a munição de um revólver (com seis cartuchos) de um cartucho para zero do que de dois para um (Binmore, 2007, 112 e s.); no entanto, em ambos os casos, tem-se o mesmo decréscimo de 1/6 no risco de morte. Aparentemente, isso parece abrangido por uma aversão hiperbólica a

---

É difícil encontrar uma definição precisa do que esse termo significa; ele inclui tanto a ideia de um estado de coisas difícil de valorar (de altíssimo impacto – em que a perspectiva de recuperação é nula, ou improvável) quanto de estimar (grande incerteza); isso é um problema teórico para a teoria da utilidade esperada subjetiva, já que esta associa a probabilidade subjetiva atribuída a um evento com a propensão do agente a apostar nele (Ramsey, 1931). Nossa noção de catástrofe aqui é, pois, contextual: dentro da prática de um mercado financeiro nacional, o cenário em que o valor de mercado de títulos do governo atinge zero é um evento catastrófico. Para a maioria dos investidores, o problema não é que tal cenário seja uma impossibilidade lógica (afinal, governos dão calotes, e às vezes deixam de existir), mas que não faz sentido apostar nem comprar seguro contra eles *dentro dessa mesma prática*. Outrossim, o resultado de recessões é suficientemente ruim, já que se estima que a perda do emprego implica, para o desempregado, um aumento de risco de morte de 63% no ano seguinte ao evento (Roelfs, Shor, Davidson & Schwartz, 2011).

Podemos exemplificar o tom “catastrófico” com que a Crise é descrita:

“[Northern Rock] was the first symptom of the global crisis, which reached the next level with the very similar collapse of Bear Stearns in March 2008, followed by the crash that really did take the entire global financial system to the brink, the implosion of Lehman Brothers on 15 September. Because Lehman was a clearing house and repository for many thousands of financial instruments from around the system, suddenly nobody knew who owed what to whom, who was exposed to what risk, and therefore which institutions were likely to go next. And that is when the global supply of credit dried up. I spoke to bankers at the time who said that what happened was supposed to be impossible, it was like the tide going out everywhere on Earth simultaneously. People had lived through crises before – the sudden crash of October 1987, the emerging markets crises and the Russian crisis of the 1990s, the dotcom bubble – but what happened in those cases was that capital fled from one place to another. No one had ever lived through, and no one thought possible, a situation where all the credit simultaneously disappeared from everywhere and the entire system teetered on the brink. The first weekend of October 2008 was a point when people at the top of the global financial system genuinely thought, in the words of George W. Bush, ‘This sucker could go down.’ RBS, at one point the biggest bank in the world according to the size of its balance sheet, was within hours of collapsing. And by collapsing I mean cashpoint machines would have stopped working, and insolvencies would have spread from RBS to other banks – and no one alive knows what that would have looked like or how it would have ended” (Lanchester, 2018).

<sup>55</sup> Uma possível explicação para essas limitações é que a prática de atribuir probabilidades *subjetivas*, interpretadas como “disposição em apostar”, a eventos existe para tomarmos decisões; fora desse contexto, não há nenhum problema em pensar em outra forma de calcular probabilidades. Nada impede que se descubra que estrelas como o Sol têm uma probabilidade *objetiva* de explodir a cada  $n$  anos, nem que, a partir de uma série de dados históricos, a probabilidade *objetiva* de que comerei um brigadeiro hoje é  $q$ .

risco, semelhante ao paradoxo de Allais – exceto quando pensamos em qual seria o valor, em dinheiro ou utilidade, de *reduzir a munição de seis para cinco*. Deveria ser a mesma quantidade de dinheiro que estaria disposto a receber para aceitar *aumentar* a quantidade de munição do revólver de cinco para seis (i.e., morte certa)?

Um segundo problema de eventos desse tipo é que é difícil estimar sua probabilidade de ocorrência a partir de informações passadas, pois são sujeitos a alguma forma de “viés antrópico”<sup>56</sup>: como riscos são estimados por estatísticas baseadas na sua frequência observada, “frequências de catástrofes que destroem ou são incompatíveis com a existência de observadores são sistematicamente subestimadas” (Ćirković, Sandber & Bostrom, 2010 – trad. livre). Não devemos esperar encontrar, num mercado informacional eficiente, uma estimativa adequada sobre a *derrocada* desse mesmo mercado.

### 1.3 O modelo do agente racional

Uma estratégia comum para escapar à discussão do elogio da passividade é objetar que pressupõe um modelo de agente racional irrealista. Afinal, é trivial que qualquer agente fisicamente possível tem capacidade de processamento limitada – e essa limitação leva a um comportamento “satisficente” (*satisficing*)<sup>57</sup>, e não estritamente *otimizador*; em que, ao invés de **um** processo algorítmico de otimização, a racionalidade humana tem também um caráter procedimental (Barros, 2010). Na verdade, as formas de pensar incluem diferentes processos de juízo (cognição) e decisão (valoração) operando em paralelo – alguns dos quais inconscientes. Frequentemente, quando operam em contextos distintos daquele no qual evoluíram como uma solução ótima, esses processos implicam vieses e potenciais falhas (Kahneman, 2012); um exemplo dramático: pilotos precisam ser treinados para resistir ao instinto natural de, no caso de piques, empinar o avião para afastar-se do solo – o que implica

---

<sup>56</sup> O nome vem do “princípio antrópico” – pelo qual não devemos nos surpreender ao observar no universo condições compatíveis com a vida inteligente; a idéia geral é que a mudança da posição do observador afeta o julgamento *mesmo quando não há alteração na informação*. Bostrom (2007) nota que isso se estende até mesmo ao problema da *Bela Adormecida*: Bela Adormecida aceita participar de um experimento no qual: a) se lança uma moeda imparcial no domingo, b) ela toma um sonífero amnésico, c) é despertada no dia seguinte e d) se, e somente se, o resultado foi *coroa*, é despertada de novo na terça-feira. A cada vez que é acordada, perguntam-lhe qual a probabilidade (subjéctiva) de que o resultado da moeda haja sido *cara*.

É evidente que, tanto no início (domingo) quanto no fim (quarta-feira) do experimento, ela deveria responder 50%; no entanto, ao ser acordada durante o experimento, ela sabe que há três possibilidades: ou é segunda-feira, e a moeda resultou em *cara*, ou é segunda-feira e resultou em *coroa*, ou é terça-feira e resultou em *coroa*. Logo, *ao ser perguntada*, ela deve responder que a probabilidade é de 1/3 (Elga, 2000).

<sup>57</sup> Mesmo soando mal, essa tradução para *satisficing* é utilizada, e.g., em Pedroso (2014, 63); a vantagem é que é fiel à tentativa de associar a noção de *satisfatório* à de *suficiente*.

reduzir a velocidade do ar, fazendo o aparelho perder altitude mais rapidamente (Lo, 2013, 631).

Nossas capacidades cognitivas (como as envolvidas em aprender com o passado e representar cenários futuros) e nossas emoções e sentimentos como medo, dor, prazer, ganância não foram desenhadas como um sistema consistente – na realidade, sequer evoluíram sincronizadamente. Algo assim ocorre quando falamos em “fraqueza de vontade” - quando Ahab (em *Moby Dick*) pede que o amarrem para lhe amputar a perna (como Ulisses se amarra ao mastro para não seguir as sereias), eleo faz porque *sabe* que suas preferências atuais em relação a seus estados futuros conflitam com as preferências que terá assim que começar a cirurgia (Schelling, 1983).

Algo análogo pode ocorrer com a cognição: quando enxergamos um triângulo de Penrose, nossa percepção visual conflita com nosso conhecimento sobre geometria euclidiana; Sorensen (2001, 38) caracteriza esse conjunto de processos distintos como “specialized homunculi who work like idiot savants”. Mas, se esses “homúnculos” não podem ser “submetidos” a um único sistema que dá a “palavra final” (que funcionaria de maneira análoga a um “ditador” da teoria da escolha social), então a situação é afim à de uma sociedade com votantes com preferências distintas (Minsky, 1986; Paglieri, 2016; Dennett, 2018), donde a observação de inconsistências pode ser vista como uma consequência (quase) trivial do teorema da impossibilidade de Arrow (1950)<sup>58</sup>.

---

<sup>58</sup> O mesmo pode ocorrer numa teoria mais próxima do senso comum e do estudo filosófico da razão prática se considerarmos que as preferências do agente são definidas por razões incomensuráveis – como em Dietrich & List (2012; 2013). Infelizmente, os problemas formais identificados pela teoria da escolha social *não* se restringem à política: diferentes tarefas *classificatórias* são formalmente isomórficas a uma função de escolha social arroviada, compartilhando do mesmo problema (v., e.g., Recamonde-Mendoza, 2016).

Uma explicação alternativa (e pouco explorada na literatura) para as inconsistências observadas é que possam ser formas de randomização, ainda que o agente não o saiba; Binmore fornece um exemplo de concorrência empresarial:

“It is natural to object that only crazy people make serious decisions at random, but mixed strategies are used all the time without anyone realizing it. [...]”

Game theory predicts that such a company will use a mixed strategy in the pricing game it has to play when the demand for vacations proves to be unexpectedly low. However, when I asked a senior executive whether his company actively randomized their prices last year, he reacted with horror at such an outlandish suggestion. So why were his prices for similar vacations so very different? His answer was instructive: ‘You have to keep the opposition guessing’ (Binmore, 2007a, 23).

Pode-se dizer que um tomador de decisão que buscasse implementar uma “estratégia mista” (na expectativa de que os demais jogadores também o fizessem) estaria substituindo a incerteza da interação estratégica vinculando sua própria decisão a um mecanismo de randomização – o que, salvo exceções (como sorteios – Dennett, 1986), não costuma ser socialmente aceito: mesmo quando se reconhecem os “limites da razão”, exige-se que escolhas sociais sejam feitas por um procedimento não-aleatório (contra tal exigência: Elster, 1988).

Outrossim, o modelo não consegue captar adequadamente a incompletude/incerteza lógica, nossa ignorância sobre estados de coisas possíveis, que inclui o que Donald Rumsfeld chamou de “desconhecidos desconhecidos” – já que, na epistemologia *bayesianista*, “desconheço que desconheço que  $p$ ” implica “conheço que  $p$ ”<sup>59</sup> (Binmore, 2009, 139). Essa incompletude permite nossa *modéstia* lógica, nosso potencial para inconsistência – como a capacidade de Sorensen (2001, 125-6) afirmar que tem crenças inconsistentes<sup>60</sup>.

O modelo do agente racional ainda pode ser útil – se não como “ponto de partida”, ao menos como “ponto de chegada” (como uma teoria normativa da qual buscamos nos aproximar). Há confirmações empíricas de que, em condições usuais, nos aproximamos dele (no sentido de que os sujeitos se comportam como se buscassem maximizar uma função-utilidade), e os desvios podem ser explicados como erros de performance (que os sujeitos corrigiriam após reflexão), ou acomodados em extensões da teoria (Gintis, 2009, 23). Ainda, pode-se defender que *tentar* atribuir probabilidades de forma precisa e consistente, mesmo quando elas não estão disponíveis, é uma demanda de *acurácia* (Schoenfield, 2015) da racionalidade; em especial, o viés de aversão à ambiguidade tende a diminuir quando se supõe que a distribuição de probabilidades decorre de um processo aleatório (Binmore et al., 2012). Isso leva a resultados superiores na atividade de efetuar estimativas – como revela a emergência de “super-preditores” (*superforecasters*) nas competições sobre previsões de Tetlock & Gardner (2017)<sup>61</sup> – e pode ser a melhor maneira de veicular a respectiva informação entre diferentes agentes, de uma maneira suscetível de ser agregada (e.g., por um mercado de informações).

---

<sup>59</sup> Isso deriva trivialmente do corolário de que o agente conhece todo o espaço lógico. Mesmo numa epistemologia logicamente mais fraca, que incorpore apenas os axiomas de *fatividade* ( $Kp \rightarrow p$ ) e *reflexividade* ( $Kp \rightarrow KKp$ ), a interpretação usual de “desconhecido desconhecido” é trivialmente falsa: um agente não pode enganar-se sobre se uma determinada proposição  $p$  pertence ou não a seu sistema de crenças – é impossível que, dentre “ $Kp$ ” e “ $\sim Kp$ ”, ele escolha a alternativa falsa.

<sup>60</sup> Assumindo que erros na atribuição de consistência correspondem a uma inconsistência (i.e., que se  $p$  é uma proposição consistente, então “ $p$  é inconsistente” é uma proposição inconsistente), então pode-se fornecer um argumento análogo ao do paradoxo do mentiroso: seja  $p =$  “é impossível ter crenças inconsistentes”; logo, se Sorensen (2001, 125) acredita em “ $\sim p$ ”, ele acredita numa crença inconsistente – mas isso implica que  $p$  não pode ser verdadeiro.

Por que o sistema de crenças de Sorensen não colapsa? Afinal, pode-se derivar qualquer coisa de uma contradição. A explicação pode apontar, novamente, para a incompletude desse sistema: ou a inconsistência é indiscriminável e restrita a um local (como no caso do estudante que faz um teste de lógica e não sabe se acertou todas as questões - Sorensen, 2001, 157), ou ela decorre de dois processos cognitivos distintos (de forma análoga à ilusão de ótica do triângulo de Penrose: a imagem formada conflita com nosso conhecimento de geometria euclidiana – Sorensen, 2001, 79) e é resolvida, quando necessário, em “tempo de execução” pelo sujeito consciente (como um compilador que corrige um pequeno erro do programador, ou um sistema operacional que arbitra entre demandas conflitantes de *softwares* distintos).

<sup>61</sup> Superiores em relação aos que não agem dessa forma. Mesmo *superforecasters* tiveram desempenho não muito melhor que a sorte, e foram incapazes de prever “cisnes negros” (Tetlock & Gardner, 2017, xx-xxi).

Mas a compreensão das limitações de nossa racionalidade sugere uma explicação sobre como, mesmo na ausência de bolhas (entendidas como a inflação do preço de um ativo sem base nos “fundamentos do mercado”), mercados financeiros podem amplificar vieses (de otimismo ou pessimismo) e apresentar “variações selvagens” (Mandelbrot & Hudson, 2008) – o que Grauwe (2008), Lo (2013) e Greenspan (2013) associam à noção de *animal spirits* de Keynes. Podemos dizer que “Stock prices contain lots of information. Markets, as Friedrich Hayek argued, are the best aggregators of information known to man. Yet mixed up amid the information in security prices is an awful lot of emotion, error, and noise” (Fox, 2011, 307). Esses problemas podem ser agravado pela inovação tecnológica – em especial, por algoritmos de transações de alta frequência (*high-frequency trading* - HFT), que supostamente deveriam tornar o mercado mais eficiente (no sentido da HME) ao esgotar as oportunidades de arbitragem rapidamente; como exemplo, citam-se o *flash crash* de 6 de maio de 2010, com perdas (em volume de negociações) estimadas em U\$ 1 trilhão, e o *crash* 23 de abril de 2013, com perdas de U\$ 136 bilhões (Kirilenko & Lo, 2013)<sup>62</sup>.

#### 1.4 Porque não é simples fugir de atribuições de probabilidades

Se nosso problema (a incerteza) é não poder estimar a probabilidade dos riscos incorridos, por que não adotamos uma concepção distinta de *risco*, que não faça apelo a probabilidades? É o caso da concepção de Pritchard (2015) de risco, que apela para uma “explicação modal”; ele a distingue da “concepção probabilística” mediante a seguinte comparação entre dois eventos equiprováveis, sobre os quais não se pode influir: suponhamos que uma bomba explodirá se, (i) numa loteria, um número específico dentre 14 milhões for retirado, ou se (ii) ocorrer uma conjunção de eventos bizarros (como, e.g., o pronunciamento espontâneo de uma frase em polaco no próximo discurso da Rainha, a vitória de uma azarão no *Grand National...*), com a probabilidade *a posteriori* de 1 em 14 milhões. De acordo com o argumento, o segundo evento seria um risco menor, porquanto “modalmente mais distante” –

---

<sup>62</sup> Este teria começado com um *tweet* da Associated Press (inserido por um invasor desconhecido) sobre uma explosão na Casa Branca, retirado do ar 5 minutos depois – tempo suficiente para ser lido por algoritmos de leitura de notícias, embora qualquer humano que porventura o tivesse visto o teria rapidamente desconsiderado como *fake news* (Kirilenko & Lo, 2013). Mas um *trading bot* não apenas tende a atribuir uma probabilidade alta a qualquer informação originada de uma fonte confiável (pois carece da habilidade humana de senso comum) como, num contexto estratégico, se presume que minimizará sua exposição sempre que atribuir uma probabilidade suficientemente grande a:

(A) um número suficientemente grande de outros agentes vai minimizar suas exposições.

Ninguém precisa, pois, acreditar na informação extraordinária; só é preciso atribuir uma probabilidade suficientemente alta a A – que se torna assim uma previsão auto-confirmatória.

i.e., está em um “mundo possível” menos plausível. Um defensor deste argumento, no entanto, deveria trocar uma perda de \$100,01 sob a condição (a) *sair um 6 num lance de um dado imparcial*, por uma perda de \$100 sob a condição (b) *sair uma sequência cara-cara-coroa em três lances de uma moeda imparcial (sendo que, caso os três lances tenham resultado idêntico, a moeda é relançada outras três vezes)*<sup>63</sup>. Embora a segunda condição seja mais complexa – e, no sentido dado por Pritchard, “modalmente mais distante” – as duas são equiprováveis, e a troca implica um deságio de \$0,01.

Não há escapatória simples para a concepção probabilística: uma vez que atribuímos probabilidades a uma classe de eventos, a atualização de crenças pela regra de Bayes é a única forma de o agente obter informação nova sem violar dinamicamente as demandas de consistência e transitividade, de forma insuscetível a um *dutchbook* (Skyrms, 1987). Então, por que a concepção de Pritchard é atraente? Uma possível resposta é que ela se serve de uma “bomba de intuição (*intuition pump* – para usar o rótulo de Dennett, 2013) potencializado por um *viés cognitivo* de representatividade: mesmo sabendo que as duas opções são equiprováveis, preferimos a mais simples de conceber. Algo análogo ocorre quando as pessoas julgam, e.g., que a frequência de uma determinada letra é maior no começo que no meio de uma palavra, uma vez que é mais fácil lembrar palavras que comecem com uma determinada letra do que palavras que a contêm em outra posição; ou quando julgam, ao jogar cara (K) ou coroa (C) com uma moeda imparcial, a “sequência K-C-K-C-C-K como mais provável do que a sequência K-K-K-C-C-C, que não parece ser aleatória” – embora, por definição, ambas tenham a mesma probabilidade  $2^{-(6)}$  de ocorrer (Tversky & Kahneman, 1979, 7; reproduzido em Kahneman, 2012, 527). Somos tentados a ver a sequência menos complexa como incompatível com a aleatoriedade do processo.

Tais vieses decorrem da extrapolação de heurísticas – processos de julgamento e decisão que costumam levar a resultados satisfatórios nas situações de racionalidade limitada em que nossa inteligência evoluiu. De fato, se acrescentássemos à situação do experimento mental de Pritchard a verdadeira incerteza do mundo real, ainda que atribuíssemos a mesma probabilidade subjetiva aos dois eventos, nós os distinguiríamos pela quantidade de informação disponível: no cenário (i), temos *informação completa* (conhecemos, tão bem como qualquer outro agente, o que há para saber sobre as variáveis relevantes – a loteria), donde nossa

---

<sup>63</sup> A probabilidade de um *cara-cara-coroa* é  $2^{-3}$ , normalmente; mas, no caso de (b), é  $1/6$ , pois estão excluídos os resultados *cara-cara-cara* e *coroa-coroa-coroa*. Pode-se simular qualquer probabilidade por meio de um processo afim.



estimativa se aproxima muito da probabilidade objetiva<sup>64</sup> do evento – ela é idêntica à estimativa de qualquer outro agente; já no cenário (ii), há menos informação e um alto grau de incerteza, donde as estimativas tendem a ser *ambíguas* (ao invés de uma probabilidade bem definida, os agentes a estimariam dentro dum intervalo relativamente amplo; afinal, como estimar a probabilidade de a Rainha falar polonês?). Ela seria percebida como menos arriscada graças à aversão a cenários de maior informação em situações de perda. Outrossim, o fato de o cenário (ii) se basear numa sucessão de eventos (mormente porque se trata de ações, e não de eventos aleatórios) o faz parecer mais evitável – uma impressão que tende a se manter em nossa apreciação intuitiva, mesmo depois de excluída pela declaração de que não podemos influir sobre ele; nessa linha, é provável que um agente preferisse prevenir-se ou comprar um seguro contra (i) a o fazer contra (ii).

É melhor interpretar a teoria de Pritchard como uma hipótese descritiva de como as pessoas, em geral, efetivamente avaliam riscos (embora isso demandasse testá-la empiricamente) – uma teoria descritiva concorrente, e.g., à *prospect theory*. Nesse caso, uma explicação mais abrangente seria vê-la como uma epistemologia da racionalidade limitada aplicada a riscos, explicando um conjunto de heurísticas efetivamente usadas na avaliação destes – adotadas porque, cotidianamente, aproximam-se o suficiente da avaliação de probabilidades por um agente racional, mas a um custo de processamento muito menor. De fato, a ideia de distância modal da epistemologia de Pritchard parece compatível com uma *teoria da decisão baseada em casos* (Gilboa & Schmeidler, 2003), em que, no lugar dos juízos bem definidos sobre probabilidades presentes na teoria da utilidade esperada, parte-se de avaliações sobre a *similaridade* entre eventos observados no passado com possíveis eventos futuros. Ao contrário de Pritchard, porém, Gilboa & Schmeidler reconhecem que:

a) sua teoria é, efetivamente, uma teoria da racionalidade limitada (Gilboa & Schmeidler, 2003, 41), combinando raciocínio estatístico, dedutivo e analógico, a respeito de como as pessoas efetivamente decidem em contextos de informação muito restrita – como

---

<sup>64</sup> É desnecessário travar uma discussão ontológica sobre o significado de probabilidades objetivas x subjetivas aqui (e mais ainda sobre o realismo modal); a observação funciona com uma noção *relativa* de probabilidade subjetiva (no estilo de Gilboa & Schmeidler, 2003): e.g., julgo que uma moeda *n* tem uma probabilidade *objetiva* de 0,5 de resultar em cara analisando o histórico de seus lances anteriores (i.e., a frequência de *cara* se aproxima, no limite, de 0,5), enquanto que uma moeda *m*, sobre a qual não tenho qualquer informação, tem uma probabilidade subjetiva de 0,5 de resultar em cara porque, no caso, aplico o princípio de razão insuficiente de Laplace – ou seja, via de regra, dizemos que um juízo de probabilidade é *mais* subjetivo na medida em que a avaliação é feita a partir de *menos informações*.

quando não podem mapear todos os cenários possíveis, nem estimar adequadamente a probabilidade ou a utilidade destes (Gilboa & Schmeidler, 2003, 1-2);

b) pode-se “embutir”, matematicamente, essa teoria na teoria da utilidade esperada, e vice-versa (Matsui, 2000) – i.e., no limite, num cenário de informação completa e perfeita, as duas teorias fariam previsões/prescrições equivalentes. Na realidade, Gilboa & Schmeidler (2003, 93) defendem que ambas são antes “diferentes estruturas conceituais” (*conceptual frameworks*), nas quais teorias específicas voltadas para determinados tipos de problema de decisão são formuladas, do que propriamente teorias concorrentes<sup>65</sup>.

## 2 Incerteza e filosofia política normativa

### 2.1 A pervasividade da objeção: teorias da justiça e legitimidade

A filosofia política normativa pode ser tentada a defletir a objeção, sob o argumento de que teorias da justiça não dizem nada sobre como os agentes fazem juízos de futuros contingentes, mas sobre como *deveriam* se comportar (proposições normativas). Contudo, crenças sobre como as pessoas se comportarão e sobre o que pode ocorrer são relevantes na medida em que definem o que é possível que ocorra (Sorensen, 1995, 267). Limitações epistêmicas podem tornar impossível identificar o ato recomendado por uma determinada teoria da justiça: para esses casos, o respectivo teórico costuma recomendar procedimentos especiais, de modo que o “ato correto” passa a ser aquele determinado por esses procedimentos; mas então, se o método e o respectivo princípio podem divergir, as recomendações do teórico assumem um caráter paradoxal, se não contraditório (Kagan, 2018). Podemos assim estender o argumento cético da teoria da ação para a teoria da justiça: uma teoria é *inaplicável* se não se pode acreditar, justificadamente, que se pode agir para obter os fins por ela prescritos, e é auto-derrotável (*self-defeating*) se, agindo de acordo com os princípios por ela prescritos, *não podemos* chegar ao fim por ela prescrito.

Um argumento desse tipo é frequentemente utilizado contra o utilitarismo, concluindo com a recomendação de que os agentes não professem o utilitarismo (se desejam atingir o

---

<sup>65</sup> A teoria baseada em casos dá lugar de destaque ao raciocínio analógico associado à prática jurídica – justamente o tipo de cenário focado por Pritchard (2015). A avaliação de provas num processo judicial é difícil de capturar num modelo que use a teoria da utilidade esperada, pois o processo de decisão é permeado por uma divisão de ônus probatório entre as partes e por regras e precedentes compartilhadas entre elas e os oficiais; tais regras visam não apenas à economia de tempo e de outros recursos escassos, mas também a valores como publicidade e equidade, restringindo a busca por exatidão e eficiência (Sunstein, 1994).

melhor estado de coisas possível); mas utilitaristas não veem problema na objeção (Elster, 2016; Sorensen, 1995, 250): aceitam a inaplicabilidade imediata do utilitarismo de atos e, a partir daí, argumentam a favor de algo afim ao utilitarismo de regras – i.e., adotar regras e políticas tendentes à utilidade geral teria maior utilidade geral esperada do que decidir, individual e constantemente, qual ação tem maior utilidade geral esperada (e.g.: Harsanyi, 1985).

É um engano pensar que a objeção se limita a teorias consequencialistas ou teleológicas. Afinal, as consequências da ação também podem ser relevantes para teorias deontológicas<sup>66</sup> – do contrário, elas são simplesmente “insanas” (Rawls, 1971, 26)<sup>67</sup>. Mesmo um “kantiano” não estaria imune à objeção: ele pode ter de se preocupar com as consequências de sua ação ao definir a máxima sobre a qual aplicar o imperativo categórico (Gibbard, 2007, 207 e s.; Dennett, 1986, 129) – o que pode, e.g., levá-lo a adotar princípios especiais para lidar com o *mal*, como na controvérsia sobre o direito de mentir em legítima defesa (Korsgaard, 1986); e mesmo concepções deontológicas de direitos individuais têm de incluir considerações sobre consequências ao tratar de riscos e danos (e.g.: Nozick, 1990, 73 e s.; Dworkin, 2011, 290).

A *objeção cética* é particularmente grave para a legitimidade da autoridade, já que esta é embasada, via *deregra*, em algo afim à *tese da justificação normal* de J. Raz (1986) – segundo a qual a autoridade é justificada se o sujeito tem mais chance de atender às razões que se aplicam a ele acatando suas diretivas do que se buscar seguir essas razões diretamente, por si mesmo. Essa tese é associada ao requisito de *dependência*: a diretiva da autoridade deve se basear em razões aplicáveis ao próprio sujeito<sup>68</sup> (Raz, 1986, 53 e s.). Nessa teoria, a autoridade nos presta um *serviço*: “terceirizamos” o papel de decidir qual a melhor coisa a fazer; contudo, se a situação padrão em questões sociais é a de *ignorância*, então a autoridade não pode conhecer melhor do que o sujeito as razões para agir.

---

<sup>66</sup> Especialmente quando se aborda justiça distributiva; no prefácio de *Sovereign Virtue*, Dworkin (2000) defende que cada ação ou omissão do estado beneficia alguém, e que podemos mais ou menos estimar de quem se trata; por isso, devemos ser capazes de justificar o status quo àqueles que estão em pior situação – já que poderíamos ter tomado decisões diferentes. Contudo, se, como aduz a objeção cética, ninguém é capaz de prever ou estimar adequadamente as consequências relevantes de uma regra ou política pública, então não podemos fornecer tal justificação.

<sup>67</sup> Pode-se discutir se a objeção cética ameaçaria inclusive o princípio da diferença de Rawls: “(...) that social and economic inequalities must be arranged so that they are to the greatest benefit to the least advantaged. Insurmountable problems about measuring social and economic status might make a mystery of what counts as satisfying this obligation” (Sorensen, 1995, 248). Contudo, assim como os utilitaristas, Rawls parece estar bastante consciente de que os *desiderata* de uma teoria política podem conflitar em tal sentido – como referimos infra, uma das principais motivações da redação do *Liberalismo Político* (v. Rawls, 2005, xvi) teria sido que a concepção de estabilidade da Parte III de *Uma Teoria da Justiça* (Rawls, 1971) conflituaria com a “doutrina filosófica abrangente” liberal do restante da obra.

<sup>68</sup> É isso o que distingue a noção de autoridade da exploração e da coerção simples.

Donde se pode derivar um argumento contra a legitimidade da regulação por uma autoridade central e técnica: à direita do espectro político, esse argumento pode ser relacionado, e.g., aos economistas da Escola Austríaca, como Von Mises (1949) e Hayek (1973, 11-12), para os quais seria melhor minimizar o número de regras e políticas produzidas por um Estado (já que não se podem conhecer adequadamente seus efeitos); à esquerda, ele pode ser encontrado nas objeções da teoria crítica à tecnocracia, a exemplo de Oliveira (2003) – para quem o Banco Central é a “instituição mais anti-republicana e mais anti-democrática”<sup>69</sup>. Afinal, a tecnocracia só pode ser baseada na “razão instrumental”; principalmente diante de incerteza, caberia abrir espaço para o que Habermas (2012) chama de “razão comunicativa”, exercida por meio da deliberação em grupo, que visa a obter “compreensão e consenso” – mais fundamental que a racionalidade instrumental, que visa apenas ao “sucesso”, a atingir um objetivo (Finlayson, 2005, 139).

O problema é que a *objeção cética* prova demais; se a incerteza nos impedisse de formar estimativas, isso não levaria apenas à impossibilidade do planejamento e da ação social (como ora apresentamos), mas à impossibilidade da ação racional *simpliciter*, já que é impossível agir sem ter expectativas sobre o futuro<sup>70</sup>: haveria “complete anarchy in the conduct of men, with experience being plainly in contradiction. Such an assumption would make the existence of economy just as impossible as economic theory” (Morgenstern, 1925, 175). Isso compromete os argumentos anti-tecnocracia do teórico crítico e do libertário; assemelham-se, pois, à pessoa que, observando que nossos planos nunca correspondem adequadamente à realidade, sugere que não façamos plano algum.

Se incerteza sobre o futuro implica irracionalidade da decisão, decisões tomadas por uma assembleia são tão irracionais quanto as tomadas por um especialista, já que também os membros da assembleia estão impedidos de fundamentar suas previsões e estimativas. Apesar dos benefícios (e.g., um senso de autonomia, igual consideração e pertencimento à comunidade) de instituições desenhadas para fomentar a participação dos

---

<sup>69</sup> A discussão é particularmente atual diante da tramitação do Projeto de Lei que visa à autonomia do Banco Central do Brasil – assim como as discussões, no âmbito da União Europeia, sobre a limitação da soberania dos países em relação à política monetária.

<sup>70</sup> É difícil exagerar no destaque a esse ponto: gestores de firmas enfrentam um contexto de incerteza pelo menos tão grave quanto o de autoridades; no entanto, libertários e economistas da escola austríaca relutariam em lhes aplicar o elogio da passividade. A diferença não é que “firmas visam ao lucro” – já que a discussão, até o momento, deu-se no âmbito da teoria do juízo e da decisão, não dos incentivos pertinentes a cada agente. Afinal, a idéia de maximização do valor do capital é análoga à de maximização dos interesses do cidadão (embora nenhuma das duas se aplique plenamente à realidade); ainda, a comparação também vale para associações sem fins lucrativos – há inclusive ONGs voltadas a objetivos de longuíssimo prazo, como o *Future of Humanity Institute* em Oxford, que assumem que, dada uma pequena tendência ao progresso, o melhor é minimizar riscos de catástrofes globais.

interessados, *objeta-se* (seguindo Elster, um teórico da democracia e crítico da racionalidade instrumental) a que:

these benefits can be the main or even the sole point of the system. This would be to turn into the main purpose of politics something that can only be a by-product. It can indeed be highly satisfactory to engage in political work, but only on the condition that the work is defined by a serious purpose which goes beyond that of achieving this satisfaction. If that condition is not fulfilled, we are left with a narcissistic theory of politics (Elster, 2016, 91-92).

O juízo do cético libertário também é injustificável, de acordo com a mesma objeção: não sabemos quais as consequências de “cruzar os braços” e não podemos compará-la com as alternativas<sup>71</sup>. Mas, se tal comparação for possível, então não precisamos sequer identificar a solução ótima para a decisão sob incerteza – basta mostrar que tende a ser melhor agir (ainda que sob algumas restrições) do que não agir<sup>72</sup>.

---

<sup>71</sup>Trata-se de uma forma de *reductio* aplicada à objeção de Hayek ao planejamento central:

"The trouble is that it also applies to unfettered market capitalism. No doubt markets transmit information in the way that Hayek claimed. But what reason is there to believe that – unlike any other social institution – they have a built-in capacity to correct their mistakes? History hardly supports the supposition. Moods of irrational exuberance and panic can, and often do, swamp the price-discovery functions of markets" (Gray, 2015).

"Von Hayek even quoted Gödel's theorem in an attempt to demonstrate that efficient planning is impossible failing to notice that if he were right, markets wouldn't be able to work efficiently either, because the whole of society can be thought of as a kind of gargantuan computing machine" (Binmore, 2009, 189-190).

Arriscando ser repetitivo: *todo* tomador de decisão tem limites especiais – “pontos-cegos” (*blindspots*). Como arguiremos ao fim do texto, se o processamento descentralizado de informações do mercado lhe permite realizar coisas que uma autoridade central não pode fazer, o inverso também é verdadeiro.

<sup>72</sup> Cabe antecipar que esse argumento **não** é um ataque à teoria crítica e ao libertarianismo em si (o que não é nosso objetivo), mas a seu uso do argumento cético no ataque à tecnocracia; porém, eles não são presa desse argumento quando, e.g., ao invés de apelar para um argumento cético geral, apresentam argumentos sobre:

(i) A teoria da escolha social – caso em que se as podem ver como duas teorias concorrentes sobre a melhor maneira de agregar a informação dispersa na sociedade e de decidir (por meio de votações e deliberações públicas vs. por um mercado livre), e assim cotejar suas respectivas vantagens e desvantagens (e.g., em Sunstein, 2006).

(ii) *Restrições* à autoridade legítima. Nesse caso, um libertário pode apelar para uma teoria *lockiana* dos direitos individuais como limites à legitimidade dum estado mínimo (como em Nozick, 1990); já o democrata pode apelar para o conceito de representação e afirmar que, embora se possa dizer que uma autoridade pode agir no interesse do sujeito (satisfazendo, pois, a tese da justificação normal) sem o consentimento deste (de forma análoga a um gestor de negócios), ela, nesse caso, não o *representa* num sentido substantivo (Pitkin, 1972, 152).

Vale destacar que mesmo Huemer (2012) reconhece que é preciso mais do que o ceticismo para concluir pela passividade, adicionando como premissas um viés de não-intervenção baseado numa aversão à coerção (ações sociais demandam coerção; omissões, não) e num “argumento entrópico”: é mais provável que uma intervenção qualquer danifique um sistema complexo (como a sociedade, ou o corpo humano), ao invés de o melhorar.

Não aprofundaremos a consideração sobre estes pontos: primeiro, para simplificar a exposição do argumento; segundo, porque o viés pró *status quo* depende da assunção de que o sistema social se encontra em alguma espécie de equilíbrio – mas **essa assunção é incompatível com o grau de ignorância** que Huemer invoca (mesmo no caso da analogia com o corpo humano e a medicina, o viés de não-intervenção só se sustenta para corpos saudáveis e intervenções aleatórias – médicos não ferem a ética profissional quando tomam decisões arriscadas no interesse de pacientes desacordados). Pelo contrário, de acordo com o que sabemos hoje, pode-se supor que alguns processos sociais importantes, como a desigualdade de renda, **não** estão em qualquer forma de equilíbrio (Adamou & Peters, 2016). Seria preciso aprofundar considerações sobre teorias do equilíbrio para apresentar esse argumento de maneira mais convincente, mas um exemplo de teoria dos jogos evolutivos, utilizado para modelar a cooperação

## 2.2 Incerteza e imparcialidade

Isso nos permite retornar à filosofia política normativa – ela pode nos fornecer outros princípios para selecionar soluções para um problema de decisão em grupo além da consistência (e da utilidade esperada) como, e.g., imparcialidade. Podemos testar a imparcialidade de nossos princípios de decisão no experimento mental de uma “posição original”, na qual, sob um “véu de ignorância”, os sujeitos desconhecem suas características e interesses particulares (Freeman, 2016). Harsanyi (1953; 1975) tem um experimento análogo (do *observador imparcial*), com uma conclusão *bayesianista*: aplicando o princípio de razão insuficiente, um participante na situação da posição original atribuiria a mesma probabilidade a ocupar a posição de qualquer indivíduo após o experimento; portanto, sem conhecer sua própria função utilidade (i.e., suas preferências pessoais), o agente racional escolheria os princípios que levariam à maior utilidade

---

social, pode ser suficiente para ilustrá-lo: como argúi Axelrod (1984), num dilema do prisioneiro reiterado indeterminadamente (ao contrário da clássica versão de *round* único), a cooperação mútua é um possível *nash-equilibrium*; uma vez que um agente agressivo sinaliza ao demais que não é confiável, ele tende a ser punido com a deserção em suas interações seguintes, incentivando os agentes a adotarem estratégias condicionadas – como cooperar na primeira interação, e depois copiar a última ação do adversário (o *tit-for-tat*, ou “toma lá, dá cá”), ou cooperar apenas até ser vítima de um agente não-cooperativo (a estratégia *severa*). Contudo, não há um equilíbrio *evolutivamente estável* (i.e., com estratégias imunes à “invasão” por parte de outras estratégias com menor prevalência na população de agentes) no dilema do prisioneiro: uma população que adote *tit-for-tat* pode ser “invadida” com sucesso por “mutantes” que cooperam incondicionalmente – os quais, por sua vez, seriam presa fácil para *defectors* (não-cooperadores) incondicionais, que então dominariam a população, até que cooperadores condicionais os superassem, reiniciando o ciclo (Binmore, 2007, 136).

Esse argumento é confirmado pela abundância de simulações de dilemas do prisioneiro; o leitor mais interessado pode encontrar diversos programas para tais simulações – um dos mais simples é o aplicativo para celular oferecido pela Pygmalion (<http://pygmalion.nitri.org/>), que permite oferece simulações gráficas (em que cada *pixel* representa um autômata-agente, cuja cor refere a estratégia adotada no *round* anterior) do jogo; de acordo com a prevalência inicial e *pay-off* oferecido a *defectors*, a simulação pode resultar em: a) equilíbrios estáveis dominados exclusivamente por cooperadores ou *defectors*, ou com “ilhas” com um tipo de agente em meio a um “mar” de outro tipo, ou b) uma infundável disputa entre os diferentes tipos de agente (resultando numa animação afim a um caleidoscópio), sem resultar em equilíbrio. Podem-se encontrar notáveis simulações análogas, nas pesquisas de Skyrms (2014), Vanderschraaf (2006, 264-6), Gintis (2006, 11-12) – que aduz que a existência de “altruísmo punitivo” (i.e., de punição não só a *defectors*, mas também aos que falham em punir *defectors*) aumenta a probabilidade de cooperação. Tais simulações ressaltam o papel de instrumentos de correlação e a possibilidade de aprendizado; nesse sentido, frisamos a surpreendente pesquisa de Dedeo (2017), que apresenta uma simulação dum dilema do prisioneiro reiterado, num contexto com contexto de memória e aprendizado limitados (“a simple language to think with and enough resources to have memories and to act on the”) e mutações (“successful machines passed their strategies to the next generation, with the occasional slight variations designed to imitate the blind distortions typical of cultural evolution”). Ao invés de confirmar a dominância de estratégias de cooperação condicional conhecidas, o resultado foi a emergência de uma forma de “tribalismo”, apelidada de *Shibboleth*:

“After an early, chaotic period, a single machine would rise rapidly to dominance, taking over its imaginary world for hundreds of generations until, just as suddenly, it collapsed, sending the world into a chaos of conflict [...]. The dominant machines had taken players’ actions to be a code by which they could recognize when they were faced with copies of themselves[...]. Because the sequence was so hard to hit upon by accident, only the descendants of ruling machines could profit from the post-code era of selfless cooperation. All others were killed off, including those using the *tit-for-tat* strategy. This domination would last until enough errors accumulated in the code handed down between generations for dominant machines to stop recognizing each other” (Dedeo, 2017, n.p.).

geral esperada. Para Rawls (1971), diante do cenário de incerteza, um agente escolheria a opção menos arriscada – incluindo, entre seus princípios de justiça, o “princípio da diferença”, pelo qual as desigualdades de riqueza e renda são arranjadas tendo em vista o benefício dos que estão em pior situação (Rawls, 1971, 72).

Nessa linha, Shrader-Frechette (2014, 194) argumenta, em favor de Rawls e contra Harsanyi, que, em tomadas de decisão em situações de alta incerteza que tenham: a) por escopo uma sociedade (e não só um indivíduo); b) potencial de catástrofe e c) desigualdade na distribuição dos efeitos adversos, o princípio de *maximin* (e, conseqüentemente, o princípio da diferença) é superior<sup>73</sup>. Um dos motivos é que, em situações de grave incerteza, não conhecemos todos os cenários possíveis – o que impede uma aplicação adequada do princípio da utilidade

---

<sup>73</sup> Estou negligenciando os “quatro argumentos” que Shrader-Frechette (2014, 190-94) lança, parafraseando Rawls, contra o utilitarismo; primeiro, por simplicidade, e segundo, porque esses argumentos almejam o utilitarismo em geral, e são pouco relevantes para o ponto de vista da posição original – embora se possa cogitar que as objeções ao utilitarismo em geral possam adicionar força à ideia de que este não fornece uma base estável para a cooperação.

Outrossim, essa exposição não é uma defesa incondicional do *maximin* e do princípio da diferença, apenas um contraste com o princípio da utilidade esperada, visando a argumentar *contra* o elogio da passividade; por isso, negligencio diversos princípios de justiça “mistos” (i.e., que aplicam o princípio da diferença sob certas condições – como até um nível mínimo de bem-estar) ou concorrentes que poderiam ser relevantes para o problema da atuação da autoridade em situação de incerteza, como o “mercado ideal de seguros” de Dworkin (2011, 352). Uma das explicações é que essas teorias capturam aspectos salientes dos argumentos do debate Harsanyi-Rawls – quando não são vistas como um compromisso entre o *desiratum* igualitário do *maximin* com a eficiência de uma análise de custo-benefício utilitarista.

Na verdade, o próprio Harsanyi admite que o *maximin* pode ser uma boa *proxy* do princípio de utilidade – fazendo uso da distinção entre “princípio filosófico básico” e princípios aplicáveis à tomada de decisão (Harsanyi, 1975, 606):

“To be sure, the maximin principle does have its valuable uses, and we must be grateful to Rawls for calling our attention to it. Even if it cannot serve as a basic principle of moral theory, it can be used as a principle of approximate validity in practical applications, such as the theory of optimal income distribution or of optimal taxation. In such applications, its relative independence of detailed interpersonal utility comparisons, and of the actual mathematical form of people's von Neumann-Morgenstern utility functions for money, is an important advantage, and can be fruitfully exploited in economic studies.

(...) These practical merits of course do not in any way provide a reason for abandoning utilitarian moral philosophy. (Basic philosophical principles must be exactly right, and not merely approximately right.) But they do provide a reason, even for a utilitarian moral philosopher, to use the maximin principle as an admissible approximation in many cases”.

Há uma heurística alternativa negligenciada nessa discussão: enquanto o princípio de utilidade esperada manda escolher a ação associada à maior média da distribuição de utilidade entre os possíveis resultados (ponderados pela respectiva probabilidade), a *heurística da mediana* manda escolher a ação com a maior *mediana* dessa distribuição em situações de incerteza: “Statistically, expected value is the central tendency of the distribution embodied in a risky gamble. For highly non-Gaussian distributions the mean is not considered a valid estimator (...). The median, an alternative estimator of central tendency, is robust to noise and is often favored for highly skewed distributions” (Hayden & Platt, 2009). Na ausência de outras informações, o agente tem a mesma probabilidade de resultar tanto na posição *acima* quanto *abaixo* da posição mediana. Considerando que, via de regra, a distribuição de renda em sociedades complexas é incompatível com uma distribuição normal (Mandelbrot & Hudson, 2008) ou mesmo com a ideia de que o processo subjacente tende a atingir alguma forma de equilíbrio (Adamou & Peters, 2016), então a heurística da mediana na posição original (i.e., escolher a sociedade que maximize a utilidade/recursos do indivíduo na posição mediana) tenderia a convergir mais para o *maximin* que para o utilitarismo – embora não coincidissem com nenhum dos dois.

esperada. Um exemplo ordinário: suponha que concordemos em lançar uma moeda qualquer amanhã. Devemos, pois, atribuir uma probabilidade  $p$  ao resultado “cara”,  $p$  a “coroa”, e  $(1 - 2p)$  a “a moeda não foi lançada, por algum motivo qualquer”. Nesse caso, como estimamos  $p$  – se tivermos de considerar a probabilidade de que a aposta não se concretize?

Sugerimos que os dois autores chegam a conclusões distintas principalmente *porque* partem de experimentos *qualitativamente* distintos, com objetivos distintos – como arguido por Moehler (2015). Em Rawls (1971, 65), há agentes em vias de definir a estrutura básica onde ocorrerá a repartição de “benefícios e ônus da cooperação” de uma sociedade estável. Já em Harsanyi, eles devem escolher qual distribuição de utilidade prefeririam - i.e., qual mundo possível prefeririam ocupar, dado que só conhecem tal distribuição (sem incluir, pois, informação sobre como ocorrerá a cooperação). De fato, assumindo uma população constante e uma probabilidade igual de ocupar o lugar de qualquer indivíduo, um indivíduo racional poderia preferir viver no mundo/sociedade com maior utilidade esperada média por agente (Harsanyi, 1975, 598) – isso equivaleria a escolher a situação em que o agente tem maior probabilidade de satisfazer em maior grau o maior número de suas preferências. Mas, nessa situação, faz pouco sentido falar em *contrato social*: ela é análoga a perguntar a alguém se ele preferiria nascer numa sociedade igualitária pré-histórica, com índice de *Gini* perto de zero e expectativa de vida de 30 anos, ou no século XXI. É a resposta certa para a pergunta errada<sup>74</sup>.

Mesmo em Rawls, a defesa do princípio da diferença envolve mais do que o simples fato da incerteza. A posição original não é apenas uma decisão sob ignorância completa, contra uma entidade fictícia sem intenções chamada “Natureza”, pois antecipa as interações de conflito/cooperação que ocorrerão após a retirada do véu de ignorância<sup>75</sup>. Nesse caso, o princípio da diferença fornece a base de um contrato social garantindo, aos cidadãos em pior

---

<sup>74</sup> Um problema dessa distinção é que permite que perguntemos se, de fato, a posição original é o melhor contexto a partir do qual determinar os princípios de justiça que *nossa* sociedade deve adotar. Esse é um ponto cotidianamente frisado por Sen (1986; 1999; 2009, 93): a questão de quais princípios devemos adotar depende do tipo de informação disponível – mesmo na posição original. Isso permite explicar por que a discussão sobre “igualdade de quê?” (recursos x utilidade) é *irrelevante* para nossa atual análise sobre a posição original (embora seja crucial para a filosofia política pós-Rawls); afinal, se os agentes *não têm* informação sobre utilidade na posição original, então a decisão sobre os princípios de justiça seria tomada com base na informação sobre a distribuição de recursos – ajustada por alguma taxa de desconto marginal de utilidade (se tivermos essa informação). Igualmente, se não tivermos informações sobre recursos, mas sobre “utilidade esperada ao longo da vida”.

<sup>75</sup> Essa caracterização pode ser controversa – mas ela parece seguir da ideia de que deve haver uma conexão entre a posição original e a sociedade após a retirada do véu. Gauthier (1984) e Binmore (2005), sim, concebem a posição original como um *jogo* (uma situação de decisão não-paramétrica); contudo, eles o modelam como um *problema de barganha* bilateral, o que poderia não representar adequadamente diversos aspectos de um contrato social com uma multiplicidade de partes, como a formação de coalizões, reputação e a evolução de convenções (Sugden, 2001, 235).



situação, que a única maneira de melhorar sua condição relativa seria colocando outrem numa posição igual ou pior. Na realidade, a justificação do princípio da diferença não repousa *apenas* no fato da incerteza, mas também na exigência de publicidade e de reciprocidade (Freeman, 2016; Rawls, 2001, xvii): a sociedade deve organizar-se em princípios *estáveis* que todos possam identificar e aceitar.

Por isso, Binmore (2005, 170) aduz que um igualitarismo liberal seria a opção de contrato social na ausência de uma entidade externa para definir e aplicar as normas sociais, na medida em que estas podem ser justificadas por razões compatíveis com diferentes doutrinas filosóficas abrangentes razoáveis<sup>76</sup>. O que uma tal sociedade perde em potencial eficiência é compensado em equidade – e, principalmente, em estabilidade<sup>77</sup>.

Outra maneira de colocar o problema é dizer que, em Rawls, a decisão na posição original é uma razão para que o indivíduo atenda ao contrato social após se levantar o véu de ignorância – pois, como frisamos, trata-se de definir os princípios sobre os quais cooperaremos. Mas isso não ocorre na teoria de Harsanyi – em que a conclusão (utilitarismo) é uma razão de *um ponto de vista destacado*:

(...) actual individuals would not see the reasons of the chooser in Harsanyi's original position as reflecting their reasons. In this way, the justificatory link between the model of rational choice in the original position and the reasons of actual individuals in society would be severed. Individuals will not see Harsanyi's justification of average utility as a justification for *them*. (Gaus & Thrasher, 2015, 57 – grifo do original)

Mesmo que o argumento de Harsanyi seja considerado consistente enquanto fundamento do utilitarismo como *doutrina filosófica abrangente* – ao menos *na situação em que não temos* outras informações sobre o problema de decisão – ele não fornece uma razão para alguém, já inserido na sociedade, apoiar as respectivas instituições utilitaristas; pelo contrário, pode parecer revoltante que os que estão em melhor situação se justifiquem aos desafortunados: “lamento seu azar, mas tínhamos a mesma chance *a priori* de sucesso, e minha felicidade compensa sua miséria”.

---

<sup>76</sup> Há aqui algo que lembra a noção de *ponto focal* de Schelling (2015) – um equilíbrio *saliente* para o qual todos podem convergir.

<sup>77</sup> Para Binmore (2005, 5-9), esse é o *desideratum* mais importante de normas sociais, por um raciocínio evolucionista – i.e., considerando o simples fato de que normas instáveis, por definição, não persistem por um longo tempo (ou elas são alteradas, ou as sociedades que regem se dissolvem).

A solução de Harsanyi, pois, subestima a necessidade de coordenação e cooperação justificadas, negligenciando seu caráter não-paramétrico (Regan, 1985; Gauthier, 1984, 71). Numa sociedade utilitarista, os favorecidos pela transferência dos benefícios e ônus da cooperação não seriam os mais necessitados, mas aqueles com maior potencial de incrementar a utilidade marginal geral; isso pode coincidir com os mais necessitados – dada a lei dos retornos decrescentes<sup>78</sup> – mas não necessariamente. Como no gerenciamento de uma atividade produtiva, uma sociedade voltada para a eficiência seria tentada a transferir recursos a agentes capazes de incrementar a “produção” de utilidade (Sen, 1979, 9)<sup>79</sup>. Algo análogo ocorre na teoria da escolha social, onde se pode defender o *princípio majoritário* como o procedimento de decisão coletiva em que cada participante tem a *mesma* e a *maior* probabilidade *a priori* de que a escolha pública satisfaça suas preferências (Taylor, 1969); contudo, deve-se frisar a cláusula “*a priori*”, que desconsidera *outras* informações sobre o respectivo cenário: numa sociedade onde uma coalizão majoritária oprime os membros da minoria, estes não têm nenhuma razão para considerar que o argumento majoritário justifica sua situação.

Buchak (2017), a partir do argumento de Otsuka & Voorhoeve (2009) de que o deslocamento da perspectiva da decisão individual (à luz da qual ora discutimos risco e incerteza) para a da decisão *interpessoal* altera os princípios que devem ser utilizados para decidir, argumenta em favor do *risk principle*: “When making a decision for an individual, choose under the assumption that he has the most risk-avoidant attitude within reason unless we know that he has a different risk attitude, in which case, choose using his risk attitude” (Buchak, 2017, 632). Isso está de acordo com o que por vezes demandamos de autoridades. Adotamos, no direito ambiental e em medidas sanitárias, o princípio de *precaução*, que tem a vantagem de suprir a assimetria de informações entre o público e o inovador deslocando para este o ônus de pesquisar e publicar os efeitos da inovação (Picavet & Lafaye, 2012). De forma

---

<sup>78</sup> “It is a perennial idea of the utilitarian school that if utilities are concave, egalitarian consequences will follow from the sum or average rules” (Mongin & Pivato, 2016, 734). No entanto, nem o utilitarismo, nem o *maximin* nos dão motivos para preferir, na posição original (ibid.):

1 distribuição de 1 *util* a A (com probabilidade de 50%), ou a B (com probabilidade de 50%), a

2 distribuição de 1 *util* a A.

Um rawlsiano escolheria a situação (i) não pelo princípio da diferença (ao menos não por uma aplicação *ex post* desse princípio – já que, em qualquer caso, apenas um dos indivíduos será premiado), mas porque condiz com o espírito do princípio de igual oportunidade.

<sup>79</sup> O princípio da diferença também permite redistribuir recursos “para cima”, a talentosos, como incentivo – mas apenas a fim de os redistribuir, posteriormente, aos em pior situação. Esse é o “argumento do incentivo”: “Supposedly, [...] the greater expectations allowed to entrepreneurs encourages them to do things which raise the prospects of laboring class. Their better prospects act as incentives so that the economic process is more efficient, innovation proceeds at a faster pace, and so on” (Rawls, 1971, 68). Ambos os princípios têm de permitir alguma forma de redistribuição contra-intuitiva para serem “compatíveis com incentivos” – para que os agentes informem, sinceramente, suas respectivas condições e cooperem com as instituições (Gailmard & Patty, 2013).

análoga, o princípio da *prudência*, em contabilidade, determina que um ativo de valor duvidoso deve ser registrado pelo menor valor de avaliação (pois cabe ao adquirente do título, e não a seus credores, avaliá-lo); o princípio foi incorporado pela prática dos profissionais (dada a vantagem da reputação dessa prudência) de tal forma que, mesmo após sua alteração em 2011, as firmas continuaram a adotá-lo no mercado de capitais brasileiro (Kronbauer et al., 2017)<sup>80</sup>.

Outrossim, isso ecoa também uma preocupação sobre a relação entre gestores e investidores; no mercado financeiro e no mercado de capitais, demanda-se que as instituições sejam mais claras na prestação de informações sobre seu apetite a risco, quanto maior for este. Ainda, demanda-se que os riscos sejam geridos visando à robustez (*robust decision making*), a fim de otimizar a resistência a *falhas* – i.e., perdas das quais uma instituição não pode se recuperar. Um sistema é robusto se ele pode continuar funcionando, a despeito de riscos internos e externos, sem alterações drásticas em sua estrutura (Lempert & Collins, 2007); por *modus tollens*, se um sistema não é robusto, num ambiente com riscos decorrentes de processos aleatórios, ou ele deixará de funcionar, ou terá sua estrutura alterada. Em algumas atividades – que implicam externalidades, como bancos, ou mesmo quando pode haver um conflito intertemporal entre os interesses de curto prazo de um grupo de *shareholders* com os de longo prazo de outros – demanda-se que os gestores adotem medidas robustas a eventos incertos, minimizando o risco de falhas que possam ameaçar a atividade econômica, mesmo que a probabilidade destas seja pequena ou desconhecida (donde seriam subestimadas numa análise cotidiana de custo-benefício).

---

<sup>80</sup> Pode-se constatar a necessidade de tal princípio analisando as situações em que ele *não* é adotado – a exemplo da falta de informação em relação a passivos contingentes nas companhias brasileiras. Segundo o Pronunciamento Contábil nº 25 do Conselho Federal de Contabilidade (basicamente, uma tradução do IAS 37), no caso de uma obrigação dependente de uma situação incerta (ou não-quantificável), como um processo judicial, as firmas devem: a) registrar uma provisão para perda, se o risco é provável – i.e., se a probabilidade de a perda ocorrer é maior do que a de não ocorrer; b) publicar a contingência em seu balanço, se o risco é “possível”; c) não publicá-lo, nem estimá-lo, se “remoto”. O problema é que não há padrões claros para a definição destes conceitos (nem pode haver uma definição geral, dada a noção de incerteza utilizada neste artigo), nem para o estabelecimento das provisões, no Brasil. Essa ambiguidade enseja à firma praticar “gerenciamento de resultado” – i.e., ela pode superestimar tais contingências para reduzir a distribuição de dividendos e se capitalizar, ou (mais frequentemente) subestimá-las para atrair o investidor ou maximizar o bônus de desempenho dos gestores. De fato, a transparência (e a provisão) de passivos contingentes é considerada insatisfatória no Brasil (Jesus & Souza, 2016). Isso pode ser confirmado contrastando as informações contábeis das firmas com outros documentos que avaliam a possibilidade de prejuízos decorrentes de danos ambientais ou ações judiciais; se uma firma adota medidas para mitigar determinado risco nessa área, é porque os gestores consideram *ao menos possível* que esse risco resulte num passivo, donde ele deveria ser mencionado entre as contingências. Durante o julgamento do processo sobre planos econômicos (ADPF 165) no Supremo Tribunal Federal, os respectivos bancos argumentaram que sua condenação poderia resultar em risco sistêmico, dada a magnitude das perdas (Ferreira, 2013) – mas isso era incompatível com o nível dos passivos contingentes informados em suas demonstrações financeiras.

Retornemos, pois, ao argumento de Shrader-Frechette sobre situações caracterizadas como catastróficas: são exemplos de cenários negativos que todos reconhecem que se devem evitar – permitindo a adoção de medidas preventivas justificáveis por parte da autoridade. E mesmo que uma sociedade que se dedique à prevenção desses cenários comprometa parte do bem-estar de alguns cidadãos, esse é um preço aceitável – já que o contrário implicaria arriscar o bem-estar dos que estão em pior situação<sup>81</sup>.

### 3 Greenspan: um estudo de caso de decisão sob incerteza

Alan Greenspan foi *Chairman* do Federal Reserve System (FED), o banco central americano, de 1987 a 2006; após sua saída, foi acusado de ser um dos responsáveis pela grave crise financeira de 2007-2008, a qual levou à chamada “Grande Recessão” – cujos efeitos econômicos se alastraram pelo mundo, com impacto especial na crise da Dívida Soberana da Zona do Euro (Krugman, 2009).

As causas de crises bancárias e do “ciclo de negócios” são conhecidas dos economistas, embora não haja consenso sobre o que fazer a respeito (Scheffczyk, 2016, 266). No caso da crise de 2007-2008, a expansão monetária (justificada para contrapor os efeitos negativos sobre a economia do 11/9 e da “bolha do *dot-com*”) e incentivos governamentais ensejaram uma bolha imobiliária, financiada por um amplo mercado não regulado (o *shadow banking*) de títulos lastreados em créditos imobiliários (como MBS – *mortgage-backed securities*), por sua vez garantidos por derivativos como CDS (*credit default swaps*). A alta alavancagem (Greenspan, 2013) e a regulação assaz permissiva (Stiglitz, 2010a) teriam tornado o sistema bancário vulnerável à crise imobiliária; a complexidade das operações e de sua contabilização dificultou a avaliação dos riscos por parte de analistas e reguladores (McMillan, 2014, 56 e s.). Finalmente, após a quebra do Lehman Brothers em setembro de 2008, governo precisou prover liquidez ao sistema e investir em instituições sob risco de falência (o famigerado *bail-out*), enquanto a economia entrou em recessão.

---

<sup>81</sup> Pode ser desnecessário, mas cabe deixar claro que esse raciocínio não pode ser extrapolado para justificar um controle autoritário, incompatível com uma sociedade liberal; afinal, pelo mesmo argumento da posição original, as partes escolheriam o princípio de igual liberdade e o princípio de iguais oportunidades – os quais têm precedência sobre o princípio da diferença. Uma sociedade autocrática pode ser vista como o primeiro risco contra o qual as partes da posição original buscariam se garantir.

Teria sido, pois, muito mais econômico<sup>82</sup> adotar medidas restritivas, como aumentar requerimentos de capital dos bancos, ou mesmo – mais impopular – elevar as taxas de juros, a fim de evitar a crise ainda em 2005; em especial, teria sido melhor para a multidão de menor poder aquisitivo (Stiglitz, 2010b). Por que Greenspan não o fez? Afinal, Bancos Centrais são autoridades *regulatórias* e *monetárias*: seu principal papel é manter a *estabilidade* econômico-financeira, por meio do controle da quantidade de moeda na economia, de operações com títulos públicos - além da regulação, da supervisão e da provisão de serviços ao sistema bancário.

A defesa de Greenspan é que *não poderia* saber – segundo ele, a Crise "representou uma crise existencial para a previsão econômica" (Greenspan, 2013, 10). Ninguém pode prever quando uma “bolha” começa e quando termina<sup>83</sup>, mesmo que saibamos que elas podem ocorrer (Mandelbrot & Hudson, 2008); e, como o mercado conhece melhor seus próprios riscos do que o regulador (uma clássica assimetria de informações), a intervenção deste (salvo para garantir o respeito à lei) seria injustificada. Sua posição condiz com a recomendação de Huemer (2012) pela passividade – e mesmo com a análise de custo-benefício sancionada pelo *Bayesianismo*: afinal, as medidas anticíclicas também teriam um custo, e a avaliação anterior a 2007 é que era extremamente improvável que a crise tivesse tal magnitude. Greenspan “confiava no interesse próprio das instituições financiadoras para proteger o capital dos acionistas”; e considerava que uma regulamentação e supervisão mais restritiva não era uma boa alternativa – pois “raramente leva a uma melhora e [...] chega a ser contraproducente, quando excessiva” (Greenspan, 2013, 89).

### 3.1 Incentivos, coordenação e segurança

Ainda que as informações do regulador fossem *apenas* um subconjunto das informações do regulado (i.e., que a assimetria de informações fosse unilateral), isso não significaria que aquele não poderia atuar sobre os incentivos deste. Há aqui um problema de coordenação

---

<sup>82</sup> Apenas o *bail-out* teria custado U\$ 700 bilhões ao governo americano (Stiglitz, 2010b, 122). Mas isso não representa o impacto sobre o crescimento econômico; nos EUA, o crescimento anual médio do Produto Interno Bruto foi de 2,99% entre 1998 e 2007, mas de apenas 0,73% nos cinco anos seguintes (2008: -0,31%; 2009: -3,1%). O crescimento do PIB mundial foi de 4,3% em 2006 e 4,2% em 2007, despencando para 1,7% em 2009. Dados do Banco Mundial, disponíveis em: <<https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG>>.

<sup>83</sup> No entanto, Rajan (2006), à época Diretor do Departamento de Pesquisas do FMI, apontou para os riscos da securitização e a possibilidade de uma crise sistêmicas - mas, como outras "Cassandras", foi considerado exagerado. Da mesma forma, Roubini (2006) defendeu que o FED deveria intervir sobre a bolha imobiliária, inclusive elevando a taxa de juros. Parte do problema aqui é o "desacordo entre pares" (*peer disagreement*); Greenspan (2013) reclama que sempre há pesquisadores alertando que "o mundo está por desabar" – mas que há pelo menos o mesmo número de estudiosos defendendo o contrário.

análogo ao do dilema do prisioneiro: executivos de companhias não têm incentivos para, unilateralmente, classificar os *ativos que já possuem* como de maior risco, pois implicaria uma desvantagem competitiva para a empresa: “as instituições temiam que, caso se retraíssem cedo demais, quase certamente perderiam uma fatia do mercado, talvez de maneira irrecuperável” (Greenspan, 2013, 71); nem se pode esperar que auditores privados o façam – porquanto só avaliam *uma firma de cada vez*, eles dificilmente demandam que o cliente adote uma postura mais conservadora do que a do restante do mercado. A situação é agravada pelo fato de que as normas regulamentares frequentemente incentivam os investidores a uma “terceirização cognitiva” da avaliação do risco dos ativos para agências de *rating*<sup>84</sup> –comprometendo a responsabilidade epistêmica daqueles; “Treating triple A rated structured debt securities as though they had a yield curve commonly associated with triple A rated corporate bonds was, as Philippe Jorion states, an ‘act of blind faith in the credit rating’”(Bruin, 2015, 171).

Isso enseja a dinâmica da bolha: o valor do ativo tende a continuar subindo até que um conjunto de investidores se desfaça de suas posições (com medo de que outros ajam antes), precipitando a queda. Reguladores, pois, teriam o dever de demandar que o conjunto de regulados aumente sua percepção do risco nesse caso, quando tiverem dúvidas – ao contrário do que sugere Greenspan (2013, 112):

Um século atrás, os inspetores podiam avaliar empréstimos individualmente e emitir um juízo sobre sua solidez. Mas no ambiente global de empréstimos contemporâneo, como um inspetor bancário americano pode avaliar a qualidade creditícia, digamos, de um empréstimo a um banco russo, e, por conseguinte, a carteira de empréstimos desse banco? Isso, por sua vez, exigiria avaliar as contrapartes do banco russo e as contrapartes dessas contrapartes, tudo para verificar a solidez de um único empréstimo. Em suma, nem um inspetor bancário nem uma agência de classificação de risco são capazes de formar um juízo. Para que se dê uma certificação, qual a profundidade necessária, nas infindáveis camadas de inspeção?

---

<sup>84</sup>“Credit securities and credit derivatives were often traded many times, between multiple institutions; they seemed to offer what had previously been taken to be an impossible combination: high returns and safe investments” (Herzog, 2017, 10). No entanto, a impossibilidade dessa combinação é um dos corolários da microeconomia ortodoxa – a correlação entre risco e retorno; por isso, Lo (2012, 2) refere que os verdadeiros discípulos da HME deveriam ser menos suscetíveis à crise, dada a correlação positiva entre risco e remuneração: se um título *x* dá uma taxa de retorno muito superior à de um título *y*, com o mesmo *rating* e a mesma maturidade, então ou HME é falsa, ou o *rating está errado*. Afinal, por que o emissor de *x* ofereceria uma taxa maior?

É difícil imaginar que um auditor utilizasse tal exemplo<sup>85</sup>; afinal, se uma instituição concedesse um empréstimo a um cliente qualquer, sem ter informações sobre a capacidade de pagamento deste, classificar-se-ia a operação como arriscada. Por que deveria ser diferente para o caso do banco russo?

Só se pode *confiar* que "as instituições sabem o que fazem" se elas forem capazes de o demonstrar – de apresentar informações justificando suas avaliações de risco; se não temos informação para julgar, então a conclusão *default* é que a operação é arriscada. Aplica-se a tal relação entre supervisores e supervisionados uma lição da teoria do *design* de mecanismos: pode haver restrições relativas a “incentivos informacionais” (elas podem ter incentivos para omitir informações) ou a “incentivos estratégicos” (elas podem ter incentivos para assumir mais riscos do que seria socialmente ótimo); essas restrições equivalem a custos transacionais que comprometem o papel de comunicação e coordenação que economistas liberais atribuem ao mercado (Myerson, 2008, 339).

Embora Greenspan conceda que reguladores e supervisores tenham a função de assegurar o cumprimento de normas legais (*law-enforcement*) e que podem emitir normas e orientações “que não exijam [...] a capacidade de prever com precisão um futuro incerto”(Greenspan, 2013, 103-04), adotar uma atitude cética e conservadora *não* demanda essa capacidade preditiva *precisa* – embora demande conhecimento especializado e a obtenção da maior quantidade de informação possível. Normas contábeis e práticas de supervisão têm um efeito “performativo” (Walter, 2012, 41-42; 2015, 25-26) – levando-nos de volta ao problema da “reflexividade” da ciência social (seção 1.1.1 supra): elas não apenas definem padrões para a cognição (para juízos justificados sobre o patrimônio de um agente), mas também *incentivam* comportamentos de gestores e investidores<sup>86</sup>.

---

<sup>85</sup> Greenspan o introduz pouco depois de narrar sua experiência pessoal em bancos de investimentos como banqueiro e economista – mostrando que teria conhecimento prático do assunto; mas essa demonstração sugere que é presa de um viés chamado por Stiglitz (2012, 56) de *captura cognitiva*: apela antes para o ponto de vista da área de negócios (que tem menor aversão à incerteza) de uma instituição financeira que para a área de controle e gerenciamento de riscos (a qual é enviesada em relação à aversão à perda). Isso pode ser associado ao risco, apontado por Tucker (2018, 17) como um dos contribuintes da Crise, de que Bancos Centrais privilegiem áreas de maior prestígio (como política monetária), em detrimento das de menor visibilidade (como supervisão).

<sup>86</sup> Isso remete ao afastamento, nas últimas décadas, em relação a normas contábeis de avaliação de ativos associadas ao princípio da prudência e a estimativas baseadas na série histórica (um fator observável, objetivo – ainda que epistemicamente limitado), em favor do *fair value*; o que levou a situações inusitadas antes da crise: o banco Goldman Sachs (credor) estimou em \$5,1 bi o valor da exposição em derivativos da AIG (devedor), enquanto esta avaliou a **mesma** posição em \$1,5 bi – e ambas as análises foram avalizadas pela auditoria externa da PricewaterhouseCoopers – PwC (Ford & Marriage, 2018). Ou seja, não apenas as instituições usaram modelos distintos (avalizados pela mesma auditoria) para precificar um conjunto idêntico de ativos, mas elas utilizaram modelos que otimizavam suas respectivas posições.

Isso leva a um segundo ponto: *se* as autoridades de regulação e de supervisão *adotam* uma postura mais cética em relação ao mercado (se não terceirizam sua responsabilidade epistêmica para o supervisionado e as agências de *rating* e de auditoria), então a assimetria de informações pode ser bilateral. Mesmo que o supervisionado esteja, via de regra, numa condição privilegiada para avaliar os riscos e benefícios de sua própria empresa, e que mercados abertos agreguem essas informações privadas, o regulador pode ter informações que o supervisionado não tem sobre o ambiente da atividade. No caso da supervisão financeira, a autoridade tem, e.g., acesso a informações privadas de todos os supervisionados – podendo, e.g., comparar seus modelos de negócio, de gestão de risco, e mesmo as classificações de risco de seus clientes. No caso de Bancos Centrais, também pode ter uma informação macroeconômica especial: os bancos podem estimar qual será a taxa de juros no futuro, mas a autoridade monetária o *decide*. Autoridades fornecem não apenas a oportunidade de processar e transmitir tais informações (nisso, frequentemente, são inferiores a mercados), mas de tomar decisões e produzir coordenação.

Ainda que possa valer a pena, do ponto de vista individual da instituição financeira (ou de seus gestores), assumir um contrato *incerto* (mormente porque as perdas da operação são limitadas ao capital em jogo), cabe prevenir que essa prática se espalhe e resulte num problema para o mercado como um todo e numa externalidade social negativa grave (uma crise financeira). Podemos, pois, retomar nosso argumento a partir da aplicação do *maximin* e do princípio da diferença na prevenção de catástrofes: considerando a gravidade que uma crise financeira tende a acarretar aos cidadãos em pior situação, a autoridade está justificada em adotar medidas para preveni-la, mesmo quando houver incerteza a respeito – ela está justificada inclusive em aplicar medidas para *evitar* ou *reduzir* a exposição social a essa incerteza.

### 3.3 Autoridade, autonomia e representação

Uma discussão mais polêmica, agora na atuação de Bancos Centrais como autoridades monetárias, é se devem "mirar" no preço de ativos, usando a política monetária para evitar bolhas. Para Greenspan, mesmo que o banco central consiga identificar uma bolha, ele não tem *legitimidade* para atuar sobre ela; ele pode, claro, demandar que as instituições reguladas sejam

---

Ninguém parece negar a existência de tal efeito performativo; no entanto, é raro que os teóricos extraiam daí a conclusão (como o faz Walter, 2012, 41; e MacKenzie, 2006): que a escolha de modelos de precificação e avaliação de ativos não é uma decisão sobre um instrumento para representar a realidade, mas a adoção de uma *estratégia* num complexo jogo com outros participantes do mercado.



robustas ao “estouro” (i.e., editar normas visando a reduzir a alavancagem) e, após esse sinistro, adotar as medidas monetárias necessárias para evitar o prolongamento da recessão (expansão monetária e provisão de liquidez ao sistema). Mas, uma vez que não se sabe quando tal bolha vai estourar, e se contaminará a economia real, e dados os potenciais custos macroeconômicos de tal política monetária (como o risco de induzir uma recessão), essa medida não é aconselhável: implicaria usar recursos do contribuinte para resolver um problema interno ao mercado financeiro (Greenspan, 2013, 143-44)<sup>87</sup>.

Contudo, se é adequado adotar uma atitude *quase*-paternalista em relação à alavancagem, ou mesmo prover a garantia da estabilidade financeira no caso de crise, por que não seria legítimo atuar para prevenir o sinistro? Afinal, "It is inconsistent and non-optimal to argue that monetary policy should react to bursting bubbles but not to rising bubbles" (Roubini, 2006, 100). A resposta não pode apelar para a "ladeira escorregadia" de que isso implicaria intervir em qualquer bolha: mesmo que a distinção entre bolhas com e sem potencial de afetar a economia não seja clara, isso não significa que ela não exista; nada impede que a autoridade "cruze os braços" diante de uma situação que certamente não afeta a economia<sup>88</sup> (como, até o momento, seria o caso das *criptomoedas*), mas não nas que julga afetarem a economia. Outrossim, não se pode objetar que prevenir bolhas implicaria risco moral, pois a atuação *ex post facto* (seja com o *bail out*, seja garantindo liquidez) o implica em ainda maior grau:

---

<sup>87</sup> Em defesa de Greenspan, pode-se dizer que o FED atuou adequadamente para prover a liquidez necessária ao mercado nos *crashes* financeiros ocorridos durante seu mandato (e.g., o *crash* de 1987, a bolha *dot-com* de 2000, as crises de 98 e 99 – nos mercados asiáticos e na Rússia), em que não houve contágio do sistema bancário. Para ele, porém, o papel do FED é não de prevenir bolhas, mas evitar que elas prejudiquem a economia real: "We as central bankers need not be concerned if a collapsing financial asset bubble does not threaten to impair the real economy, its production, jobs, and price stability. Indeed, the sharp stock market break of 1987 had few negative consequences for the economy".

Para Krugman (2009), justamente o otimismo resultante da “Grande Moderação” pode explicar a *hubris* que impediu as autoridades de adotar uma estrutura regulatória adequada e se antecipar à crise de 2007 (Krugman, 2009); ou, como o próprio Greenspan coloca: "What do we get by being very successful in forecasting? And I say a bubble. In fact I think the evidence probably is conclusive that a necessary and sufficient condition for a bubble is a prolonged period of economic stability, stable prices, and therefore low risk spreads, credit-risk spreads" (Fox, 2014).

No entanto, uma hipótese que não encontrei na literatura é a da falta do "efeito halo" da figura de Greenspan, o *maestro*, em 2008; talvez a crise não houvesse se espalhado tão rapidamente se o mercado tivesse maior confiança no presidente do FED (embora Bernanke, o sucessor, seja um economista respeitado, ele não era “o maestro”). Não faltam anedotas sobre como sutilezas da comunicação e da psicologia social tem efeitos sobre mercados financeiros: durante o *crash* de 1987, a Casa Branca sugeriu que Greenspan desmarcasse seu primeiro discurso como Presidente do FED em Dallas, a fim de permanecer em Washington, perto da "ação"; Greenspan manteve a agenda e a calma durante todo o período, sob o argumento de que alterar seus planos passaria a "mensagem errada" aos investidores (Noble, 1996).

<sup>88</sup> Mesmo quando se trata de regulação e supervisão, adotam-se regras e procedimentos distintos para empresas e atividades com riscos diferentes. Não se deve, pois, confundir o argumento de Roubini com a "tese-espantalho" (alvo de Posen, 2006) de que a prevenção de bolhas não acarretaria custos e riscos e de que a política monetária pode ser um substituto à manutenção de uma estrutura regulatória que garanta a estabilidade financeira.

"exactly because liquidity insurance should be properly used in times when systemic risk is acute, it is necessary to avoid distorting investors' incentives in good times" (Roubini, 2006, 102); mesmo Greenspan (2013, 145 e 276) considera o fato das instituições "grandes demais para falir" uma séria ameaça fiscal e econômica. A decisão sobre se vale a pena intervir, portanto, é uma decisão sobre qual a política monetária ótima; e, após a crise, Bancos Centrais e governos adotaram políticas monetárias não convencionais, numa escala até então inédita.

O argumento de Greenspan torna-se mais atraente, porém, quando nos perguntamos sobre quem teria legitimidade para tomar tal decisão e sobre suas consequências políticas: aumentar requerimentos de capital não é uma medida considerada particularmente gravosa, e é pouco controversa (do ponto de vista ideológico); prover liquidez ao mercado, *custe o que custar* (como afirmou Mario Draghi num célebre discurso - Heijden, Beetsma & Romp, 2017), num cenário extremo, é visto como uma medida especial e atrai poucas pressões e suspeitas – todos podem reconhecer que se trata de uma exceção. Mas o mesmo não vale para decisões macroeconômicas impopulares, como, e.g., elevar taxas de juros *antes* do sinistro ocorrer. Nessa linha, Greenspan (2013, 130) lembra a constante ameaça ao FED por parte da pressão política; pode-se conjecturar que esta será tanto maior quanto mais (a) epistemicamente duvidosos forem os fundamentos da decisão (como é o caso diante de incerteza sobre o futuro), (b) mais alternativas houver, e (c) maior o impacto potencial das medidas adotadas.

Uma autoridade podem ser vistas como um nó central que processa informações e toma decisões dentro de uma rede de cooperação – o que permite "ganhos de escala" obtidos pela centralização; mas isso também permite ver o tipo especial de *risco* criado pela existência de autoridades – redes dependentes de um nó central têm o problema do "ponto único de falha" (*single point of failure*), que as torna vulneráveis a ataques e ao comprometimento desse nó<sup>89</sup>. No caso da autoridade, esse "comprometimento" corresponde à extrapolação de sua competência (e, no limite, à tirania), à corrupção e à ineficiência – problemas que ameaçam a continuidade dessa rede<sup>90</sup>. Bancos Centrais não são autoridades eleitas; no entanto, gozam de

---

<sup>89</sup> Por isso a tecnologia de *blockchain* é considerada promissora para redes com informações sensíveis – a exemplo do sistema de pagamentos (Burgos et al., 2017). Vários Bancos Centrais estudam a possibilidade de adotar a tecnologia de contabilidade distribuída (*distributed ledger*), mas ainda não se tem notícia de seu emprego efetivo; em especial, parece haver um *trade-off* entre custos de processamento (aí incluso o tempo), privacidade (deseja-se mais sigilo do que o uso de "pseudônimos", como é o caso do *bitcoin*) e segurança (Del Río, 2017).

<sup>90</sup> Podemos associá-lo ao que Fukuyama (2011, 312 e s.) chamou de o "problema do mau imperador", decorrente da centralização da organização política sem, em contrapartida, as limitações decorrentes do estado de direito (*rule of law*) e prestação de contas (*accountability*); não apenas a autoridade não tem incentivos para investir na satisfação do público, como os tem para abusar de seu poder. Sem essas limitações, a centralização tende a degenerar no que Acemoglu & Robinson (2013) chamam de *instituições extrativas*, atraindo disputas de poder entre oligarcas que visam a benefícios à pessoa da autoridade e a seu sub-grupo.

maior autonomia (se não de direito, ao menos de fato – pois os governos não querem correr o risco de interferir nesses órgãos e gerar instabilidade) em relação ao poder político do que outras instituições com funções análogas – como, e.g., reguladores setoriais, como os de telecomunicações. O impacto de sua atuação sobre a economia pode ser alto, imediato e difícil de prever – com consequências para o contribuinte e para a distribuição de renda e riqueza. Por um lado, sua atividade implica guardar e processar um conjunto imenso de informações importantes, e reagir a elas; por outro, demanda-se que atue com cada vez mais transparência – que tanto as decisões macroeconômicas quanto as regulatórias não acarretem grandes surpresas, e que sejam publicamente justificadas<sup>91</sup>.

O problema é análogo ao do bordão *quis custodiet ipsos custodes*: agências autônomas (ou independentes – não precisamos fazer a distinção aqui) existem para diminuir a incerteza relativa à atuação dos regulados; há questões graves de política pública em jogo, e é necessário estabelecer um “compromisso crível” de que certos objetivos serão perseguidos de maneira técnica e estável, sem captura pelo mercado, seminterferência política imediata, e sem a necessidade de aguardar uma demorada decisão parlamentar. Mas isso implica novos riscos, causados pela concentração de poder e de informação necessária para desenvolver suas atividades:

There appears to be a paradox here. On the one hand, delegation is designed to help the democratic state deliver better results by sticking to the people’s purposes: in that sense credible commitment is enabling of democratically generated purposes. On the other hand, the people have to remain free to change their purposes. The resolution has to be either that there are some commitment problems where democracy, as ordinarily understood, should be suspended or, alternatively, that an institution designed to enable credible commitment cannot be absolute. (Tucker, 2018, 222)

Em especial, há uma questão de confiança – análoga à discussão sobre os poderes dos juízes na interpretação da lei<sup>92</sup>: os cidadãos particulares, em nome e no interesse de quem essa

---

<sup>91</sup> Apesar da ênfase em transparência na área econômica, observada ao menos desde a Lei de Acesso à Informação (Lei nº 12.527/2011), as apresentações utilizadas nas reuniões do Comitê de Política Monetária do Banco Central do Brasil eram classificadas como *reservadas* até 27 de maio de 2018 – sob o argumento de que “a divulgação imediata das apresentações no curso das reuniões do Colegiado pode introduzir, indeliberadamente, significantes limitações na capacidade de coordenação das expectativas dos agentes econômicos”. Fonte: <<http://www.bcb.gov.br/acessoainformacao/DESCCLASSIFICADASRESERVADAS.pdf>>

<sup>92</sup> Para Shapiro (2011), a aplicação do direito é uma forma de realizar e interpretar algo afim a um planejamento conjunto (ainda que vaga e descentralizadamente), e os poderes da respectiva autoridade são definidos pela respectiva *confiança* com que se pode esperar que ela desempenhe essa tarefa: “Roughly speaking, the Planning Theory demands that the more trustworthy a person is judged to be, the more interpretive discretion he or she is

autoridade age, podem se ressentir de quão pouco são representados por ela (que não elegeram, e que pouco parece responder a suas expectativas); seus representantes eleitos, alinhando-se aos anseios de seus representados, tendem a pressioná-la – diante do quê, investidores tendem a responder com desconfiança, sub-investimento e fuga de capitais.

O que está em questão, pois, é um *trade-off* entre valores tão distintos (e parcialmente conflitantes) quanto eficiência, representatividade, *accountability* e transparência; mas solucioná-lo não é a proposta desse texto<sup>93</sup>, que visa apenas a refutar a objeção cética e defender o princípio da diferença como complemento à teoria da decisão *interpersonal* sob incerteza – o que justifica tomar as medidas necessárias para prevenir crises. Mas concedemos parcialmente o ponto de Greenspan: ainda que defendamos que uma autoridade executiva possa chegar à conclusão de que é racional intervir (e que ela tem discricionariedade, no sentido jurídico, para tanto), há espaço para discutir se *ela* é *a* autoridade politicamente legítima para tal intervenção, ou mesmo se é politicamente conveniente que *ela* a adote.

#### 4 Conclusão

Começamos com a *objeção cética* de Huemer (2012); vimos que, embora Huemer tenha razão em chamar atenção para a incerteza que permeia a ação social, isso, por si só, não é suficiente para justificar seu elogio da passividade (e o respectivo argumento libertário). Arguimos que a objeção cética é *pervasiva*: qualquer teoria sobre como devemos agir deve lidar com o problema da decisão sob incerteza – sob pena de tal teoria ser inaplicável.

O problema, na realidade, é que há mais de um candidato a complemento da teoria da decisão sob ignorância. Defendemos que, em situações de incerteza com riscos sociais, a adoção do *maximin* permite justificar a ação a todas as partes envolvidas; além disso, como notado pelo próprio Harsanyi, esse princípio é mais fácil de aplicar que o princípio de utilidade, pois tem uma demanda informacional menor – é epistemicamente mais simples identificar e evitar os piores resultados. Donde nossa crítica à inércia de Greenspan em relação à bolha imobiliária e ao sistema financeiro americano; mesmo sabendo que o mercado é suscetível a bolhas e a

---

accorded; conversely, the less trusted one is in other parts of legal life, the less discretion one is allowed. Attitudes of trust are central to the meta- interpretation of law, I argue, because they are central to the meta- interpretation of plans — and laws are plans, or planlike norms” (Shapiro, 2011, 331).

<sup>93</sup> Nem o é a questão sobre o desenho do ambiente regulatório. Uma das reclamações após a crise, por exemplo, é que a existência de diversos tipos distintos de reguladores tornou o trabalho desses ainda mais complexo e permitiu às instituições praticar arbitragem regulatória – i.e., adequar seu modelo de negócios à regulação mais fácil de seguir.

crises, ele achava que a autoridade pouco poderia fazer a respeito – uma vez que sempre conheceria menos do que o mercado; argumentamos em sentido contrário: essa atuação pode ser necessária para que os participantes do mercado conheçam seus próprios riscos e respondam de maneira condizente.

Podemos arriscar uma generalização aqui: mercados são “eficientes” no sentido informacional *porque* há investidores que não confiam *exclusivamente* nele, que pesquisam novas informações a fim de encontrar oportunidades de arbitragem; algo análogo ocorre quando acrescentamos agências de *rating* e autoridades ao sistema – ele só funciona adequadamente se tais agentes *não* terceirizam suas responsabilidades, confiando cegamente nas avaliações dos demais.

No entanto, Greenspan apresenta um argumento distinto (e mais forte) em favor da auto-restrição quando aborda sua experiência prática na relação entre a macroeconomia e a política; um dos pontos a pesar, portanto, é a competência legítima e o impacto político e econômico de uma intervenção do FED visando a conter uma bolha. Conquanto essa seja uma das discussões mais importantes – em macroeconomia e em teoria da regulação, ao menos – após a crise financeira, não temos como ser conclusivos a esse respeito, mormente porque foge ao escopo deste texto. Destaquemos o limite desse escopo: uma aplicação do *princípio da diferença* a uma classe bastante restrita de eventos (catástrofes evitáveis, como crises financeiras). Há uma explicação para isso: o problema da tomada de decisões sociais sob incerteza não deveria ser tratado sob uma teoria abstrata e geral (como a da conclusão libertária pela *passividade*) que nos impeça de aprender com erros passados e de analisar idéias alternativas de desenho institucional. Podemos extrapolar a analogia de Huemer (2012) com a medicina: os profissionais da saúde juram “não fazer mal” – mas continuam tomando decisões sob incerteza e aprendendo com erros passados; a diferença é que hoje o fazem de maneira um pouco mais aberta e transparente.

## Capítulo 5: A Crise Moral: a responsabilidade de administradores de instituições financeiras e o argumento da inevitabilidade

**Sumário:** 1 Introdução: a explicação cognitiva e a explicação moral da crise financeira. 2 A crise inevitável. 2.1 A ideia-alvo: justificação a partir do inevitável. 3 Krishna, Shakespeare e Tuld: do fatalismo à teoria da decisão racional. 3.1 Resultado e considerações de responsabilidade pessoal. 3.2 Fatalismo e absurdo. 3.3 O paradoxo do agente fatalista. 4 Conclusão.

**Resumo:** Este ensaio argumenta, mediante análise conceitual, contra uma objeção a reprovações dirigidas a financistas após a crise de 2007-8: a ideia de que não poderiam ter agido de outra forma (ao menos não racionalmente) e de que ninguém deve ser responsabilizado por um fato que não poderia ter evitado. Se correta, ela ameaçaria a *justificabilidade* da responsabilidade social da empresa e da responsabilidade jurídica de administradores. Identificada como a “tese da inevitabilidade”, essa objeção é ilustrada por uma análise do filme *Margin Call* (2011) e associada a outras investigações sobre ética e responsabilidade. A tese-alvo decorre de uma confusão entre diferentes noções de responsabilidade (por uma tarefa, por uma decisão, por causar um evento e por repará-lo) e conduz a uma forma incoerente de fatalismo. Sugere-se, por fim, que a invocação da “inevitabilidade” pode ser uma forma de racionalizar a decisão, obscurecendo as razões em contrário.

[...] But their jobs often entailed exploiting others or living off the results of such exploitation. There was individualism but no individual responsibility. [...] society cannot function well if people do not take responsibility for the consequences of their actions. “I was just doing my job” cannot be a defense. (Stiglitz, 2010b)

1 **Introdução: a explicação cognitiva e a explicação moral da crise financeira**

Anos após a quebra do conglomerado financeiro Lehman Brothers, ainda se contam (pelo menos) dois tipos de histórias sobre a “Grande Recessão” (Kling, 2010), a primeira grande crise econômica do Século XXI; segundo a “explicação cognitiva”, a crise foi ocasionada por um ‘problema epistemológico’: os executivos de instituições financeiras (além de reguladores, auditores e o restante do mercado) subestimaram os riscos incorridos e o grau de alavancagem das instituições. Isso foi agravado pela falta de conhecimento público do volume de obrigações “fora do balanço” (*off-sheet balance*) e dos riscos sistêmicos decorrentes da engenharia financeira, como a securitização de créditos hipotecários (que, supostamente, deveria mitigar riscos, ao permitir sua diluição dentro do mercado). Seria, portanto, um caso especial de falha no gerenciamento de risco, que ou poderia ter sido prevenido mediante níveis maiores de capital para absorver possíveis perdas, ou, numa visão mais pessimista, simplesmente não poderia ser prevenido de qualquer forma – afinal, as análises de economistas são incapazes de lidar com o que Taleb & Martin (2012) chamam de ‘cisnes negros’ (uma classe de eventos imprevisíveis com consequências devastadoras)<sup>94</sup>.

Por outro lado, de acordo com o que chamaremos de ‘explicação moral’, a estrutura de remuneração dos executivos de instituições financeiras os teria estimulado a colocá-las em risco excessivo a fim de, no curto prazo, auferir bônus maiores (Stiglitz, 2010a, xx). Trata-se, portanto, de uma falha de governança, e o desastre poderia ter sido evitado se a estrutura organizacional dessas instituições ensejasse maior transparência e controles efetivos em relação às decisões dos gestores, ou se eles houvessem agido de forma mais responsável – adotando, ex., níveis mais prudentes de alavancagem.

A segunda linha de explicação é bastante popular, e tem sido politicamente explorada desde o início da crise. Ela exemplifica a usual crítica de economistas políticos ao fato de que economistas e analistas de mercado, em geral, ignoram restrições morais (vide, e.g., Sen, 1977; Delfim Netto, 2016, A02). Ainda, fundamenta o recente debate na política americana (uma rara convergência entre o programa de democratas e republicanos) sobre a adoção de uma legislação mais rigorosa (como a divisão de grandes bancos e a reintrodução da plataforma Glass-Steagall) no setor financeiro (Weissmann, 2016), bem como, no nível internacional e no Brasil, a ênfase na adoção de princípios de governança corporativa e de responsabilidade social por parte das

---

<sup>94</sup> Isso não significa que Taleb & Martin privilegiem a explicação cognitiva; na realidade, ela seria complementada pela explicação moral. Eles escrevem, e.g., que a “melhor regra de gerenciamento de risco” já escrita é a prescrição, pelo Código de Hamurabi, da condenação do construtor à morte, caso uma casa caia, matando seu morador (Taleb & Martin, 2012, 50). É o que chamam de “heurística da pele em jogo” – i.e., regras que garantam que os responsáveis pela criação e gerenciamento de um risco suportem as consequências do acontecimento.

instituições financeiras (a exemplo da Resolução nº 4.327, de 25/4/2014, do Conselho Monetário Nacional).

Por outro lado, há quem defenda que, mesmo que a “explicação moral” seja verdadeira, isso não justifica a reprovação dirigida aos financistas. Richard Posner (2010, 111), e.g., comenta de forma crítica o clamor público gerado pela crise financeira de 2007-8: “*Calling bankers greedy for taking advantage of profit opportunities created by unsound government policies is like calling rich people greedy for allowing Medicare to reimburse their medical bills*”. Posner chama atenção (corretamente) para a necessidade de que o governo e seus oficiais assumam responsabilidade moral – afinal, é dever dos reguladores estabelecer normas (e dos supervisores, aplicá-las) a fim de evitar, dentro do possível, eventos desse tipo. Mas esse autor parece pressupor que tal dever exclui a responsabilização dos homens de negócios – dos quais só se pode esperar que ajam visando a maximizar seus próprios interesses e os dos respectivos acionistas; desde que observem a legislação, eles podem alegar que estavam apenas “fazendo seu trabalho”.

Nossa hipótese é que essa tentativa de isentar os agentes econômicos dessa reprovação é fundamentada por um tipo de argumentação (chamemo-la de *argumento da inevitabilidade*) mais geral sobre o conceito de responsabilidade: uma vez que um agente acredita, justificadamente, que um fato ocorrerá e que ele não o pode evitar, é permitido contribuir para com esse fato, a fim de colher os respectivos benefícios – donde seria injustificado imputar-lhe responsabilidade pela ocorrência do referido fato. De uma maneira mais sintética: “Se  $x$  não pode evitar a situação  $S$ ,  $x$  não é responsável por  $S$ ”.

A solução desse problema é relevante tanto para uma teoria da responsabilidade social da empresa quanto para a teoria da responsabilidade pessoal (civil, ou mesmo criminal) de administradores. Afinal, se a tese da inevitabilidade fosse correta, a instituição de disposições jurídicas que incorporem a “heurística da pele em jogo” (como a responsabilidade solidária dos administradores por danos causados durante sua gestão, em caso de liquidação da instituição financeira, prevista no art. 40 da Lei nº 6.024/1974, e a punição pelo crime de gestão temerária – art. 4º, parágrafo único, da Lei nº 7.492/1986) e de uma cultura de *business ethics* seria moralmente injustificável – mesmo que constituísse uma maneira eficiente de incentivar comportamentos desejados.

Ao expor a tese da inevitabilidade na seção a seguir, buscaremos ilustrá-la com algumas cenas extraídas do filme *Margin Call* (2011); a partir daí, traçaremos um paralelo com os



comentários de Amartya Sen (2006) em *Ideas of Justice* sobre o diálogo entre Arjuna e Krishna no épico religioso *Baghavad-Gita*. A comparação permite mostrar que a tese da inevitabilidade depende de uma confusão entre diferentes noções de responsabilidade e de alguma forma de fatalismo – como analisado nas seções 3.1 e 3.2. Concluimos com a aposta de que o eventual defensor da tese da inevitabilidade não pretende aceitar a consequência desse tipo de fatalismo: o esvaziamento da própria noção de deliberação.

## 2 A crise inevitável

Duas cenas do filme *Margin Call* (2011) podem ilustrar o que está em jogo: na primeira, executivos de um banco de investimentos de New York debatem sobre o que fazer diante da descoberta de que a volatilidade dos seus derivativos securitizados por hipotecas está prestes a superar os níveis históricos (i.e., o valor de mercado desses ativos tende a cair em breve); considerando a alavancagem da firma, isso os levará à falência. John Tuld (interpretado por Jeremy Irons), CEO da instituição, decide desfazer-se de todos esses títulos o mais rápido possível, mesmo que isso implique contaminar seus clientes (e o mercado em geral), e justifica sua decisão com o consagrado ‘lema do filme’: *there are three ways to make a living in this business: be first, be smarter, or cheat*.

Em outra cena, mais próxima do fim do filme, Tuld é confrontado por Sam Rogers (Kevin Spacey), chefe dos *traders*, o qual, com remorso por ter participado da venda dos ativos (precipitando uma crise financeira e causando a ruína de investidores que confiavam na firma), pede demissão. O CEO responde diminuindo a gravidade do que fizeram: é o mesmo que vinham fazendo há anos; e enumera diversas crises econômicas ao longo da história, aduzindo sua inevitabilidade – tudo que eles poderiam fazer seria aproveitar a oportunidade:

**John Tuld:** It's just money; it's made up. Pieces of paper with pictures on it so we don't have to kill each other just to get something to eat. It's not wrong. And it's certainly no different today than it's ever been. 1637, 1797, 1819, 37, 57, 84, 1901, 07, 29, 1937, 1974, 1987 [...], 92, 97, 2000 and whatever we want to call this. It's all just the same thing over and over; we can't help ourselves. And you and I can't control it, or stop it, or even slow it. Or even ever-so-slightly alter it. We just react. And we make a lot money if we get it right. And we get left by the side of the side of the road if we get it wrong.

Tuld não está sendo cínico (ao menos não completamente<sup>95</sup>); ele não enxerga qualquer razão para recriminação em sua conduta – vide a ênfase dada na dissociação anterior em relação ao ‘trapaceiro’. A firma estaria apenas agindo de acordo com as regras e a ética do mercado, fazendo o necessário, o que qualquer outro faria; e, se isso vier a causar algum infortúnio geral, Tuld não se responsabiliza por isso. A forma de pensar de John Tuld **não** é uma exceção; em boa medida, ele pode ser identificado com a conduta de executivos do Banco Goldman Sachs, o qual, no início da crise de 2008, vendia derivativos atrelados a créditos hipotecários a seus clientes e, simultaneamente, comprava *swaps* apostando contra esses derivativos – o que lhe rendeu um processo da *Securities and Exchange Commission*, encerrado por um acordo de U\$550 milhões (Solomon; Morrison; Wilhelm Jr., 2012).

Tuld não é sequer um ideólogo do livre mercado – mesmo porque sua visão sobre o ciclo de negócios é pessimista. Ele poderia ser como o financista (e filantropo) George Soros, o qual traça uma distinção entre sua atuação como agente de mercado e como um cidadão. Enquanto agente no mercado financeiro internacional, o qual compete com os demais visando a maximizar seus ganhos, Soros é um conhecido especulador, famoso por ter lucrado em ao menos duas crises financeiras de consequências internacionais – a da ‘quarta-feira negra’ da Libra de 1992 e a crise financeira da Ásia de 1997 (o que ele comenta em Soros, 2000, 208-9). Se agisse de outra forma, ele seria suplantado pelos demais agentes<sup>96</sup>. Porém, enquanto cidadão e intelectual, Soros critica o ‘fundamentalismo de mercado’ e defende que o Estado deve regular e supervisionar o sistema financeiro, além de dispor de instrumentos para evitar que a ganância individual leve a situações desastrosas – o que ele considera, atualmente, a maior ameaça à democracia (Soros, 2008, xi-xii). Sua posição, portanto, reflete parcialmente o argumento de Posner, exposto na Introdução supra.

Tanto Tuld quanto Soros estão ‘tirando o corpo fora’: desde que ajam de acordo com as regras do jogo, eles não se consideram responsáveis pelos resultados a que deem causa – em especial porque, caso deixem de aproveitar as oportunidades de lucro, outros o farão (como diz John Tuld, *be the first*), donde a abstenção não teria efeito prático. Ou seja, não é que estejam apenas expondo um argumento a ser sopesado contra outras razões em contrário, como numa

---

<sup>95</sup> Contudo, estamos omitindo uma cena em que Tuld discute com dois executivos subordinados e revela que, na realidade, eles sabiam do risco dos ativos em questão. Isso não altera o cerne da posição de Tuld (i.e., a escusa de que não violara qualquer regra regendo mercado), apenas exclui uma possível alegação de ignorância.

<sup>96</sup> Investidores agem sob uma forma de pressão evolucionária: “*stockbrokers who do less well than their competitors go bust. The rules-of-thumb that stockbrokers use are therefore subject to the same kind of evolutionary pressures as the genes of fish or insects*” (Binmore, 2007, 16-17).

deliberação prática usual; na verdade, essas outras razões<sup>97</sup> sequer são consideradas, donde o próprio conflito moral e o remorso que o personagem de Sam Rogers alega experimentar não teriam razão de ser.

Nesse aspecto, podemos ser tentados a associar Tuld com o personagem de Trasímaco, o ‘grande vilão’ do Livro I da *República* de Platão, para quem a justiça seria apenas o interesse do mais forte (Platão, I, 338c):

[...]o homem justo é em todos os lugares inferior ao injusto. Em primeiro lugar, no comércio, quando se associam um ao outro, nunca descobrirás, ao dissolver-se a sociedade, que o justo ganhou, mas que perdeu; em seguida, nos negócios públicos, quando é preciso pagar contribuições, o justo paga mais do que os seus iguais, o injusto menos[...]. Por isso, Sócrates, a injustiça levada a um alto grau é mais forte, mais livre, mais digna de um senhor do que a justiça e, como eu dizia a princípio, a justiça significa o interesse do mais forte e a injustiça é em si mesma vantagem e lucro.

É claro que Trasímaco defende algo um pouco mais perverso que Tuld, o qual se recusa a trapacear. Mas os dois são semelhantes quando, no cerne de seus respectivos argumentos, apresentam a situação que se avalia moralmente como um cenário determinado: é um **fato**, para John Tuld, que haverá crises, e que os primeiros (os que as precipitam) ‘limpam a mesa’; também é um fato, afirma Trasímaco, que os injustos se saem melhor que os justos. Por isso, Simon Blackburn identifica Trasímaco como o ancestral de uma forma perniciosa de pensar:

No drama platônico, Trasímaco na verdade é o porta-voz dos emissários atenienses<sup>98</sup>. Eles representam os homens maquiavélicos da *real politik*, sabendo que vivem num mundo onde as pessoas se entredevoram adaptando-se a ele. Eles e seus sucessores deixam um longo rastro rubro na história humana[...]. São os ancestrais diretos da *blitzkrieg*, do terrorismo, da adoração ao livre mercado e da ética da escola de comércio. São também os ancestrais diretos do “neoconservadorismo” norte-americano [...].

---

<sup>97</sup> Da perspectiva desse agente, elas sequer são reconhecidas como razões, já que não tem qualquer função na deliberação. A contrapartida “meta-ética” dessa discussão é precisamente se só podemos chamar de razões aquilo que o agente toma como tal (ou que ele tomaria, em seu juízo perfeito) – i.e., o subjetivismo e o internalismo – ou se há razões que não dependem desse reconhecimento pelo agente – i.e., uma teoria objetivista.

<sup>98</sup> Blackburn, ao mencionar os “emissários atenienses”, está referindo o episódio da Conferência de Melos, contado por Tucídides (1874) no Cap. XVII da *História da Guerra do Peloponeso*, em que os emissários atenienses fizeram uma “proposta irrecusável” à cidade de Melos (uma pequena ilha mediterrânea inofensiva): rendição incondicional ou a destruição da cidade. Os Mélicos resistem, e os atenienses cumprem a ameaça: matam todos os homens adultos e escravizam as mulheres e crianças.

No século XIX, Trasímaco recebeu um grande impulso com o advento do darwinismo, que muitas vezes foi interpretado como uma demonstração de que o mundo dos que se entredevoram não apenas é moralmente justificável, mas também de alguma forma inevitável [...]” (Blackburn, 2006, 42-3).

### 2.1 A tese-alvo: o argumento a partir do inevitável

A ideia geral ora atacada é a seguinte: dada a situação fática  $S$  (o mercado de capitais, a vantagem do injusto, o conflito bélico, o estado de natureza, etc.), segue-se inevitavelmente o resultado/estado de coisas  $R$ ; e, *dado que*  $R$  é (relativamente ao agente) inevitável,  $R$  é (moral ou praticamente) justificado – ou, alternativamente, é justificado (ou ‘correto’, ‘melhor’, ‘de direito natural’, ‘permitido’, ‘obrigatório’: um termo normativamente carregado qualquer) contribuir para/agir de acordo com  $R$  –, de modo que as razões que eventualmente houver em contrário são irrelevantes<sup>99</sup>.

Formalizando o raciocínio, pode-se provar que a tese da inevitabilidade leva à conclusão radical de que ninguém é responsável por um evento, a menos que sua ação seja uma condição necessária deste:

- i) Se  $S$  ocorreria sem o agente  $x$ ,  $x$  não é responsável por  $S$ .

**[inevitabilidade]**

- ii) Se  $S$  ocorreria sem o agente  $x$ , também ocorreria sem o agente  $y$ <sup>100</sup>; [i.e., nenhum sujeito é condição necessária para  $S$ ]
- iii)  $S$  poderia ocorrer sem o agente  $x$ . [equivale a “ $S$  é inevitável”]
- iv) **Conclusão:** ninguém é responsável por  $S$  [por (i), (ii) e (iii)].

A conclusão em (iv) é uma consequência válida das premissas; contudo, ela é inaceitável: há diversas situações em que um evento decorre das ações de um conjunto de agentes, sem que qualquer deles seja essencial para que ele ocorra. O que a tese da

---

<sup>99</sup> Isso não nega, óbvio, que o agente tenha uma motivação para produzir  $R$ ; o ponto é se existe outra alternativa racional. Nessa linha, a argumentação de Trasímaco é mais completa do que a de Tuld, pois aquele assume explicitamente a premissa de que o auto-interesse é a única motivação do agente racional – e Sócrates vai atacá-lo mostrando que ou essa teoria da motivação é factualmente incorreta (dando o exemplo do médico, cuja arte visava bem do paciente), ou a concepção egoísta de auto-interesse é (normativamente) incorreta – o que dá o tema da continuação do diálogo, após a intervenção de Trasímaco: a teoria da superioridade da vida do homem justo.

<sup>100</sup> Note como esse “passo indutivo” equivale a “ $S$  poderia ocorrer sem  $x$ , ou sem  $y$ , ou sem  $z$ ...” – i.e., nenhum agente em especial é condição necessária para  $S$ . Não se trata, pois, da generalização universal “ $S$  poderia ocorrer sem  $x$ , e sem  $y$ , e sem  $z$ ...” – o que implicaria “ $S$  poderia ocorrer sem ninguém”.

inevitabilidade acarreta é que tratemos essas situações de forma semelhante a eventos naturais: ninguém é responsável por elas. Imaginemos alguém justificando sua participação num linchamento: “não foi culpa minha; o sujeito morreria de qualquer forma...”. Portanto, só nos resta negar uma das premissas – e a candidata mais plausível é o condicional em (i).

Não são apenas os ‘vilões’ de nossa história intelectual, enumerados por Blackburn, que pensam assim; alguns ‘heróis’ da filosofia moral aceitam o argumento, mas aduzem que, em alguma situação relevante, a premissa (iii) não é verdadeira – i.e., o agente pode e deve fazer algo para evitar *S*. Um exemplo da tradição contratualista é Thomas Hobbes (o qual, tendo traduzido Tucídides antes de escrever suas obras políticas, estava a par do argumento dos atenienses na Conferência de Melos); no capítulo XIV dos *Elements of Law*, ele deduz que, no estado de natureza, todos têm “direito a todas as coisas” (Hobbes, 2002, 95) e que o forte tem o direito de subjugar o fraco (Hobbes, 2002, 96-97), usando como premissa a noção de “necessidade da natureza” (Hobbes, 2002, 94) e uma análise ao estilo da teoria dos jogos<sup>101</sup>. Dessa forma, o estado de natureza levaria a algo que Hobbes (2002, 96) chama de “contradição”: uma vez que o ‘direito de todos a tudo’ coloca os homens em conflito, deixando todos em pior situação assim, quem desejasse viver nesse estado contradiria a si mesmo – razão por que devem fazer o possível para passar ao estado civil.

### 3 Krishna, Shakespeare e Tuld: do fatalismo à teoria da decisão racional

#### 3.1 Resultado e considerações de responsabilidade pessoal

Podemos ilustrar melhor nossa argumentação traçando um paralelo entre a discussão de Sam Rogers e John Tuld com o diálogo entre o General Arjuna, que hesita em tomar partido na guerra entre Pandavas e Kauravas, e seu conselheiro Krishna (na realidade, o avatar do deus *Vishnu*), conforme narrado no *Bhagavad-Gita*<sup>102</sup> – parte do épico religioso hindu *Mahabharata*. Talvez seja sobejo alertar: não nos interessa, evidentemente, fazer uma exegese do hinduísmo ou da filosofia oriental (mesmo porque não é de nossa competência), nem sugerimos que

---

<sup>101</sup> Não à toa, os teóricos dos jogos apelidaram algumas situações beligerantes, nas quais os agentes atacam por medo de serem atacados (os “dilemas de Schelling”), de “armadilhas hobbesianas” (v. Baliga & Sjöström, 2012).

<sup>102</sup> Uma vez que se trata de um texto clássico, indicaremos a referência ao *Bhagavad-Gita* por BG, seguido da informação do capítulo e dos respectivos versos.

advogados, economistas ou outros cientistas sociais devam fazê-lo a fim de compreender ou criticar a atuação de financistas; nossa analogia visa a comparar as propriedades formais das duas discussões.

A interpretação clássica da obra é que se trata de um confronto entre uma concepção consequencialista de moralidade, professada por Arjuna (o qual lamenta o sofrimento que a guerra trará), e a ética do dever defendida por Krishna. Mas essa exegese tradicional de Krishna como “o deontologista máximo”, focado apenas no dever, e de Arjuna como um “consequencialista típico”, constituiria uma interpretação assaz estreita, segundo Sen (2009, 216), na medida em que obscurece as considerações de Arjuna sobre a ideia de que uma pessoa deve assumir responsabilidade especial pelas consequências de suas ações:

The issue of responsibility is central to the debate between Arjuna and Krishna, though the two present quite different interpretations of how Arjuna’s responsibilities should be seen. Arjuna argues that the results of one’s choices and actions must matter in deciding what one should do, whereas Krishna insists that one must do one’s duty no matter what happens, and that the nature of one’s duty can be determined, as in this case, without having to examine the consequences of the chosen actions (Sen, 2009, 213).

Para Krishna, uma vez que o dever de um agente está plenamente determinado pelo seu papel e por seu destino, a responsabilidade desse agente pode ser definida simplesmente em função do cumprimento desse dever. Analisa-se primeiro se o agente observa a conduta devida; se sim, ele não pode ser reprovado pelas consequências decorrentes (as quais não desempenham papel direto na determinação desse dever).

Essa é **uma** noção de responsabilidade. Segundo Dworkin (2011), há pelo menos quatro noções distintas (mas compatíveis) de responsabilidade de um agente – responsabilidade causal (por causar um evento), “decisional” (*judgemental* - i.e., a responsabilidade que toda pessoa assume ao tomar uma decisão), por uma tarefa (*assignment*), e a responsabilidade pela reparação de um evento (*liability*):

A person is causally responsible for an event, we say, if some act of his figures (or figures substantially) in the best causal explanation of that event. I would be causally responsible for an injury to a blind beggar if I shoved him to steal his money or collided with him absentmindedly or while drunk or deranged or even just accidentally. But not when someone else has pushed me into him, because then no act of mine has contributed to the injury. (My body is part of the causal chain, but I am

not.) Someone has assignment responsibility for some matter if it is his duty to attend to or look after it. The last person to leave a room, we say, is responsible for turning off the lights, and the sergeant is responsible for his platoon. Someone has liability responsibility for an event when he is required to repair, compensate for, or absorb any damage flowing from the event. I have liability responsibility for the damage I cause by my careless driving; an employer may have liability responsibility for any damage his employees cause. Causal, assignment, and liability responsibility are all, finally, to be distinguished from judgmental responsibility. Someone has judgmental responsibility for some act if it is appropriate to rank his act on some scale of praise or criticism. I have judgmental responsibility for walking past the beggar, giving him nothing, but not for the harm when someone else pushes me into him. These different senses of relational responsibility are conceptually independent: an employer may have liability responsibility for damage caused by his employees' negligence even though he is neither causally nor judgmentally responsible for that damage (Dworkin, 2011, 103).

Arjuna está referindo o que poderíamos chamar 'responsabilidade causal' – a relação causal entre o agente e o evento<sup>103</sup>; ao passo que Krishna privilegia a noção de responsabilidade como tarefa ou atribuição (*assignment*) – a ideia de que analisar a responsabilidade moral do agente implica avaliar se cumpriu um determinado papel (o que independe das consequências e dos antecedentes de sua ação).

A tensão entre os dois pontos de vista só é resolvida, em definitivo, no Livro XI, quando Krishna, após comover Arjuna com revelações sobrenaturais (do universo como um todo), mostra seu aspecto temível e destrutivo da *Vishvarupa* (a famosa “forma universal” de múltiplos braços) a devorar os soldados que combaterão na guerra (BG, 11: 26-27) e demanda que o general se submeta a seu destino:

**Lord Krishna:** I am death, the mighty destroyer of the world. I have come here to destroy all these people. Even without your participation in the war, all the warriors standing arrayed in the opposing armies shall cease to exist.

---

<sup>103</sup> Essa noção de responsabilidade causal também não é isenta de problemas, mormente quando se tem em conta a diferença entre fatores que aumentam o risco relativo de um evento (“carne processada causa câncer”), fatores que aceleram sua ocorrência (catalizadores ou *hastenings*), fatores que o precipitam (a “última palha que quebra as costas do camelo”). Isso não nos concerne, contudo, já que, nos exemplos usados, o agente é um caso paradigmático de responsável causal.

Therefore, get up and attain glory. Conquer your enemies, and enjoy a prosperous kingdom. I have already destroyed all these warriors. You are a mere instrument, O Arjuna<sup>104</sup>. (BG 11:32-33)

### 3.2 *Fatalismo absurdo*

Agora podemos explicitar melhor a analogia entre Arjuna e Sam Rogers: ambos, preocupados com suas participações como protagonistas do desastre, questionam a aparente falta de sentido de suas ações e reprovam a participação num evento desastroso. Ao passo que tanto Tuld quanto Krishna, detentores de um conhecimento privilegiado<sup>105</sup>, respondem com fatalismo e com uma instigação à “glória” – ou, no caso de Tuld, ao lucro. Para o executivo, o prejuízo dos investidores já é uma realidade - assim como o é, para Krishna, a morte dos guerreiros<sup>106</sup>.

Metafísicos costumam objetar ao fatalismo alegando que é inaceitável que o futuro esteja “determinado”. Isso é irrelevante para nossa discussão; o que nos concerne é a tese moral de que o fatalismo esvazia a noção de que as ações do agente têm alguma relação especial com o resultado (“*you are a mere instrument*”, segundo Krishna), caso ele seja inevitável. O valor da ação torna-se totalmente dependente, na realidade, da forma como ela está inserida numa “visão cósmica”, em alguma concepção de como o mundo “realmente é” – o que quer que isso signifique. E sem esta, a conclusão é o *nihilismo*; o agente se enxerga numa situação afim à do Rei Macbeth diante da notícia da morte de sua esposa: “ela teria morrido em outro momento. [...] A vida [...] é um conto narrado por um idiota, cheio de som e fúria, que não significa nada” (Shakespeare, Ato V, Cena 5, versos 15-25, tradução livre).

---

<sup>104</sup> Trata-se de uma passagem particularmente conhecida por ter sido citada por J. Robert Oppenheimer (que, versado em sânscrito, afirmara ter sido altamente influenciado pelo texto) numa entrevista em que falava sobre a sensação de testemunhar o primeiro teste nuclear. Disponível em: <[www.youtube.com/watch?v=lb13ynu3Iac](http://www.youtube.com/watch?v=lb13ynu3Iac)>, acesso em 18 de agosto de 2016.

<sup>105</sup> Em outra passagem, Tuld menciona que é justamente por ‘adivinhar’ antes dos demais o que está para ocorrer que ele é o CEO da empresa:

**John Tuld:** *I'm here for one reason and one reason alone. I'm here to guess what the music might do a week, a month, a year from now. That's it. Nothing more. And standing here tonight, I'm afraid that I don't hear - a - thing. Just... silence.*

<sup>106</sup> Claro, a diferença fundamental é que, enquanto Krishna alega que somos todos instrumentos de uma vontade divina, vivendo num ciclo ilusório de nascimento e morte (a *samsara*), Tuld nos apresenta como entregues ao sabor do ciclo de negócios, das crises financeiras (algo análogo à *samsara*), num mundo onde os perdedores são deixados à sarjeta. Embora a “visão cósmica” (e o que Tuld talvez não perceba é que seu pessimismo, se não é uma “visão cósmica”, cumpre a mesma função) da realidade seja diferente em cada um dos casos, a forma dos dois argumentos é idêntica, assim como a consequência – a anulação da noção de responsabilidade pessoal.



Essa relação especial entre o agente e o resultado é capturada pelas noções de responsabilidade causal e “judgamental” (*judgemental*). É isso que nos permite compreender, e.g., a tragédia de Otelo; o fato de que ele é o responsável (causador) pelo homicídio de Desdêmona lhe dá uma perspectiva distinta da de qualquer outro agente (Sen, 2009, 220). Ele não lamenta apenas a morte da esposa, ou ter realizado uma ação censurável, mas principalmente por ser o autor do fato. A descoberta da culpa de Iago é tão incapaz de afastar essa responsabilidade quanto o seria a descoberta de que Desdêmona teria uma vida curta de qualquer maneira (por uma doença incurável, ou porque seria assassinada por Iago de outra forma); Otelo não pode dizer, como Macbeth, “ela teria morrido em outro momento”, pois, embora ele haja sido um instrumento de Iago, o qual previra exatamente como o mouro reagiria, as decisões e ações cruciais ainda foram de Otelo – e podemos reprová-las na medida em que expressam falhas de seu caráter (i.e., de traços de sua personalidade que são passíveis de avaliação moral, como virtudes e vícios – Michelon Jr., 2006, 49).

### *3.3 Fatalismo e cegueira*

Um “fatalista” pode tentar escapar a essa noção de responsabilidade decisional; imagine que Tuld alegasse que a própria decisão era inevitável: ela era a única possibilidade consistente, segundo a teoria dos jogos (ou outra teoria da decisão racional), já que maximizava a utilidade esperada – a sua, a dos acionistas, a dos demais executivos e empregados, até a da própria firma, na medida em que melhor satisfazia as preferências desses agentes.

Na melhor das hipóteses, uma alegação do tipo seria circular; afinal, em teoria dos jogos, “preferência” é um termo técnico: segundo a tese da preferência revelada, o comportamento de escolha é primitivo – i.e., a escolha revela a ordenação de preferências do agente (trata-se, pois, de uma relação lógica, e não causal – Blackburn, 2010, 12); apenas pressupõe-se que essa ordenação seja consistente e, pressupondo a comparabilidade dessas preferências, agir de forma consistente pode ser descrito como maximizar uma propriedade formal chamada utilidade (Binmore, 2007, 7). Dizer “agirei de tal maneira porque isso maximiza minha utilidade” é, portanto, circular – equivale a dizer “agirei de tal maneira porque é o que decido fazer”.

Isso sugere que uma teoria da decisão não deve pretender ser, simultaneamente, normativa e preditiva (ou descritiva); em especial, ela não pode modelar uma deliberação – ao

menos não uma que contenha alguma forma de dilema, em que a pergunta “o que devo fazer?” precisa ser respondida depois de “o que **quero**, ou o que **devo** querer fazer?”. Ou, como argumenta Blackburn:

[...] in game theory as it is now being conceived, nothing can be translated into advice. For suppose we are "advised" to follow the dominant strategy. This is null advice, equivalent to: behave so that a tautology is true of you. So if we don't follow the advice, then our choice reveals that it wasn't that game. But if it wasn't that game then the advice was inapplicable, and if the advice was inapplicable, then there was no point in following it in any event, for the game theorist had failed to model the situation properly. As Wittgenstein might have said, anything could accord with the advice, and that means that no advice was given. The economists' slogan "Maximize!" turns out to be no injunction at all, for nothing could count as failing to follow it. (Blackburn, 2010, 20).

Logo, em nosso exemplo, o que John Tuld poderia afirmar seria um “imperativo hipotético” do tipo “se desejo sobreviver à crise, devo fazer tal coisa”, mas não poderia apresentá-lo com uma “roupagem” categórica, como em “devo fazer tal coisa”. Antes de decidir, Tuld não pode considerar que a premissa do condicional já é verdadeira, já que suas preferências (no sentido técnico) estão em questão tanto quanto suas ações – e avaliar e formular essas preferências é a finalidade da deliberação moral. Ele lembraria um exemplo de Cohen (2008, 67) de um sequestrador que dissesse “*Gee, I’m sorry, but the fact is that, unless you pay, I will not release your child*” – o que indica um estranhamento de suas próprias intenções.

Seria necessário, portanto, mergulhar mais fundo no fatalismo, alegando algo como “sou assim, não poderia ter agido de outro modo”. Ocorre que uma proposição desse tipo não tem função numa deliberação moral (Dennett, 1984); um agente não pode acreditar, ao deliberar, que suas preferências (no sentido técnico) já estão definidas de modo independente de suas escolhas, nem que é impossível escolher entre agir ou não de determinada forma. Nesse caso, não parece haver sequer a ilusão de uma deliberação: um sujeito que tentasse agir dessa forma não estaria propriamente **deliberando**, mas tentando apenas **preversuas** ações<sup>107</sup> – ou

---

<sup>107</sup> Podem-se imaginar exceções, no estilo do episódio “Ulisses e as sereias” da *Íliada*. Por exemplo, um prisioneiro antecipa que, torturado, confessará, ainda que não queira fazê-lo. Mas, nesse caso, ou ele está: a) prevendo que a tortura vai torná-lo um “animal guinchante”, sobre o qual não é capaz de exercer controle, o que explica que visualize a confissão de forma “descompromissada”, como um “fato objetivo” (ou como a ação de um estranho), e não como uma ação propriamente sua, ou b) está considerando que a tortura quebrará sua força de vontade, colocando-o numa situação usual de *akrasia* – uma situação que, embora levante discussões filosóficas por si só, não costuma lançar quaisquer suspeitas sobre a ideia de responsabilidade. Ou seja, em (a), não há ação que se

talvez seja melhor dizer “o que ocorrerá com seu corpo”. Segundo Dworkin (2011, 223-224), não há como um agente(i.e., alguém que age) comportar-se como se, de fato, acreditasse nisso.

Essa posição absurda **não** é consequência de uma “metafísica determinista”, mas da incompatibilidade entre conceber a si mesmo, simultaneamente, como um objeto cujo movimento pode ser previsto de uma perspectiva inteiramente paramétrica ou “externa” e conceber essa mesma previsão como uma razão para agir da maneira prevista – o que revela, novamente, um raciocínio circular. Esse é um problema antigo – presente, e.g., na interpretação de tragédias gregas cujos personagens frequentemente agem sob a égide de oráculos e profecias:

Could Eteocles say, "It has been fixed by the gods that I shall do it, so my decision is that I shall do it?" As a decision, surely, this is incoherent. Agamemnon indeed put on the harness of necessity—he could do it, and there was a reason for him to do it. But he could not have put it on for the following reason: that he had seen that he was already wearing it. It is something like this that Eteocles would be doing if we read his last line as at once recognising an externally imposed necessity and offering it as a reason for his decision. (Williams, 1993, 137-138 – grifo nosso).

Segundo Williams, nas tragédias, não é a previsão contida na profecia que é uma razão para agir, mas uma obrigação sobrenatural do personagem de atender ao destino traçado pelos deuses<sup>108</sup> – como no exemplo supra de Arjuna. Isso, porém, é confrontado com as demais razões que o agente tem – as quais, porém, são mais fracas que essa razão sobrenatural; é isso que nos permite enxergar as ações previstas como verdadeiras ações.

De forma análoga, ao invés de conceber a argumentação de Tuld como excluindo a possibilidade de ações alternativas, é mais simples descrevê-la como algo semelhante ao que Palazzo, Krings e Hoffrage (2012) chamam de “cegueira ética”: o agente faz uma racionalização de sua escolha descrevendo a situação de uma forma que obscurece sua avaliação; os exemplos usuais são de homens de negócios que exageram a gravidade de suas

---

possa atribuir ao sujeito (poderíamos dizer que a identidade pessoal ao longo do tempo foi rompida); em (b), há ação, mas trata-se de um caso ordinário de responsabilidade.

<sup>108</sup> Isso é ilustrado com a hesitação final de Orestes, na tragédia *Coéforas*, antes de matar Clitemnestra como uma forma de cumprir uma **obrigação** para com o Oráculo:

“Orestes hesitates at the last instant before killing his mother and turns to Pylades—“Shall I do it?”; and he, in his one intervention, says, ‘Where will the trusty oracles of Loxias be for the rest of time. . . ?’ thus linking obedience to the oracle’s commands with saving its predictive credibility. There is complexity here, and an obscurity that comes with the very idea of an oracle, but in the *Choephoroi* Aeschylus revolves the elements of prophecy, supernatural plan, human motive, and divine command with an assurance that prevents them from falling together into immediate fatalism” (Williams, 1993, 140).

ações comparando o mercado a uma guerra, ou que, no outro extremo, diminuem essa gravidade comparando tudo a um jogo (como Tuld quando diz a Rogers que é “apenas dinheiro”, “*it’s made-up*”: “são apenas notas coloridas...”).

Algo semelhante ocorre quando uma pessoa apresenta suas escolhas como fatos inevitáveis. Segundo Jackall (1988, 39), executivos racionalizam suas decisões dessa forma: “*decisions are only made when they are inevitable*” – i.e., quando a decisão “faz a si mesma”. Uma afirmação desse tipo, numa situação cotidiana, pode ser aceita como uma figura de linguagem, como uma forma mais dramática de dizer “era a melhor coisa a fazer”<sup>109</sup> – mas, nesse caso, ela pressupõe uma escolha e a possibilidade de deliberação (e, com isso, a responsabilidade pela ação). Eis o principal contraste entre Tuld e Rogers: este se questiona se é melhor assumir o risco de falir, ou destruir o mercado.

#### 4 Conclusão

Williams (1993, 164) observa que:

Napoleon remarked to Goethe that what fate was in the ancient world, politics was in the modern, and in the same spirit Benjamin Constant said that the significance of the supernatural in ancient tragedy could be transferred to the modern theatre only in political terms.

De fato, Napoleão e Constant, tendo atravessado um dos períodos mais turbulentos da idade contemporânea, têm motivos para tal afirmação: para eles, a política, e não apenas a diplomacia (parafraseando o aforisma de Clausewitz), era a continuação da guerra por outros meios. Em nossos tempos, poder-se-ia atribuir ao mercado esse papel de “pano de fundo trágico”, com uma vantagem: enquanto comunidades e instituições políticas podem ser vistas como agentes racionais (ao menos em teoria, ou de forma idealizada), donde se lhes pode atribuir a capacidade de deliberação racional, refletida e pública, o mercado é *visto como* um agregado de decisões individuais, imprevisível e desprovido de intencionalidade.

---

<sup>109</sup> Como observa Roger Ebert em uma resenha do filme, a própria vida de suas personagens torna-se desprovida de significado: “*I think the movie is about how its characters are concerned only by the welfare of their corporations. There is no larger sense of the public good. Corporations are amoral, and exist to survive and succeed, at whatever human cost. [O elenco] reflect the enormity of what is happening: their company and their lives are being rendered meaningless*”. Disponível em <<http://www.rogerebert.com/reviews/margin-call-2011>>, acesso em 19.10.2011. De fato, o espectador atento pode notar como por vezes, nos diálogos, os personagens mencionam que “não tinham escolha” – embora estejam se referindo a casos paradigmáticos de decisão racional.

Defendemos supra que esse cenário “trágico” não afasta a responsabilidade moral de seus agentes – mesmo porque sequer a causalidade sobrenatural excluía dos heróis das tragédias antigas a responsabilidade por suas decisões. Guerreiros não podem isentar-se da responsabilidade por seus atos de guerra alegando a inevitabilidade desta – eles podem (o que é muito diverso) tentar alegar que agiram sob circunstâncias especiais justificatórias (o que corresponde, no direito, ao estado de necessidade – circunstâncias que tornam uma ação correta) ou exculpantes (circunstâncias que precluem a atribuição de culpa, em consideração às limitações humanas do agente – como, alegadamente, seria o caso da “névoa da guerra”). Essas são distinções usuais para juristas e filósofos morais.

Algo semelhante vale para investidores: mesmo que John Tuld estivesse correto ao dizer que não podemos prever, nem evitar, as crises financeiras, que só podemos “reagir”, isso não implicaria que a única forma de reagir é *sendo o primeiro*. Ele e Rogers tinham razões para agir de outra forma, e poderiam tê-lo feito; poderiam ter escolhido até mesmo falir a precipitar a ‘destruição’ do mercado.

Uma última comparação entre o filme e o épico hindu: não é o argumento fatalista de Krishna que põe fim à discussão do *Baghavat-Gita*, mas a aceitação de sua autoridade sobrenatural por Arjuna (que assim aceita seu próprio papel naquela “visão cósmica”). De forma distinta, Rogers se despede de Tuld reiterando que não concorda com ele, que não foi seu monólogo que o convenceu a ficar; ao invés disso, ele se conforma e permanece no emprego apenas porque precisa do **dinheiro**.

Contrariamente à citação inicial de Posner, a reprovação a financistas gananciosos é justificada, na medida em que é uma consequência do reconhecimento de que tomaram decisões criticáveis; isso não significa que essa reprovação seja útil: Posner, um pragmático confesso, quer dizer, na realidade, que o risco de ser alvo de reproche não constitui incentivo suficiente para evitar a ganância. Contudo, saber que essa reprovação é adequada nos permite defender uma cultura de ética corporativa, além de instituições que visem a dissuadir, por meio da coerção ou da responsabilização civil, condutas gananciosas socialmente deletérias.

## Capítulo 6: As *Eumenides* e a Crise: responsabilidade, risco moral e dissuasão no sistema financeiro

I'm not sure if it's philosophically possible for an action to be both necessary and a disaster, but that in essence is what the bailouts were.  
[...] life did not get harder for banks and for the financial system. In the popular imagination, the people who caused the crisis got away with it scot-free, and, as what scientists call a first-order approximation, that's about right. (Lanchester, 2018)

**Resumo:** Após a crise de 2008, houve diversas críticas ao *bail-out* e à falta de responsabilização de financistas; uma objeção a essas críticas é que estariam associadas a uma necessidade psicológica de encontrar culpados por sinistros. Deveríamos, pois, evitar punir instituições e gestores (mormente se solventes), pois: a) não caberia ao regulador tutelar os interesses dos investidores (objeção do *paternalismo*), b) haveria instrumentos mais eficientes e menos gravosos, associados à regulação prudencial (objeção da *ineficiência*).

Contra (a), a insolvência de instituições financeiras acarreta uma externalidade negativa: crises financeiras prejudicam de tal forma a economia que governos, fundos garantidores e órgãos de regulação são obrigados a intervir para impedi-las ou mitigá-las – o que corresponde a uma *garantia implícita* aos credores. Isso é polêmico do ponto de vista da moralidade política: implica utilizar recursos do público (ou de usuários do sistema financeiro) em benefício de pessoas em melhor situação que *escolheram* incorrer em certos riscos. Na primeira seção, estudamos uma possível justificação desse arranjo a partir do “argumento do incentivo”, decorrente do princípio da diferença de Rawls; concluímos, porém, que tal argumento não exclui considerações sobre a idéia de responsabilidade.

Contra (b), arguimos que não podemos prescindir completamente de instrumentos coercitivos – mormente porque a *garantia* acarreta risco moral *por tornar* improvável a liquidação de instituições. Logo, é preciso que a ameaça de responsabilização *tenha um caráter punitivo e preventivo*, se estendendo ao descumprimento de normas sobre risco mesmo *antes* de gerar efeitos econômicos catastróficos.

## Introdução

No curta-metragem cômico *The Last Laugh* (2008), um banqueiro (John Bird) diz a um entrevistador (John Fortune) que o colapso financeiro poderia ser evitado se *os governos devolvessem aos bancos o dinheiro que estes perderam com suas próprias especulações*; ao que o entrevistador pergunta, intrigado, o que aconteceria se não o fizessem. Ao que se segue a réplica: “... Eu diria a você o que pessoas como eu sempre dizem – que não seremos nós a sofrer com isso, mas o seu fundo de pensão”. A anedota *mostra*, mais do que explica, o que há de errado com certo raciocínio: por um lado, o que o banqueiro diz (e suas implicações) parece inaceitável de um ponto de vista moral – uma redistribuição de renda em favor duma elite econômica; por outro, é empiricamente verdadeiro que, sem a atuação do governo (ou de alguma instituição sob sua égide) como garante ou prestamista de última instância, o sistema financeiro pode colapsar, com consequências graves para toda a economia (principalmente para os mais pobres).

A *sketch* supraalude à crise de 2008, quando o governo americano interveio diretamente, a fim de auxiliar (ou de financiar a aquisição de) instituições com problemas de liquidez; o *bail-out* foi controverso “on both sides of the ideological aisle: on the right, because it meant a *de facto* admission of market failure, and on the left, because the bailout privileged the needs of Wall Street over Main Street” (Murray, Manrai & Manrai, 2017, 186). Mesmo um pragmático como Posner (2010, 111 e s.) reconhece que a indignação popular a respeito do *bail-out* foi parcialmente justificada (ao menos em relação a instituições que distribuíram lucros após o socorro financeiro, como o Goldman Sachs). Por trás dessa indignação está a idéia do esvaziamento da noção de responsabilidade, de que não se pode auferir benefício da própria torpeza: “no longo prazo, a sociedade não pode funcionar bem se as pessoas não assumem responsabilidade pelas consequências de suas ações”, i.e., se temos “individualismo sem responsabilidade individual” (Stiglitz, 2010b, 120 – tradução livre). Ressoa a ideia de que uma economia baseada no gerenciamento de risco (i.e., na especulação) é “corrosiva para as normas morais e civis com as quais vale a pena se importar” (Sandel, 2013, 335).

Preocupações afins têm ocupado a pauta sobre a reforma de legislações bancárias: sob a influência da crise, os EUA aprovaram a lei Dodd-Frank em 2010, adotando uma regulação financeira mais restritiva – com destaque para inovações na proteção ao consumidor, na regulação do mercado de derivativos e no regime de resolução bancária (Hanson, Kashyap & Stein, 2011, 24); em 2018, porém, a legislação foi substancialmente alterada, sob o argumento

de que se tornara um ônus para atividade econômica (Borak, 2017). No plano internacional, além das mudanças sobre requerimentos de capital do Acordo de Basileia III, nota-se uma tendência a enfatizar normas sobre governança e responsabilidade social, visando a alinhar os incentivos de administradores e acionistas<sup>110</sup> com os interesses de longo prazo das instituições financeiras (ou da sociedade), e reformas na legislação sobre resolução bancária visam à utilização de recursos dos próprios acionistas e credores da instituição, em caso de problemas de liquidez, a fim de prescindir do uso de fundos garantidores<sup>111</sup> ou de recursos públicos (Parker, 2011).

A principal abordagem para o problema é a *prudencial*<sup>112</sup>; ela contrasta com outra, com viés *coercitivo*, que enfatiza a necessidade *deresponsabilizar* (por meio de sanções –

---

<sup>110</sup> Em companhias abertas com capital difuso, os interesses de curto prazo de um acionista (dividendos e alto valor de mercado da ação) podem conflitar com o de longo prazo da pessoa jurídica (estabilidade) – além de conflitar com os de outros acionistas (um rótulo que pode se aplicar tanto a um pequeno investidor conservador quanto a um especulador fundo de participações) com portfólios diferentes, com os de diretores (que sofrem pressão por *performance*) ou com os do Conselho de Administração (que intermedeiam a relação entre sócios e diretores). A ideia de que a governança corporativa deveria privilegiar uma “democracia de acionistas” é facilmente desafiada por Fox e Lorsch (2012):

“[...] shareholders aren’t quite the same as owners. A simple illustration: if you own a car, you’re liable for damages in an accident even if they exceed the value of the car. But shareholders are on the hook only for what they’ve invested. And although some shareholders behave much like owners, most of them are effectively renters—often ultra-short-term renters. In real estate, renters are entitled to legal protection but seldom given a formal say in how a property is managed or whether it can be bought or sold. That seems appropriate for short-term shareholders as well”.

Mais uma vez, podem-se referir exemplos da Grande Crise: a remuneração variável de executivos era atrelada ao valor de mercado das instituições (mediante bônus de desempenho, ou simplesmente por ações ou opções de compra), de forma que eles perderam dinheiro quando estas perderam valor (Lo, 2012, 1); contudo, isso só implica um alinhamento de interesses no curto prazo – ainda seria vantajoso arriscar a solvência de uma instituição a fim de auferir ganhos suficientemente grandes, mormente se a remuneração não é afetada negativamente por perdas (Stiglitz, 2010b, 119-20).

<sup>111</sup> Fundos garantidores desempenham um papel ambíguo, a exemplo o Fundo Garantidor de Créditos (FGC) do Brasil: por um lado, trata-se de uma associação privada que assegura o pagamento de depósitos de valor limitado e financia operações que visem a evitar a insolvência de instituições (como a venda de instituições com dificuldades de liquidez); dessa forma, evitar-se-ia a crítica sobre a utilização do Banco Central e de recursos públicos para salvar bancos privados – o que é vedado pelo art. 28 da Lei de Responsabilidade Fiscal (Ferreira, 2012). Por outro lado, o FGC desempenha um serviço de utilidade pública, é regido por normas do Conselho Monetário Nacional e do Banco Central, e é financiado por contribuições obrigatórias, calculadas sobre as operações dos clientes (como a emissão de cheques, ou o volume de depósitos) dos bancos.

Há vantagens nessa estrutura privada, mas a solução não é propriamente inovadora; na verdade, ela evoca o passado: Bancos Centrais em geral foram instituições privadas até a primeira metade do Séc. XX (Piketty, 2014, 542 e 613). Uma das principais defesas de fundos garantidores privados é que é financiado por recursos do próprio sistema financeiro (a contribuição incide sobre o montante das contas de depósitos, conforme dispõe a Circular nº 3.666/2013), ao invés de recursos públicos; esse argumento é atraente, mas negligencia que:

- a) o conjunto de contribuintes tende a ser o mesmo nas duas possibilidades, pois, em economias modernas, os indivíduos que pagam impostos têm depósitos em bancos;
- b) impostos são (ou ao menos se espera que sejam) *progressivos*: contribuintes com maior renda ou riqueza arcam com uma alíquota maior; já a contribuição de uma instituição financeira a um fundo é embutida nos custos administrativos do negócio e tende a ser repartida entre seus clientes de forma *regressiva* – uma vez que clientes com transações de maior volume têm poder de barganha para negociar condições mais favoráveis (como remuneração maior por depósitos e isenção de taxas).

<sup>112</sup> Tanto a *micro* quanto a *macro* prudencial: “A microprudential approach is one in which regulation is partial equilibrium in its conception and aimed at preventing the costly failure of individual financial institutions. By



administrativas, civis, ou mesmo criminais) instituições e gestores que violem normas. No extremo desta segunda abordagem, a “melhor regra de gerenciamento de risco” já escrita seria a prescrição, pelo Código de Hamurabi, da condenação do construtor à morte, no caso de uma fatalidade produzida pela queda de uma casa (Taleb & Martin, 2012, 50).

A retórica da responsabilização, porém, é considerada suspeita em outros contextos, como no da seguridade social: a responsabilidade pela própria pobreza ou incapacidade não deveria excluir alguém do escopo de nossa assistência (Mounk, 2015, iv). Da mesma forma, no contexto da punição criminal: a justificação meramente *retributivista* violaria a neutralidade liberal, pois implicaria impor dano a outrem *apenas* para expressar reprovação (Tadros, 2011, 79-83). Trata-se de um eco das *Erínias*, as “Fúrias Vingadoras”: deusas gregas que perseguiram impiedosamente os homicidas.

Portanto, a justificativa da responsabilização punitiva (e da estrutura regulatória e de supervisão necessárias para aplicá-la) deve apelar para efeitos socialmente benéficos, como o efeito dissuasório da sanção<sup>113</sup>, a fim de ser justificável a todos. Mas tal justificação tem de enfrentar duas objeções: a) caberia aos próprios investidores exercer a “disciplina de mercado”, e não caberia às autoridades tutelá-los – o que seria uma forma de *paternalismo*; b) haveria outros instrumentos mais eficientes (e menos drásticos) que a imposição de sanções – a abordagem *prudencial* supra; i.e., sanções são *ineficientes*.

Por esses argumentos, a atuação dos reguladores e supervisores financeiros deveria limitar-se a prescrever o mínimo de normas prudenciais e contábeis (como o estabelecimento de limites mínimos de capital para a operação das instituições) necessárias para o funcionamento estável e coordenado do mercado e a evitar que sejam fraudadas (fiscalizar seu cumprimento). A realidade contrasta com essa conclusão: o mercado financeiro é um dos setores mais regulados da economia, e supervisores têm poder para impor drásticas medidas, preventivas (e.g., recusar a contratação de administradores ou suspender as atividades de uma

---

contrast, a “macroprudential” approach recognizes the importance of general equilibrium effects, and seeks to safeguard the financial system as a whole. In the aftermath of the crisis, there seems to be agreement among both academics and policymakers that financial regulation needs to move in a macroprudential direction.” (Hanson, Kashyap & Stein, 2011, 3).

O paradigma da abordagem coercitiva seria a regulação de conduta, voltada para a avaliação da observância, por parte da instituição e de seus gestores, da regulação atinente. Claro, embora contrastemos a abordagem prudencial com a coercitiva, as duas não são concorrentes; nem enfatizamos a segunda por considerá-la, em qualquer aspecto, mais *importante*.

<sup>113</sup> Isso *não* implica uma justificação consequencialista. Na teoria de Tadros, a punição é justificada (i.e., o agente tem o dever de submeter-se a punição, e o Estado tem o direito de punir) porque, se o dano resultante da violação não pode ser plenamente compensado pelo violador, então este tem o dever de proteger a vítima contra futuras violações; num sistema de justiça, a maneira institucional de o fazer é submeter-se à punição a fim de dissuadir futuras violações (Tadros, 2011, 277-80). Nosso argumento seguirá uma linha análoga: como não é possível uma compensação adequada pelos riscos sociais criados pelo sistema financeiro, a sanção é necessária para mitigá-los.

instituição) ou punitivas (e.g., proibir o exercício de cargos no setor ou cassar a autorização para funcionamento de uma instituição).

Na seção a seguir, concedendo um pouco de razão ao banqueiro do *The Last Laugh*, arguiremos em favor do compartilhamento social dos riscos da atividade financeira – i.e., da atuação de autoridades públicas visando a evitar risco sistêmico; isso equivale a uma forma de *garantia* (ainda que implícita) do sistema financeiro. Procederemos a partir do “argumento do incentivo”, derivado do princípio da diferença de Rawls, e da análise das objeções lançadas por G. Cohen. Embora o foco da discussão de Cohen seja a distribuição de renda associada à tributação, assumimos que ela pode ser extrapolada para o compartilhamento de riscos e ônus financeiros. Isso afasta, pois, a objeção do paternalismo supra: uma vez que o sistema financeiro implica uma potencial externalidade, não é apenas o interesse dos investidores que deve ser considerado.

Na seção posterior, arguimos que os sistemas jurídicos devem estabelecer sanções efetivas aplicáveis a agentes do sistema financeiro a fim de mitigar o risco moral (*moral hazard*) criado por tal *garantia*. A ideia é que a regulação e supervisão de riscos por parte do poder público, ao contrário de uma estrutura de instituições privadas, tem não só a vantagem de estabelecer um sistema de seguros compulsórios, mas também mitigar os riscos sociais mediante regulação, inspeção e punição (Hansson & Skogh, 1987, 144).

### **1 Distribuindo riscos de acordo com o princípio da diferença**

De acordo com o *princípio da diferença*, as desigualdades de riqueza e renda devem ser justificadas tendo em vista o benefício dos que estão em pior situação (Rawls, 1971, 72). Uma das várias críticas que Gerald Cohen (2008) dirige a esse princípio é que, numa interpretação permissiva, ele implica o chamado “argumento do incentivo”, segundo o qual é justificado favorecer os que já estão em melhor situação (ricos e talentosos), mesmo sem qualquer consideração relativa a mérito, a fim de incentivá-los a produzir, consumir ou assumir riscos. De fato, Rawls menciona algo afim:

Supposedly, [...] the greater expectations allowed to entrepreneurs encourages them to do things which raise the prospects of laboring class. Their better prospects act as incentives so that the economic process is more efficient, innovation proceeds at a faster pace, and so on (Rawls, 1971, 68).

Do ponto de vista empírico, é discutível até que ponto incentivos aos que estão em melhor situação resultam numa vantagem considerável para os que estão em pior situação – o

que seria o cerne do argumento da *trickle-down economics* (Quiggin, 2010, 136-37). Pode-se contrapor que, historicamente, o crescimento econômico não tem apresentado uma correlação estável com a igualdade de renda ou riqueza, mormente se a taxa de retorno sobre o capital for maior que o crescimento (Piketty, 2014, 555). Mas o ponto de Cohen independe disso: mesmo se essa premissa empírica for correta, ela não pode ser associada aos princípios de justiça. Isso não implica que, por vezes, não seja útil ou necessário adotar políticas na linha nesse argumento – mas, nesses casos, lamentavelmente, acabamos por comprometer a justiça, em função de outro valor ou por alguma urgência.

A principal objeção é que esse argumento não passa num *teste interpessoal*: imaginemos que a pessoa em pior situação (doravante, o “pobre”) ouvisse da pessoa em melhor situação (que chamaremos de “rico”) na sociedade o seguinte argumento (Cohen, 2008, 59):

- i. Políticas públicas devem otimizar a situação dos que estão em pior situação.
- ii. Se os impostos aumentarem, os ricos tendem a investir e a produzir menos, de modo que a posição dos mais pobres tende a piorar<sup>114</sup>.
- iii. Logo, os impostos não devem ser aumentados.

O problema com a premissa (ii) é que, embora possa ser uma razão para os pobres aceitarem o corte de impostos (afinal, como a resposta do banqueiro interpretado por John Bird, ela é análoga a uma ameaça<sup>115</sup>), não pode ser uma razão para os ricos defenderem-no<sup>116</sup>. Embora os primeiros possam enxergá-lo como um mero fato (como um movimento dentro do jogo da escolha social), essa alternativa não está aberta para estes: são os ricos que *tornam* a premissa (ii) verdadeira; eles precisam de uma justificativa não para aceitar, mas para impor a desigualdade que o argumento defende (Cohen, 2008, 60).

A acusação de Cohen é de que haveria uma estranha forma de alienação no rico que explicasse, de forma descompromissada, por que os pobres deveriam votar a redução de impostos – como se não fossem responsáveis por suas próprias intenções. Podemos analisar melhor o problema ao colocá-lo na “primeira pessoa”: não posso adotar essa atitude

---

<sup>114</sup> O banqueiro de *The Last Laugh* pode ser interpretado como usando um argumento com a mesma estrutura – trocando, porém, a premissa (ii) pela afirmação de que, se não permitirmos que as instituições financeiras assumam riscos e não nos dispusermos a cobrir as respectivas perdas, haverá uma recessão.

<sup>115</sup> Note a semelhança do argumento com o raciocínio dos atenienses na Conferência de Melos – a alegação de que a submissão é o melhor para os mélicos, já que a alternativa é a destruição (Tucídides, 1874).

<sup>116</sup> Cohen tem uma resposta pronta para o argumento contrafactual de que “você faria a mesma coisa se estivesse em meu lugar”: “I have always thought that the right reply to a white South African who says to an anti-apartheid advocate ‘YOU would see things differently if you were in my position’, is: Quite! I’m sure it does blind one’s vision” (Cohen, 2008, 61).

descompromissada para com minha decisão porque não posso *prever* uma decisão sem tomá-la, da mesma forma como não posso antecipar a aquisição de uma crença antes de adquiri-la. Pensar “farei isso, mas não decidi que o farei” é afim a pensar “*p*, mas não acredito que *p*”, o que, supondo-se a identidade pessoal ao longo do tempo, é uma instanciação do paradoxo de Moore, “for that prediction would show that I already have the true belief that *p*. The truth cannot wait” (Sorensen, 2014). Afinal, a justificação para julgar que *p* no futuro justifica a crença em *p* desde sempre, seja *p* um juízo teórico (sobre fatos empíricos, e.g.), seja prático (sobre o que devo fazer) – a diferença, nesses dois casos, é o que conta como justificação para *p*.

O rico só pode, pois, acreditar na premissa (ii) se *decidir* torná-la verdadeira. Mas para isso, não basta apontar para, p. ex., a literatura econômica e aduzir os efeitos adversos dos impostos sobre o consumo e o investimento; nem apontar para o fato de que, como os demais seres humanos, os ricos são auto-interessados. Ele não pode *justificar* sua decisão meramente apontando para qualquer fato; a “Lei de Hume” o impede: nenhum conjunto de proposições sobre “como o mundo é” pode prover uma justificativa, por si só, para uma conclusão sobre o que deve ser o caso – a menos que haja um juízo de valor “escondido” no meio do raciocínio (Dworkin, 2011, 44). Isso é necessário não apenas para garantir a independência do domínio do valor em relação ao dos fatos, mas também o inverso; afinal, se admitirmos que um juízo de fato é suficiente para implicar um de valor, temos de admitir (por *modus tollens*), que a negação deste implica a daquele – i.e., que um juízo de valor implica um factual.

O problema é que alguma espécie de “princípio de incentivo” parece decorrer da posição original, a versão *rawlsiana* do estado de natureza da filosofia contratualista (Rawls, 1971, sec. 1:3), onde as partes não têm conhecimento de suas condições particulares, mas apenas de fatos gerais a respeito da natureza humana. Dado esse “véu da ignorância”, elas escolheriam por princípios de justiça aqueles tendentes a *otimizar* a posição de cada indivíduo numa sociedade bem-ordenada. Como ainda não sabem quais serão as posições ocupadas por cada um após o levantamento do véu, o argumento do incentivo tenderia a ser aceito por beneficiar *ex ante* tanto o rico (que ainda não sabe que é rico) como o pobre (que ainda não sabe que é pobre).

Há um contraste entre o contexto “apriorístico” da *posição original* (onde cada parte é algo análogo a uma variável cujo valor ainda é desconhecido), donde se derivam os princípios de justiça, e a posição conhecida e determinada (após levantar-se o “véu da ignorância”) onde Cohen situa sua objeção interpessoal – que assume um caráter *ad hominem*. Esse contraste pode ser observado também na divergência em relação ao escopo dos princípios de justiça: enquanto

Rawls (1971, 72) alega que o *princípio da diferença* tem por escopo a estrutura básica da sociedade, e não a legislação ordinária, Cohen (2001, 134 e s.) insiste que qualquer princípio de justiça que mereça o nome deve ser invocável para além dessa estrutura, mesmo *depois* de se levantar o véu da posição original. Sem isso, por abster-se de considerações de mérito ou de justiça substantiva, Rawls não teria nenhum princípio com base no qual pudesse criticar o corte de impostos (ou, em nosso exemplo, o socorro a investidores imprudentes).

Na posição original, é plausível que as partes adotem um ponto de vista destacado em relação ao que ocorrerá depois de levantado o véu da ignorância; elas podem julgar (um juízo de fato) que, dada a natureza humana, algumas pessoas (i.e., “existe um x tal que...”) exigirão recompensas maiores para agir de modo socialmente benéfico – do mesmo modo, um economista pode prever que uma determinada barganha (ou algum outro *jogo*) terá certo resultado – e que isso justifica (um juízo de princípio) que assim as recompensemos. Mas assim como **eu** (uma “variável identificada”) não posso justificar uma exigência afim em benefício próprio, não posso considerar justo que *A* (outra variável identificada) assim o exija.

#### *A relação entre incentivos e princípios*

Mesmo Cohen admite que, ao menos na política real, não se pode escapar de discussões sobre incentivos – eles são o principal instrumento de governo moderno (mesmo sanções, arguiremos posteriormente, funcionam como uma forma de *incentivo negativo*); logo, ainda que o argumento de Cohen fosse correto, ele não nos ensinaria uma restrição moral/prática geral que funcionasse melhor do que o *princípio da diferença*. Nessa linha, este tem a vantagem de ser um princípio útil, ou conveniente – um compromisso entre o ideal de justiça com a eficiência. A crítica de Cohen (2011, 229) é que atos justos não precisam ser eficientes dessa forma<sup>117</sup>: seria uma verdadeira prova de virtude se, e.g., eu decidisse queimar um “maná do

---

<sup>117</sup> Uma das vantagens do princípio da diferença é que, do ponto de vista da posição original, ele pode ser considerado Pareto-eficiente (Rawls, 1971, 59); aparentemente, isso não valeria para a igualdade radical de Cohen. Na realidade, uma situação em que as pessoas *voluntariamente* tendessem à igualdade seria Pareto-eficiente, porque os cidadãos de alta-produtividade manteriam o mesmo nível de produção, em benefício dos mais desfavorecidos (Cohen, 2008, 90; Atria & Michelon Jr., 2007) – satisfazendo a preferência de todos os cidadãos igualitaristas por uma distribuição igual (na realidade, isso viola *outra* premissa da teoria de Rawls – do desinteresse mútuo).

O problema é que **não** vivemos em tal mundo, não temos qualquer idéia de como chegar até ele e, mesmo que isso ocorresse, não saberíamos como mantê-lo de forma estável – ele cria um *risco moral* que indivíduos egoístas poderiam explorar em benefício próprio. Por outro lado, também não vivemos num mundo sem qualquer preferência intrínseca (moral ou não) por igualdade; Binmore (2005, 143-45) aduz que essa é uma preferência natural, que evoluiu em sociedades pré-históricas, a partir de práticas sociais como a divisão de comida. Isso condiz com a análise de Gintis (2006, 11-15) sobre a existência de equilíbrios cooperativos em simulações da *tragédia dos comuns*, indicando a existência de uma “reciprocidade forte” - i.e., um viés pró-cooperação, contrário ao auto-interesse, em uma ampla parcela da população. Essa parece ser uma característica esperada quando interpretamos

céuindivísivel” que chovesse sobre meu terreno, pelo simples fato de que é um benefício injustificado. Isso é verdadeiro mesmo se admitirmos que a igualdade é um ideal inalcançável, por razões empíricas (históricas, sociais ou biológicas), e que, dado este fato, “incentive-based inequality [...] is regrettably unavoidable, if not *tout court*, then at least if we do not want to depress everyone’s condition” (Cohen, 2001, 121). O importante, para ele, é separar, conceitualmente, considerações de eficiência e de fato (como uma explicação biológica ou sociológica do egoísmo humano) das considerações de princípio, a fim de não confundir as duas áreas.

A primeira resposta, portanto, que um defensor do argumento do incentivo poderia dar a essa objeção é que, se a disputa se limita à *análise do conceito* de justiça (e não sobre o que devemos fazer, na prática), ela perde um pouco do interesse para a filosofia política – e perde *todo* o interesse para a ciência social aplicada. No entanto, é preciso reconhecer que autoridades e cidadãos não podem ter um ponto de vista tão destacado quanto o da posição original, pois suas decisões e juízos influenciam as expectativas e demandas dos agentes; se, adotando a tese da *supply curve*, reconhecermos que é permitido cortar impostos para beneficiar a situação dos desfavorecidos, também *incentivaremos* os ricos a demandarem maiores cortes de impostos. O resultado é que *não beneficiaremos* os desfavorecidos; como observa Cohen em outra situação (ao tratar do abandono de princípios socialistas por parte do Partido Trabalhista inglês na década de 90), “It can be inexpedient to abandon principle for expediency, because it is hard to hide the fact that you are doing so” (Cohen, 2011, 217). Logo, mesmo que o princípio da diferença seja uma inferência correta da destacada posição original, isso não significa que ele é *suficiente* para justificar as decisões sociais sobre incentivos – ao menos não sem supor previamente as regras de alguma instituição, ou sem remeter a algum outro princípio de justiça, como mérito, responsabilidade, necessidade, solidariedade, etc.

A segunda alternativa para um *rawlsiano* é “morder a bala”: *não afetaria* a teoria da justiça admitirmos que o princípio da diferença (mesmo incorporando o argumento do incentivo) compromete, parcialmente, a justiça com a eficiência<sup>118</sup>, nem conceder que os cidadãos (além das instituições) também devem observar princípios de justiça. Essa alternativa se segue da ideia de que “When all strive to comply with these principles and each succeeds, then [...] their nature as moral persons is most fully realized” (Rawls, 1971, 463; cf. Cohen, 2001, 126). A teoria da justiça também não seria afetada se admitirmos que *adiscussão*

---

a dinâmica evolutiva como operando sobre conjuntos de estratégias (que, em simulações, instanciam um algoritmo – i.e., “memes”), e não sobre indivíduos biológicos.

<sup>118</sup> Na realidade, uma das vantagens do princípio da diferença é que ele seria Pareto-eficiente (Rawls, 1971, 59).

*políticas* sobre o sistema tributário, o sistema financeiro, a seguridade social, etc., pode apelar para outros princípios, como alguma noção compartilhada de responsabilidade<sup>119</sup>, de mérito ou necessidade, desde que se respeitem os limites da razão pública – i.e., que os argumentos dessa discussão procedem de “razões ou premissas que, razoavelmente, pensamos que outros poderiam aceitar para conclusões que eles, também razoavelmente, poderiam aceitar” (Freeman, 2004, 1-2, trad. livre). Isso tira força tanto da crítica a partir do argumento do incentivo, quanto da objeção de que não há sociedade justa sem alguma forma de *ethos* igualitário – o cerne da crítica de Cohen (2008, 116; 2011, 246-47) à ideia de que os princípios de justiça se restringem à estrutura básica.

Portanto, conclui-se que, ainda que o argumento do incentivo possa dar a “primeira palavra” – e quiçá a mais importante – em discussões sobre distribuições, isso não implica que ele seja a última: outros princípios, que se apresentem de acordo com a razão pública, devem ser considerados.

## 2 Garantia implícita e risco moral

Nosso problema inicial, ilustrado pelo *The Last Laugh*, trata da justificação do *bail-out* – e da polêmica redistribuição de recursos que ele implica. Vimos supra a possibilidade de ele ser defendido a partir do argumento do incentivo: uma vez que crises sistêmicas são ainda mais danosas para a população em pior situação, é justificado utilizar recursos para evitá-las, ainda que isso implique favorecer os que estão em melhor situação. Contudo, a objeção do teste interpessoal destaca o risco de que esse argumento possa sofrer abuso.

A questão remete ao *trade-off* entre o problema da *seleção adversa* e o do *risco moral* na teoria econômica sobre *design* de mecanismos (Myerson, 2008). Trata-se de duas formas de *restrição informacional*; no caso do risco moral (associado a atividades em que parte dos riscos é suportada por indivíduos que não tomam a respectiva decisão – e.g., empreendimentos garantidos por seguro, ou empresas estatais), gestores com pouco a perder não têm incentivos para minimizar custos nem aumentar a produção. Na seleção adversa (presente em contextos em que os riscos graves recaem sobre os tomadores de decisão – de modo que agentes aversos a risco se retiram da atividade), os gestores têm incentivos *demais* para manipular ou omitir informações a fim de minimizar suas perdas e maximizar seus ganhos pessoais<sup>120</sup>.

---

<sup>119</sup> Pelo contrário, é pela capacidade de assumir responsabilidade inclusive por seus fins (suas preferências) que os cidadãos são vistos como mais do que “passivos portadores de desejos” (Rawls, 2005, 186).

<sup>120</sup> Ou mesmo os ganhos da firma – o que pode ser incentivado pela cultura corporativa: “Some corporate cultures actively conceal their flaws and irregularities, not only from the public or from regulators but also from others

Em nosso exemplo, a ausência de uma garantia em relação a créditos de instituições insolventes implicaria seleção adversa: ou investidores prudentes reduziriam suas exposições, ou as instituições teriam de oferecer maiores prêmios de risco. Na verdade, mesmo essa situação é improvável, pois não é crível que, dada a ameaça de uma crise sistêmica, o governo se absteresse de intervir, caso uma instituição sistemicamente relevante estivesse em risco – o que implica a *garantia implícita* de instituições “grandes demais para falir”<sup>121</sup>. Mas isso leva ao *risco moral* exposto em *The Last Laugh*, que constitui uma potencial ameaça à estabilidade e à justiça de nossas instituições; poderíamos “internalizar” tal externalidade<sup>122</sup> sem alterações drásticas no sistema financeiro?

A abordagem *prudencial* da regulação assim recomenda:

Banks finance themselves with government-insured deposits. While deposit insurance has the valuable effect of preventing runs [...], it creates an incentive for bank managers to take excessive risks, knowing that losses will be covered by the taxpayer. The goal of capital regulation is to force banks to internalize losses, thereby protecting the deposit insurance fund and mitigating moral hazard. Thus, if the probability of the deposit insurer bearing losses is reduced to a low enough level, microprudential regulation is doing its job (Hanson, Kashyap & Stein, 2011, 4).

Ela é análoga, pois, ao princípio do *poluidor pagador*, que recomenda precificar o custo (individual ou social) causado pela poluição e estabelecer meios para que o poluidor a compense por meio de indenizações (como sugere Nozick, 1993, 79-80) ou mesmo por tarifas ou por um mercado de direitos de emissão negociáveis (Tirole, 2013, 38)<sup>123</sup>. Nessa linha, as instituições

---

within the corporation itself because of the risk that wider knowledge of these issues might undermine the firm’s position” (Lo, 2015, 30).

<sup>121</sup> A economia americana teria pagado caro com a recusa em socorrer o Lehman Brothers (Posner, 2010, 6), que foi o estopim de uma crise mundial nos mercados financeiros. Na realidade, o que Greenspan (2013, 16-17) chama de *exuberância irracional* pode ser associado às noções de risco moral e seleção adversa: a *euforia* guarda relação com o risco moral – em ambos os casos, investidores agem de maneira pouco sensível ao risco de perda; já o *pânico* corresponde à seleção adversa: investidores temerosos demandam prêmios de risco altos demais.

Cabe notar que o argumento a seguir independe de uma relação causal entre a existência de instituições “grandes demais” e risco moral; mesmo que este seja causado por outros fatores, a perspectiva de punição poderia constituir um desincentivo eficaz.

<sup>122</sup> Uma definição informal e prática de externalidades negativas: “a certain class of unintended, harmful side effects, things that happen because they produce benefits to some corporation and that are not stopped because the corporation in question is not where the bad effects are felt” (Passos & Goodwin, 2005, 3).

<sup>123</sup> No limite, a ambição de que os agentes devem ser adequadamente responsáveis pelos custos de oportunidade de suas decisões leva a uma teoria da justiça distributiva como igualdade de recursos:

“Equality of resources [...] offers the idea of a fair distribution of opportunity costs, not as derivative from other reasons for allowing deviation from flat equality but as itself a reason for deviating and limiting the scope of such deviation. It defines true opportunity costs recursively as those measured by prices in a market in which all have equal resources and in which insurance against risks of different sorts is marketed on equal terms. The yield of that market then structures, through taxation and redistribution, future markets in which prices set true opportunity



financeiras devem contribuir para fundos garantidores na medida do risco a que os expõem – algo afim a uma “taxa pigouviana”. Analogamente, exige-se maior capital social de acordo com uma função que pondera o risco dos ativos da respectiva instituição – i.e., o regulador torna o ativo mais “caro” para o acionista, de acordo com a probabilidade de *default*. Donde o risco social relacionado à instituição estaria (ou deveria estar) devidamente precificado.

Soluções para “internalizar externalidades” podem ser associadas ao *teorema de Coase*: na ausência de *custos de transação*, e na presença de um livre mercado com direitos de propriedade bem definidos em relação às externalidades, a barganha levaria a um resultado eficiente (Coase, 1991, 436). Logo, se pudéssemos estabelecer um sistema de direitos disponíveis (como propriedades) entre indivíduos racionais num mercado sem tais custos, então não precisaríamos nos preocupar com proibições. Mas economistas são os primeiros a reconhecer que as condições para aplicação do teorema raramente estão presentes no mundo real, como destacado pelo próprio Coase (1960, 43-44; 1991, 436). Restrições *informacionais* (como a seleção adversa, o risco moral e os limites cognitivos naturais dos indivíduos reais) equivalem a custos de transação que não podem ser completamente afastados (Myerson, 2008, 131); em tal cenário, um sistema de compensações pode ser inferior a um sistema que as mistura a proibições e regulação: “Transactions costs typically imply that the private system does not eliminate all externalities and risks. The resulting second-best outcome may be improved by an alternative system of public insurance, regulation and punishment of offenders” (Hansson & Skogh, 1987, 143).

Não se trata de uma diferença marginal entre a realidade e a teoria que usamos; nesse caso, os aspectos omitidos (custos de transação) são essenciais para a própria *função da teoria*<sup>124</sup>. Mercados são instituições para possibilitar trocas – para o “encontro” entre oferta e

---

costs. So the ambition to make people responsible for their choices is at work in that conception of distributive justice right from the start” (Dworkin, 2011, 479).

Poderíamos discutir amplamente sobre as divergências entre as concepções de igualdade de Rawls (a cujos princípios de justiça se chega por meio do exercício da razão instrumental num contexto idealizado), Dworkin (para quem a igualdade é satisfeita por uma distribuição igual de recursos, tendo em vista a necessidade fundamental de respeitar a responsabilidade dos indivíduos por suas escolhas) e Cohen (cuja concepção de igualdade, em última análise, incorpora o princípio marxista “de cada um segundo sua capacidade, a cada um segundo sua necessidade”). No entanto, isso fugiria demais ao escopo do presente texto; em especial, a ambição da concepção *dworkiniana* não pode ser adequadamente transplantada para a divisão dos ônus e riscos do sistema financeiro, ao menos num sistema econômico afim ao atual: instituições financeiras não podem compensar-nos plenamente (ao menos não sem transferir o respectivo ônus aos clientes) pelo custo de oportunidade gerado pelo risco sistêmico – o melhor que podemos fazer no lugar disso é mitigá-lo mediante uma estrutura regulatória adequada.

<sup>124</sup> Numa situação hipotética onde vingasse o *laissez-faire* extremo, qual seria a função da análise econômica do direito?

“This approach inevitably leads to a looseness of thought since the nature of the alternatives being compared is never clear. In a state of *laissez-faire*, is there a monetary, a legal or a political system and if so, what are they? In

demanda (Herzog, 2017); restrições delimitam os “contornos” dentro dos quais as transações podem ocorrer (inclusive a proscricção básica de “trapaça”)<sup>125</sup>. Donde o problema com modelos de análise econômica que pressupõem ausência de restrições informacionais: não fosse a grave incerteza e as assimetrias informacionais dos mercados reais, os agentes não teriam incentivo para buscar novas informações e possíveis oportunidades de arbitragem – assim as esgotando e agregando nova informação ao mercado (Grossman & Stiglitz, 1980); trata-se do “paradoxo dos mercados eficientes” (Lomasky, 2011, 150). Traçando uma analogia com a física, pode ser racional desconsiderar o atrito quando se calcula uma queda livre de curta distância; mas essa variável é fundamental para deslocamentos horizontais, como caminhada, patinação e automobilismo: um patinador que lamentasse a existência do atrito simplesmente ignoraria o princípio básico que torna possível sua atividade<sup>126</sup>.

Outrossim, economistas reconhecem que há regras sociais relativas a outros aspectos da vida em comum (e.g., a *repugnância moral* em relação a alguns tipos de transações<sup>127</sup>, como vendas de órgãos – Roth, 2007). Isso implica outros limites à ideia de compensação: tratar algumas relações sociais como transações pode levar a um resultado sub-ótimo *porque se trata*

---

an ideal world, would there be a monetary, a legal or a political system and if so, what would they be? The answers to all these questions are shrouded in mystery and every man is free to draw whatever conclusions he likes. Actually very little analysis is required to show that an ideal world is better than a state of laissez faire, unless the definitions of a state of laissez faire and an ideal world happen to be the same. But the whole discussion is largely irrelevant for questions of economic policy since whatever we may have in mind as our ideal world, it is clear that we have not yet discovered how to get to it from where we are” (Coase, 1960, 43).

De forma análoga, teorias utópicas têm pouca (ou nenhuma) relevância para a filosofia prática; e.g., como aduz Williams (2005, 163) se todos fossem perfeitamente altruístas, a investigação moral seria desnecessária – nesse sentido, a noção de circunstâncias da justiça (escassez moderada, altruísmo limitado) apontadas por Rawls (1971, §22): sem elas, a justiça seria impossível ou desnecessária.

<sup>125</sup> “Stock and produce exchanges are often used by economists as examples of perfect or near-perfect competition. But these exchanges regulate in great detail the activities of traders (and this quite apart from any public regulation there may be). What can be traded, when it can be traded, the terms of settlement and so on are all laid down by the authorities of the exchange. There is, in effect, a private law. Without such rules and regulations, the speedy conclusion of trades would not be possible” (Coase, 1991, 437).

No limite, a moderna ambição de *smart contracts* (baseados em *blockchains*) de que *code is law*: que todas as regras estejam incluídas *dentro* dos algoritmos que representam as transações – substituindo ao máximo a necessidade de *enforcement* externo (i.e., por parte de pessoas). Um exemplo a sugerir ceticismo sobre essa ambição é o da *The DAO (Decentralized Autonomous Organization)*, um tipo digital de fundo de investimento de capital de risco (*venture capital fund*) instanciado no *blockchain* do *Ethereum*; em junho de 2016, menos de um mês após seu lançamento, usuários exploraram uma vulnerabilidade no código para transferir U\$ 50 milhões para suas próprias contas – alegando que não haviam quebrado qualquer lei, já que a DAO não é sediada em nenhum país, nem se submete a qualquer regulação.

<sup>126</sup> “We have got onto slippery ice where there is no friction and so in a certain sense the conditions are ideal, but also, just because of that, we are unable to walk. We want to walk so we need friction. Back to the rough ground!” (Wittgenstein, 1953, §107).

<sup>127</sup> Isso não significa que a repugnância em si seja justificada, nem que seja uma justificativa para proibir a transação (muito pelo contrário, a ética pode demandar ação política para mitigar os efeitos dessa repugnância – e.g., se ela decorrer de preconceito); o ponto é que *mesmo sem considerar a justificacção moral da transacção*, a repugnância deve ser encarada como um custo de transacção, afetando o funcionamento desse mercado.

de atividades cujo valor é diminuído pelo ato de transacionar (Tirole, 2013, 39). Sandel, usando um exemplo de Gneezy & Rustichini (2000), assim argumenta para distinguir *tarifas* e *multas*:

A study of some child-care centers in Israel shows how this can happen. The centers faced a familiar problem: parents sometimes came late to pick up their children. A teacher had to stay with the children until the tardy parents arrived. To solve this problem, the centers imposed a fine for late pickups. What do you suppose happened? Late pickups actually increased.

[...] Introducing the monetary payment changed the norms. Before, parents who came late felt guilty; they were imposing an inconvenience on the teachers. Now parents considered a late pickup as a service for which they were willing to pay. They treated the fine as if it were a fee. (Sandel, 2012, 43)

No exemplo, o fato de a escola estar sendo adequadamente remunerada pelo atraso paterno relaxa a pressão social pela pontualidade; mas tal arranjo negligencia justamente os interesses das crianças – que deveriam ser a parte interessadíssima relevante em contextos do tipo (Tirole, 2013, 38). Não é que normas sociais, como a reprovação ao atraso, escapem à noção de escolha racional; mas elas demandam uma consideração distinta do usual “jeito econômico de ver o comportamento” (Becker, 1993)<sup>128</sup>. Elas abrangem noções deontológicas como direitos e restrições (*constraints*) e o caráter *preemptivo* de regras sociais e de ordens de autoridades – em contraste com o caráter opcional de ameaças e conselhos (Migotti, 2015, 389). Ao menos na maioria das sociedades que conhecemos, a conexão entre as noções de obrigação, reprovação e sanção são profundas, e expressadas em conjunto:

It is plausible that punishment is needed to trigger concepts of obligation just as exposure to language is needed to trigger understanding of syntax. To be punishment rather than mere training, the punisher must believe that an obligation has been violated, in this case, the obligation not to hit your sister just for fun. If some obligations can only be known after we are punished for violating them, they must have been in force and yet unknowable at the time we violated them. (Sorensen, 1995, 264-5)

Isso nos leva ao problema da ausência de sanções para financistas, talvez “the most important legacy of the Great Recession. Absent individual prosecutions for criminal behavior,

---

<sup>128</sup> Novamente, isso não é novidade para economistas:

“[...]it is, of course, desirable that the choice between different social arrangements for the solution of economic problems should be carried out in broader terms than this and that the total effect of these arrangements in all spheres of life should be taken into account. As Frank H. Knight has so often emphasized, problems of welfare economics must ultimately dissolve into a study of aesthetics and moral” (Coase, 1960, 44).

agents face a moral hazard, with the cost borne by principals (shareholders) and society”<sup>129</sup> (Murray, Manrai & Manrai, 2017, 186). Tal fato contribuiu para a alienação e a falta de confiança do público em relação às instituições econômicas e políticas – cujas consequências, pode-se arguir, ainda estão presentes (Lanchester, 2017). Não deveríamos subestimar o apelo psicológico da ausência de sanção; tal importância pode ser exemplificada pela recorrência do tema em tragédias gregas que se desenrolam em torno de tramas de vingança e expiação, como na *Oresteia*, em *Édipo Rei* e na *Antígona*.

Nosso argumento, porém, não apela para uma posição *moralista*, nem *retributivista*: não é somente porque o público demanda a responsabilização, com base em intuições morais, que o sistema jurídico deve implementá-la. Na realidade, pode-se sugerir o inverso: tais intuições morais decorrem de uma forma de *altruísmo punitivo* (a ideia de que se devem punir desvios da norma social, mesmo que isso acarrete custos), que evoluiu *junto* com nossas práticas de cooperação a fim de mitigar o risco moral e a existência de *free riders* (Gintis, 2006, 11-12); portanto, a permanência da reprovação moral à ausência de punição é um *signal* (mas não uma demonstração) de que ainda há um problema de risco moral a ser resolvido.

O que o estado de direito opera é uma transformação do altruísmo punitivo intuitivo num sistema de justiça com sanções condicionadas a um devido processo – algo análogo à transformação das *Erínias*, as “Fúrias”, no drama de Ésquilo (1996); na peça, as repulsivas divindades da vingança perseguem Orestes para castigá-lo pelo assassinato de sua mãe, Clitemnestra, até que a deusa Atena as convence a submetê-lo a um júri de cidadãos, os quais o inocentam. Inicialmente, elas se ressentem do resultado e ameaçam amaldiçoar a terra, mas acabam sendo persuadidas a proteger a cidade como guardiãs da justiça; o que as leva a uma transformação em aparência (deixam de ser abjetas) e em caráter (sua raiva é contida), donde passam a ser chamadas de “Bondosas” (*Eumenides*), destacada por Nussbaum:

---

<sup>129</sup> Ora defendemos a previsão de sanções individuais, ou de mecanismos afins com finalidade dissuasória; é exagerado, porém, concluir que as instituições financeiras não sofreram qualquer sanção ou impacto após a crise; além do dano de imagem, foram-lhes imputadas diversas multas nos anos seguintes: “In 2014, the US Justice Department collected a record \$24.7 billion in fines and penalties from banks and other financial institutions for fraud and misconduct in connection with the 2008 financial crisis, and this was on top of the \$21 billion it had already collected from 2012 through 2013 (Reuters 2014). During the same period, other federal and state law enforcement agencies also collected eye-popping amounts, and from many of the same institutions (Silver-Greenberg and Eavis 2013, 2014). The fines kept coming in 2015 (Protest 2015a and b; Protest and Ewing 2015), and by October the grand total collected by US federal and state authorities since the financial crisis began had risen to a whopping \$204 billion (Cox 2015). And this does not include the fines imposed by various European regulators, which have also been substantial (see Sterngold 2014). Even now, early in 2016, the grand total continues to grow by leaps and bounds, with huge new fines being assessed against or agreed to by Goldman Sachs, Wells Fargo, and Morgan Stanley (Goldstein 2016, Koren 2016, Popper 2016a)” (Reiff, 2017, 125).

Aeschylus shows that a democratic legal order can't just put a cage around retribution, it must fundamentally transform it from something hardly human, obsessive, bloodthirsty, to something human, accepting of reasons, something that protects life rather than threatening it. The Furies are still needed, because this is an imperfect world and there are always crimes to be dealt with. But they are not wanted or needed in their original form. They must become instruments of justice and human welfare. The city is liberated from the scourge of vindictive anger, which produces civil strife. In its place, the city gets forward-looking justice. (Nussbaum, 2017)

### *Supervisão e risco moral*

Punições são *prima facie* ineficientes, pois, ao invés de uma transferência de recursos, há apenas a imposição de custos ao próprio sujeito (a pena) e àqueles que se encarregam de aplicá-las. No entanto, é uma das estratégias sugeridas para mitigar o risco moral; Greenspan<sup>130</sup>, citando Hugh McCullough, primeiro Controlador da Moeda americano, aduz que não há risco moral caso a lei seja

emendada de modo que a falência de um banco nacional seja declarada *prima facie* fraudulenta, e que os diretores e executivos sob cuja administração tal insolvência ocorra sejam considerados pessoalmente responsáveis pelas dívidas do banco, e sejam criminalmente punidos, a menos que da investigação se conclua que seus negócios foram honestamente administrados (Greenspan, 2013, 110),.

As disposições correspondentes do direito brasileiro se aproximam dessa proposta: a Lei nº 6.024/1974 considera os gestores solidariamente responsáveis (desviando, pois, do princípio da responsabilidade limitada) pelo passivo da instituição<sup>131</sup>, e a Lei nº 7.492/1986 considera crime a gestão temerária de instituições financeiras; ainda, cada liquidação é acompanhada por uma comissão de inquérito encarregada de investigar as causas do evento –

---

<sup>130</sup> Na verdade, essa não é a medida favorecida por Greenspan; ele sugere que seria preferível que bancos deixassem de ser companhias abertas e voltassem a ser sociedades fechadas (*partnerships*) – pois isso aproximaria o acionista da gestão da sociedade e o faria responsabilizar-se por ela. No entanto, isso implica perder as vantagens de companhias abertas em aspectos como transparência e sensibilidade à disciplina de mercado; além disso, se o sócio só responde pelo capital em risco, permanece o risco moral associado ao problema da “pele em jogo”.

<sup>131</sup> Embora a redação do dispositivo seja dúbia (contendo, inclusive, um aparente erro tipográfico): “Art. 40. Os administradores de instituições financeiras respondem solidariamente pelas obrigações por elas assumidas durante sua gestão, até que se cumpram.

Parágrafo único. A responsabilidade solidária se circunscreverá ao montante e [sic] dos prejuízos causados”. Primeiro, o *caput* sugere uma aplicação anômala da noção de responsabilidade objetiva; por isso, a jurisprudência nacional considera que se trata de responsabilidade subjetiva, mas com presunção *júris tantum* (STJ, REsp 819217-RJ, Rel. Min. Massami Uyeda, 3ª Turma, DJe 06/11/2009). Segundo, as noções de prejuízo (associada a ativos que resultaram em perda) e de obrigações (passivos) são completamente distintas; e.g., suponha que, na gestão A, uma instituição aceite uma quantidade qualquer de depósitos, e que, na gestão seguinte B, use esses recursos para conceder créditos de forma negligente, o que resulta em insolvência: a gestão B é a causadora dos prejuízos, mas foi a gestão A que assumiu as respectivas obrigações.

algo afim à comissão de investigação independente sugerida por Lo (2009)<sup>132</sup>. Todavia, ainda que o sistema financeiro haja se mostrado resistente a *diversas* crises bancárias (Lundberg, 1999), isso não o tornou *immune* a risco moral – ou elas dificilmente teriam ocorrido.

O problema é que a proposta de McCullough se restringe ao caso da liquidação de uma instituição; conforme exposto supra, as autoridades têm incentivos e instrumentos para *prevenir* a liquidação de uma instituição em dificuldades – e. g., financiando sua aquisição por outra, como no caso do Bear Sterns, adquirido pelo J. P. Morgan. Mas isso agrava a concentração do mercado, como ocorreu na Crise: “The failing banks were eaten by surviving banks, with the result that the surviving banks are now bigger, and the too big to fail problem is worse” (Lanchester, 2017). Logo, mantém-se o risco moral, porquanto improvável a liquidação<sup>133</sup> de uma instituição que ofereça risco sistêmico – o caso do Lehman Brothers foi a “exceção que confirma a regra”: a corrida bancária que se seguiu à sua quebra levou as autoridades americanas a esforços inéditos para evitar outras falências e recuperar a economia (Posner, 2010, 6).

Ainda, restringir a punição a instituições insolventes implicaria incentivar a utilização de critérios contábeis e a realização de investimentos intencionalmente duvidosos (e, no limite, de fraudes), além da adoção de atitudes *pró-cíclicas* (como “liquidações” – *fire sales* – de ativos com preços em queda), agravando o risco de crise sistêmica. Afinal, ao reconhecerem a possibilidade de quebra da instituição, os gestores terão mais um incentivo para mantê-la operando a qualquer custo – mesmo que isso implique adotar critérios contábeis especiais e realizar transações apenas para *esconder* as perdas (a exemplo do uso do ardid do *Repo 105* por

---

<sup>132</sup> Uma importante diferença, contudo, é que as comissões de inquérito são formadas por servidores do Banco Central, nomeados por este – o que comprometeria sua independência em relação ao regulador.

<sup>133</sup> Um temor correlato é o de que instituições “grandes demais para falir” também se hajam tornado grandes demais para punir/processar, escapando a sanções. Por exemplo, o art. 11 da Lei nº 13.506/2017 outorga ao BCB a faculdade de “deixar de instaurar ou suspender, em qualquer fase que preceda a tomada da decisão de primeira instância, o processo administrativo destinado à apuração de infração” – desde que a infração não seja tipificada como grave, e a instituição se comprometa a cessar a prática sob investigação ou os seus efeitos lesivos. O problema é que “o termo de compromisso não importará em confissão quanto à matéria de fato, nem em reconhecimento da ilicitude da conduta analisada”.

Esse instituto é inspirado nos acordos (*settlement agreements*) sem admissão de ilicitude firmados pela *Security and Exchange Commission* (SEC), e pelo compromisso de cessação (aplicável à defesa da concorrência) introduzido na legislação nacional no art. 53 da Lei nº 8.884/1994. Nestes dois casos, porém, tal prática tornou-se malvista, pois não seria suficientemente dissuasória: a SEC tem, via de regra, incluído cláusulas de admissão de responsabilidade nos seus acordos (Schwab, 2014), e o art. 85 da Lei nº 12.529/2011 (que substituiu a Lei nº 8.884/1994) não faz previsão expressa da não admissão de ilicitude.

Parte do problema é a lacuna epistêmica sobre os verdadeiros efeitos incapacitatórios e dissuasórios da punição – afinal, não podemos fazer estudos randômicos controlados com potenciais infratores. Mesmo na justiça criminal, isso é objeto de polêmica – Roodman (2018) faz uma extensa análise de estudos sobre a deterrence do encarceramento, concluindo que falham em demonstrar tal efeito. Estudos que o confirmam enfatizam, na realidade, que o incremento da probabilidade de punição tem um impacto muito maior que o incremento da sua gravidade (Bun; Sarafidis; Kelaher, 2016).

parte do Lehman Brothers<sup>134</sup>), ou contaminar o mercado e agravar a probabilidade de crise sistêmica (como no exemplo do filme *Margin Call*, em que um banco de investimentos liquida todos os seus derivativos associados à hipotecas, levando a uma crise no respectivo mercado).

Portanto, a única maneira de assegurar que a expectativa de punição sirva como dissuasão é se houver a possibilidade de responsabilização *antes* da insolvência da instituição ou da ocorrência de crise<sup>135</sup>. Isso demanda a manutenção de uma estrutura de regulação e de supervisão robusta; em especial, demanda normas que permitam atribuir responsabilidades não só à pessoa jurídica, mas também a indivíduos. Trata-se de um aspecto enfatizado pelo *Financial Stability Board*, que sugere que as instituições devem ter um “mapeamento de responsabilidades” individuais: “Increasing individual accountability and responsibility can facilitate both *ex ante* prevention and remediation and more effective enforcement actions on individuals as well as firms” (FSB, 2017, 7).

Cabe ressaltar, porém, duas sérias desvantagens desta conclusão:

- a) Essa estrutura é um ônus para a economia (Haldane & Madouros, 2013) – o custo de observância não apenas implica menor eficiência do sistema (afinal, é um custo de transação), mas também serve de barreira à entrada de novos agentes, agravando a concentração no setor regulado e, indiretamente, *incentivando* a sua substituição por serviços não regulados (o *shadow banking*);
- b) Essa estrutura implica outorgar competências às autoridades de regulação e de supervisão – o que suscita considerações sobre concentração de poderes, captura regulatória, controle e legitimidade política<sup>136</sup>. Isso pode ser exemplificado pelo

---

<sup>134</sup> “Some corporate cultures actively conceal their flaws and irregularities, not only from the public or from regulators but also from others within the corporation itself because of the risk that wider knowledge of these issues might undermine the firm’s position. For example, let us look at Lehman Brothers’ use of the “Repo 105” accounting trick. Briefly, this was a repo, or repurchase agreement, valued at \$1.05 for every dollar, that was designed to look like a sale. Lehman Brothers paid more than five cents on the dollar to temporarily pay down the liabilities on its balance sheet before it repurchased the asset. The firm used this accounting trick in amounts totaling \$50 billion in late 2007 and 2008 to give itself a greater appearance of financial health—which, of course, was ultimately a failure.

[...]Lehman Brothers omitted its use of Repo 105 in its quarterly disclosures to the SEC and also neglected to tell its outside disclosure counsel” (Lo, 2015, 30).

<sup>135</sup> No limite, isso implicaria a definição de “infrações de mera conduta” (nas quais basta verificar a realização da conduta proibida, sem necessidade de que se concretize o risco que a lei busca prevenir); de fato, a jurisprudência atual tende a considerar a gestão temerária um crime de mera conduta (STF, HC 87.440/GO, Rel. Min. Carlos Britto, DJ 02.03.2007). Mas nosso argumento é mais fraco: a norma pode exigir que se prove a ocorrência de perdas ao investidor ou sócio (ou a instituições garantidoras), ainda que não se verifique a ocorrência de insolvência ou crise sistêmica.

<sup>136</sup> Não apenas por considerações de *pura* moralidade política, mas também econômicas:

“[...]the politics that we see operating everywhere else around us also determines whether societies suffer repeated banking crises (as in Argentina and the United States), or never suffer banking crises (as in Canada). [...] We mean [...] the way that the fundamental political institutions of a society structure the incentives of politicians, bankers, bank shareholders, depositors, debtors, and taxpayers to form coalitions in order to shape laws, policies, and

Banco Central do Brasil, que tem competência para supervisionar e regular o sistema financeiro (Lei nº 4.595/1964), suspender as atividades de instituições financeiras por insolvência ou pela prática reiterada de infrações (Lei nº 6.024/1974) e instaurar processos para aplicar àquelas e a seus gestores sanções administrativas – como multa, cassação da autorização para funcionar e inabilitação para ocupar cargo de administrador no setor (Lei nº 13.506/2017). Outrossim, embora não tenha atribuição para julgar ou oferecer denúncias sobre processos criminais, também tem a atribuição de informar a ocorrência de crimes financeiros ao Ministério Público e a faculdade de atuar como assistente da acusação (arts. 26 e 28 da Lei nº 7.492/1986).

Nosso escopo no presente texto, porém, não contempla tais objeções – em especial, porque a abordagem prudencial da regulação produz efeitos afins (custo regulatório e outorga de poderes a reguladores insulados da política). Ora nos limitamos à discussão sobre a necessidade de “salvar” instituições a fim de evitar crises sistêmicas (com fundamento no argumento do incentivo) e sobre a necessidade de estabelecer instrumentos coercitivos para mitigar o conseqüente risco moral. Para tratar das objeções supra, seria necessário investigar se (i) elas não são plenamente compensadas pela maior estabilidade econômica (em tese) oferecida por uma estrutura adequada, e (ii) se não existem maneiras de mitigá-las – e.g., agindo diretamente sobre a concentração de mercado ou adotando uma regulação mais favorável à entrada de novos agentes e de instituições que não ofereçam risco sistêmico (o que já ocorre), ou estabelecendo controles e repartindo as competências do regulador.

### **Conclusão**

Começamos com o problema de justiça distributiva gerado pelo *bail-out* – como pode ser justificado transferir recursos dos mais necessitados para banqueiros e investidores? Ao menos no sistema que ora temos, isso é uma consequência do argumento do incentivo; defendemos esse argumento contra as críticas de G. Cohen – reconhecendo, porém, o *peso moral* destas (em especial, a atenção a considerações morais no desenho de instituições).

Com isso, afastamos o argumento de que a assunção de riscos por instituições financeiras é um problema *apenas* para os seus respectivos investidores. Ainda, arguimos que o estabelecimento de limites *prudenciais* (i.e., a imposição de custos, como capital, para

---

regulations in their favor – often at the expense of everyone else. In this view, a country does not ‘choose’ its banking system: rather it gets a banking system that is consistent with the institutions that govern its distribution of political power” (Calomiris & Haber, 2015, 4).



investimentos arriscados) não é suficiente para mitigar o risco moral gerado pelo estabelecimento de garantias contra o risco sistêmico; isso nos levou à defesa do estabelecimento de normas coercitivas dissuadindo a assunção de riscos em excesso – e de sua aplicação mesmo antes da ocorrência dos riscos que procuram evitar (i.e., da insolvência de instituições ou da ocorrência de crise).

O que poderia, enfim, ser dito como resposta ao banqueiro de *The Last Laugh*? Podemos reconhecer que o que ele diz é verdade: que são as economias do público em geral que sofrem com crises financeiras; e que, justamente por isso, ele não deveria dirigir uma instituição financeira – pois seus gestores deveriam estar cientes de sua responsabilidade especial perante o público. Caberia ao direito e às instituições de regulação e supervisão lembrá-los disso.

## Conclusão

As duas epígrafes desta tese fornecem um contraste entre duas visões, de sucesso e de fracasso, do filósofo analítico e do intelectual em geral. A primeira foi retirada de uma cômica fantasia de Terry Pratchett, na qual um grupo de ratos, após comer restos de itens mágicos, adquire habilidades cognitivas e comunicativas em nível humano e, de “brinde”, os problemas existenciais e sociais correspondentes. A citação revela a epifania do pragmático Bronzeado Intenso<sup>137</sup>, líder do grupo de “desarmadores de armadilhas”, o qual, após um dilema moral, percebe a importância das reflexões abstratas de seu colega Perigoso Feijão, que desarma as complexas armadilhas conhecidas como *ideias*. Corresponde ao filósofo analítico que identifica um paradoxo e mostra que se trata de um “pseudo-problema”; ao fazê-lo com sucesso, pode ser tentador pensar que o empreendimento foi uma perda de tempo. Como o goleiro e o gestor de riscos, a importância dessa atividade só é destacada quando há uma falha; é o caso da segunda epígrafe em que, mesmo com a melhor das intenções (descobrir a proposição mais *importante* do universo), os filósofos são vítimas de sua própria astúcia com descrições definidas, enredando-se num paradoxo de auto-referência<sup>138</sup>.

O receio de sofrer uma sorte afim ronda os textos supra; não visamos à qualquer tese revolucionária, mas a defender alguns juízos comuns dos problemas filosóficos que os ameaçam. Respectivamente, por capítulo:

- 1) Autoridades são úteis e suas ordens podem ser legítimas, ainda que discordemos delas – ainda assim, precisamos conservar a independência de nossas próprias avaliações dessas ordens a fim de verificar sua legitimidade.
- 2) Normas jurídicas podem incorporar conceitos morais; ao invés de eles *sempre* servirem como uma lacuna para que intérpretes exerçam discricionariedade, podem servir como um parâmetro para interpretar e conhecer o significado da norma.
- 3) O problema de futuros contingentes é menos a indeterminação de seu valor de verdade (como podemos acreditar em juízos futuros sem considerá-los verdadeiros?) que a inafastável incerteza que ronda alguns juízos – em especial, os que embutem a representação da decisão do próprio agente.

---

<sup>137</sup> Tendo aprendido a ler antes de aprender o significado das palavras, os ratos escolheram seus nomes a partir dos itens presentes no depósito de lixo.

<sup>138</sup> De forma análoga, mas sem o brinde do paradoxo lógico, no *Guia do Mochileiro das Galáxias* de Douglas Adams, uma civilização avançada constrói o mais perfeito computador para responder à “questão fundamental” – que nunca é enunciada, mas poderíamos pensar que é equivalente *salva veritatem* à “qual a resposta para a pergunta fundamental?” – e obtém como *output* “42”.

- 4) Incerteza não implica que a autoridade deva manter-se passiva; pelo contrário, uma de suas funções é reduzir a incerteza de interações sociais – em especial quando se trata de supervisão bancária.
- 5) Não se pode defletir a reprovação moral por uma decisão por meio do fatalismo – alegando que as consequências, ou que a própria decisão, eram “inevitáveis”. O agente é responsável por contribuir para o resultado e pelas preferências que tem.
- 6) É legítimo instituir sanções para financistas que incorram em riscos excessivos; a disciplina de mercado e a regulação prudencial não são suficientes para afastar os riscos sociais correspondentes.

Contudo, ao esboçar tais defesas, mergulhamos em investigações relacionadas à teoria da decisão racional e à filosofia política. É possível que, após estes textos, tenhamos mais dúvidas do que no começo – em especial quanto a questões importantes que apenas tangenciamos.

### **Racionalidade**

Primeiramente, usamos o termo *racionalidade* ao longo dos textos sem especificá-lo; por ele, referimos, os axiomas básicos da teoria da utilidade esperada subjetiva em contextos de informação completa. Isso captura uma noção “fraca”, subjetivista (ela não diz nada sobre os fins do agente) e precisa de racionalidade como *coerência* entre meios e fins. Essa noção se aplica a *problemas de decisão* caracterizados por:

- a) um agente (a “unidade básica de decisão” dentro de um modelo – Guerrien, 2002, 21);
- b) um conjunto de ações (mutuamente excludentes) que ele pode tomar;
- c) um conjunto de estados de coisas (mundos possíveis).

A decisão envolve uma “regra de aprendizado”, que define de que forma o agente representa estados de coisas: a) começa com uma distribuição de probabilidade *a priori* que atribui um valor correspondente a cada mundo possível; b) à medida que recebe nova informação, o agente atualiza sua distribuição de probabilidades mediante *condicionalização* (usando uma regra apropriada, como o teorema de Bayes) – o que resulta numa distribuição de probabilidade *a posteriori* (Bostrom, 2014, 10). Outrossim, o agente precisa de uma regra de decisão, que define como ele *valoriza* estados de coisas; assim, atribui-se-lhe uma função-utilidade – uma ordem *completa, contínua e transitiva* de preferências que associa um número

(utilidade) a cada estado de coisas representando sua “desejabilidade” (Bostrom, 2011, 11). Ao decidir, o agente seleciona a ação com maior utilidade esperada; para tanto, ele deve computar: a) para cada mundo possível  $M$  e cada ação  $A$ , a respectiva probabilidade *a posteriori*  $p$  de atingir o mundo possível  $M$  condicionada à escolha da ação  $A$ ; b) para cada mundo possível  $M$ , o produto ( $p \times u$ ) dessa probabilidade  $p$  e da respectiva utilidade  $u$ ; c) para cada ação, a soma de cada um desses produtos (Bostrom, 2011, 11)<sup>139</sup>.

Essa concepção pode ser estendida, com pequenas alterações, para teorias distintas da teoria da utilidade esperada, como a *utilidade esperada ponderada pelo risco* de Buchak (2013) – em que a aversão a risco é computada como um parâmetro adicional (e não apenas como a “curvatura” da função utilidade); ela contrasta com definições vagas e amplas, como a de que racionalidade é “fazer melhor o que fazemos naturalmente” (Zagzebski, 2012, 30 – trad. livre). Também se distingue de aspectos frisados pela prática filosófica, como o uso normativo/avaliativo do termo e a capacidade de *dar razões* – somos o “animal que dá razões” (Parfit, 2011, 31). Seria tentador, para frisar tal distinção, substituir o termo “racional” por algo conexo e sem “carga normativa”, como “inteligência” – a exemplo de Bostrom (2013, 107-08); no entanto, esta noção também é objeto de contenda (vide, e.g., a crítica ao argumento de Turing por Searle, 1980), e frequentemente definida em termos de uma teoria da decisão racional, como em Russell & Norvig (2010, 36 e s.).

Paradoxos surgem justamente quando identificamos essa definição/descrição do conceito de racionalidade com um uso normativo geral do conceito – i.e., a ideia de que racionalidade implica *performance ótima* (a capacidade de dar “a melhor resposta”) em *qualquer* situação possível: “To be rational is to be perfectly rational. Choosing less than the best is therefore irrational” (Sorensen, 2006, 216). Pode ser estranho negar essa tese afirmando, e.g.: “Some agents who are led to ruin in this way are perfectly rational. It is just that certain situations exploit rational agents who are unable to self-bind.” (Arntzenius et al, 2004, 269). É plausível que, para qualquer definição suficientemente *precisa* de racionalidade, pode-se construir, dentro do conjunto potencialmente infinito de experimentos mentais, alguns cenários (servindo de “bomba de intuição” – para usar a metáfora de Dennett, 2013) onde agentes racionais (segundo tal definição) performam pior que agentes “limitados”. Não mais trivial, um *deus ex machina* pode premiar agentes que cometam contradições com uma frequência específica. O que está por trás dessa sugestão é que, se uma concepção de racionalidade é tão

---

<sup>139</sup> Ou seja, a utilidade esperada de uma ação  $A$  equivale à *média* das respectivas utilidades por mundo possível, ponderadas pela probabilidade (condicionada à adoção da ação  $A$ ) de cada um destes.

precisa que pode ser instanciada por um determinado *algoritmo* (como a AIXI de Hutter, 2005), ela só é *ótima* no conjunto de mundos possíveis onde as especificações de sua “regra de aprendizado” são apropriadas; mundos com um *deus ex machina* caprichoso tendem, e.g., a desfavorecer agentes *bayesianos* com um viés indutivo<sup>140</sup>.

O preço pago pela amplitude do conceito *normativo* de racionalidade é torná-lo, “num sentido importante, hermenêutico ou interpretativo” (Elster, 1993, 179). Isso nos permite dizer que *pode ser* racional (no sentido normativo) cooperar no primeiro *round* de um dilema do prisioneiro reiterado *n* vezes *se os agentes têm a possibilidade de fazê-lo*, ainda que agentes com “conhecimento comum da racionalidade” jamais o fizessem; o que é contra-intuitivo é que estes agentes “ilimitadamente racionais” parecem ter o que, na prática, é uma *limitação* (que os impediria, e.g., de apelar para estratégias de destruição mútua) – análoga à limitação de agentes incapazes de randomizar em jogos em que os únicos equilíbrios de Nash envolvam estratégias mistas. Nisso, divergimos de Elster (1993, 182), para quem o caráter paradoxal de jogos sequenciais necessariamente implicaria a ausência de uma “solução”; eles são contrastados com problemas matemáticos (Elster, 1993, 179), como a *Hipótese do Continuum*, que são indecidíveis em relação a um conjunto de axiomas (i.e., prova-se sua indecidibilidade ao demonstrar que são independentes de tais axiomas) – mas que podem ser resolvidos pela inclusão de novos axiomas. Recusamos tal contraste, porque: a) a possibilidade de diversos equilíbrios não é, em si, paradoxal; b) uma definição mais precisa do cenário e das crenças dos agentes pode dissolver o paradoxo de jogos sequenciais (i.e., o agente é mais propenso a cooperar à medida que aumenta o número de *rounds*, ou a diferença de *pay-off* entre as soluções de conflito e de cooperação, ou a probabilidade atribuída à cooperação dos adversários, etc.).

Não estamos adotando a posição de que a noção de racionalidade serve *apenas* para prever ou descrever o comportamento; como Elster, podemos dizer que “We care about rationality because we want to be rational and want to know what rationality requires us to do” (Elster, 1993, 189). Contudo, o que a racionalidade *demand*a pode depender duma representação mais precisa da situação do agente. Algo afim se aplica a como complementar o problema de decisão quando há restrições informacionais: nenhum agente possível poderia representar todos os mundos possíveis relevantes e atribuir-lhes uma distribuição de probabilidades e uma ordenação de preferências completa; como preencher essa lacuna, porém, pode variar de acordo com outras especificações/limitações do agente e do problema.

---

<sup>140</sup> Ao definir uma regra de aprendizado, não estabelecemos de que forma o agente forma sua distribuição de probabilidade *a priori*; em computação, busca-se definir uma regra, viesada em favor de “mundos possíveis mais *simples*”, que se aproxime de um princípio de indução.

## **Incerteza e Autoridade**

O que distinguiria a autoridade, na *concepção de serviço* de Raz, de outras divisões hierárquicas é que seus enunciados são *peremptórios* (pretende-se que substituam as razões para agir do sujeito), *dependem* das razões de decidir do próprio agente (o que a distingue de coação explícita), e a obediência é *normalmente justificada* em função destas mesmas razões. A autoridade se justifica porque é *útil*, como uma instituição ou uma tecnologia. Vimos, porém, que essa concepção pode levar a problemas quando houver desacordo entre a autoridade e o sujeito (capítulo um), mormente quando este se encontrar numa situação *informacionalmente* equivalente ou superior (capítulo quarto).

Uma das possíveis respostas a esses óbices é abandonar completamente a concepção de serviço, assumindo que o problema é com o viés instrumental do qual ela depende; contudo, esse caráter instrumental é justamente a força dessa concepção, tornando-a, *prima facie*, compatível com teorias da decisão como a teoria da utilidade esperada. Na realidade, tais obstáculos, arguimos, referem-se a questões derivadas da própria teoria da racionalidade (como enfatizamos no primeiro e quarto capítulos); novamente, pode-se reiterar a crítica: dever-se-ia abandonar, na realidade, a própria concepção de racionalidade instrumental – i.e., a ideia de que uma decisão/ação é racional quando os meios são coerentes com os fins. Há algo importante nessa crítica: assim como a teoria da decisão “muda” quando o agente é inserido num contexto estratégico (opções que seriam irracionais em contextos paramétricos, como randomização e adoção de estratégias pareto-inferiores, tornam-se opções em jogos), ela também pode “mudar” em contextos sociais (com redes de cooperadores de diferentes tamanhos e complexidades), onde há sinalização, aprendizado, formação de reputação, estratégias correlatas, etc. Além disso, problemas de decisão enfrentados por conjuntos de seres humanos têm características próprias, relativas às limitações específicas desse tipo de agente. Mas não se pode simplesmente abandonar a concepção de racionalidade instrumental, sob pena de contradição (pois ela nada mais é que uma teoria da racionalidade como coerência), nem oferecer uma concepção *concorrente* que leve o sujeito a uma variedade de respostas – pois então precisaríamos de uma meta-teoria para explicar como se arbitra entre as conclusões discordantes de ambas as concepções.

Outro caminho para resolver tais óbices é enfatizar o papel de autoridades como soluções *potencialmente* estáveis, eficientes, e públicas para problemas de coordenação (Myerson, 2009, 1122-23) – em especial, o problema da incerteza sobre como conduzir-se em

grupo. Essa foi uma das alternativas para responder ao dilema de Greenspan, e é uma ideia usual em filosofia do direito: em Hobbes (2002, 95-97; Cap. XIV, seções 10-12), um dos principais problemas no estado de natureza é que, como cada homem é juiz de si mesmo, não se pode confiar que todos agirão de modo pacífico. Mesmo se este problema é parcialmente solúvel por normas consuetudinárias (como as que regeram nossas práticas sociais durante a maior parte da história – e ainda regem práticas comuns fundamentais, como linguagens naturais), como no que Hart (1994, 79 e s.) chama de “sistema jurídico primitivo”, ainda há problemas que a introdução de regras secundárias (i.e., “normas sobre normas”: sobre quem pode aplicá-las ou alterá-las) visa a resolver: a incerteza, o desacordo e a rigidez em relação às regras primárias de obrigação. As regras que definem os poderes e deveres de autoridades são um exemplo importante de regras secundárias – e, formando um *loop* ainda muito debatido em filosofia do direito, um problema do “ovo e da galinha” (Shapiro, 2013, 39-40), uma das funções dessas autoridades é justamente tomar decisões sobre esse tipo de regras. Isso permite que sistemas baseados em autoridades processem informações e tomem decisões de maneira refletida – o que não ocorreria em “ordens espontâneas”, como, alegadamente, seria o caso de normas costumeiras e de mercados.

### **Crise e Política**

Esse papel de “ponto focal” – como um mecanismo de coordenação que permite aos agentes desviarem de “maus equilíbrios” – se estende à atuação de “autoridades” macroeconômicas mesmo que, nesse papel, *não* ajam como autoridades jurídicas (que criam ou reconhecem normas). Pode-se, por precisão conceitual, dizer que essas instituições não são propriamente autoridades (não fornecem razões protegidas, *preemptivas*), ou só o são num sentido derivado; porém, isso obscurece o fato de que elas acarretam problemas filosóficos e práticos afins. Assim como um sistema jurídico pode ser visto como uma hierarquia de instituições, de oficiais e de sujeitos mediados por normas, um sistema financeiro pode ser visto como uma hierarquia de instituições e sujeitos mediados por acordos e relações econômicas – e, nos dois casos, pode-se perguntar se os resultados e a respectiva divisão de poder são adequados, equitativos ou justos (Pistor, 2017, 185-86).

Outrossim, o poder de sanção jurídica funciona antes pela *ameaça* de seu uso que pelo uso de fato; os poderes de autoridades monetárias têm um efeito indireto análogo: tanto nos EUA quanto na União Europeia, a Crise começou a arrefecer quando os respectivos Bancos Centrais e autoridades governamentais assumiram o *compromisso crível* de que fariam o

necessário para reverter a crise – notavelmente, a crise do euro arrefeceu após o discurso do “custe o que custar” de Mario Draghi (Heijden, Beetsma & Romp, 2017). Essas instituições se destacam por seus poderes econômicos peculiares; elas não criam riqueza, mas a redistribuem (e é admissível indagar se a distribuição resultante é justificada), por meio de operações de crédito ou aquisição de ativos (Piketty, 2014, 669), com potencial de criação monetária, *prima facie*, infinito:

[...] um banco central pode, no espaço de um segundo, criar tantos bilhões quanto desejar e entregá-los a uma empresa ou a um governo em necessidade.

[...] Bancos Centrais poderiam decidir comprar todas as empresas de um país, todos os imóveis, financiar a transição energética, investir nas universidades e ter controle de toda a economia (Piketty, 2014, 673-74).

É interessante situar esse ponto nas discussões contemporâneas sobre autonomia ou independência de Bancos Centrais. A existência de tais instituições e sua insulação do poder político são justificadas principalmente como um mecanismo para limitar a capacidade do governo central de emitir moeda e de, mediante a produção de inflação, reduzir seus passivos definidos em moeda nacional; porém, em sistemas de reservas fracionárias e afins, instituições financeiras também são emissores *privados* de moeda – logo, não basta limitar a capacidade de emissão pública para garantir a estabilidade. Assim, Bancos Centrais operam como “bancos dos bancos”, como *prestamistas de última instância* que provêm liquidez ao sistema; muitos Bancos Centrais (e agências reguladoras com essa função específica) também desempenham o papel de regulação e supervisão do sistema financeiro. Essas três funções (autoridade monetária, provedor de liquidez e regulador do sistema financeiro) cresceram como resposta à Crise. Pode-se discutir se há efeitos negativos em tal crescimento e acúmulo; e.g., se o desempenho de uma dessas tarefas não implica comprometer, ainda que parcialmente, o desempenho de outra (Souza, 2009, 123). Ainda, a concentração de poder é vista com suspeição; na ausência de controle e regulação externos (Tucker, 2018, 391), atrai a pecha de “cidadãos super-poderosos” (*overmighty citizens*). Sob esse argumento, em 1997, ao conceder independência ao Banco da Inglaterra enquanto autoridade monetária, o Reino Unido retirou-lhe a função de supervisão financeira (Tucker, 2018, ix).

Apenas tangenciamos tais discussões sobre desenho institucional, ainda que sejam a “continuação natural” das discussões deste trabalho. Limitamo-nos a apresentar razões (e a afastar algumas objeções) que justificariam a *existência* de autoridades regulatórias e de alguns arranjos institucionais especiais – sem, contudo, defender uma posição sobre como essas



autoridades devem se relacionar com outras instituições estatais e com valores políticos, como controle democrático e participação popular.

Esse silêncio não se deve apenas à complexidade do tema, nem às chamadas crises da democracia e do capitalismo<sup>141</sup>, mas ao fato de que implicaria entrar em questões específicas em relação ao contexto atual. Nossas instituições, e algumas de nossas normas sociais, podem ser vistas como tecnologias– artefatos úteis para cooperar de forma estável, eficiente e equitativa; o que chamamos de moeda, estado, participação e direito tem mudado continuamente, e é lamentável que, por vezes, as vejamos como naturais ou eternizáveis – ou, pior ainda, que precisemos lutar para impedir sua degradação e o retorno a arranjos institucionais mais primitivos. Ao assim vê-las, caímos no tipo de *armadilha conceitual* contra a qual busquei alertar ao citar o caso de Perigoso Feijão – na epígrafe desta tese e no começo desta conclusão.

Nos últimos 100 anos, nossas sociedades mudaram de formas até entãojamais imaginadas: a expectativa de vida dobrou, as distâncias encurtaram e o conhecimento (geral) se disseminou. Esses fatos nos permitem uma boa dose de otimismo; mas, ao pensar nesses termos, é notável que pouco haja mudado na maneira como escolhemos nossas autoridades e representantes e como agregamos nossas opiniões (sendo que desde Condorcet se conhecem sistemas de votação formalmente melhores);que a concentração de poder econômico seja tratada como um problema muito menos relevante que(e por vezes, sem conexão com) o problema da concentração do poder político.No entanto, tanto falhas políticas quanto desigualdades contribuem para a ocorrência de crises financeiras e seus efeitos perversos (Dagher, 2018); sem progressos nessas áreas, é difícil ser otimista em relação a inovações institucionais marginais – elas não serão suficientes para evitar a próxima crise.

---

<sup>141</sup> Quando se estuda a história desses dois conceitos, nota-se que é um lugar-comum falar em “crise”.

## Referências

- Acemoglu, D.; Robinson, J. A. (2013). *Why nations fail: The origins of power, prosperity, and poverty*. London: Profile.
- Adamou, A.; Peters, O. (2016). “Dynamics of inequality”. *Significance* 13(3): 32-37. doi:10.1111/j.1740-9713.2016.00918.x
- Aristóteles. (1963). *Categories and De Interpretatione*. Trad. J. L. Ackrill. Oxford: Clarendon Press.
- Aristóteles. (1966). *The Nicomachean Ethics of Aristotle*. Trad. W. D. Ross. London: Oxford University Press.
- Arntzenius, F.; Elga, A.; Hawthorne, J. (2004). “Bayesianism, Infinite Decisions, and Binding”. *Mind*, Volume 113, Issue 450, Abril/2004, pp. 251–283, <https://doi.org/10.1093/mind/113.450.251>.
- Arntzenius, Frank (2008). “No regrets, or: Edith Piaf revamps decision theory”. *Erkenntnis*, 68 (2):277-297.
- Arntzenius, Frank (2014). “Utilitarianism, decision theory and eternity”. *Philosophical Perspectives* 28 (1):31-58.
- Arrow, K. (1950). "A Difficulty in the Concept of Social Welfare". *Journal of Political Economy*. 58 (4): 328–346. doi:10.1086/256963
- Arrow, K. (1963). “Uncertainty and the Welfare Economics of Medical Care”. *The American Economic Review*, 53(5), 941-973. Disponível em: <<http://www.jstor.org/stable/1812044>>
- Arrow, K.; Hurwicz, L. (1977). Appendix: “An optimality criterion for decision-making under ignorance”. In K. Arrow & L. Hurwicz (Eds.), *Studies in Resource Allocation Processes*, pp. 461-472. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511752940.015
- Artemov, S.; Kuznets, R. (2014). “Logical omniscience as infeasibility”. *Annals of Pure and Applied Logic*, 165(1), 6-25. doi:10.1016/j.apal.2013.07.003
- Atria, F.; Michelon Jr.; C. F. (2007). “Una crítica al principio de diferencia”. In: Squella, A. (Org.). *Filosofía y Política en Rawls*, Valparaíso: Edeval, v. 1, pp. 214-231.
- Aumann, Robert J. (1976). “Agreeing to Disagree”. *Annals of Statistics*, v. 4, nº 6, pp. 1236-1239. Disponível em: <<https://projecteuclid.org/euclid.aos/1176343654>>.

- Baliga, S.; Sjöström, T. (2012). “The Hobbesian Trap”. In: *The Oxford Handbook of the Economics of Peace and Conflict*, Oxford Handbooks Online. doi:10.1093/oxfordhb/9780195392777.013.0005.
- Barbosa Filho, B.(1999). “Saber, Fazer e Tempo: Uma Nota sobre Aristóteles”. In: Rocha, Ethel M. et al.(org.). *Verdade, Conhecimento e Ação: ensaios em Homenagem a Guido Antônio de Almeida e Raul Landim Filho*. São Paulo: Loyola, pp. 15-24.
- Barbosa Filho, B. (2004). “Nota sobre o conceito aristotélico de verdade”. *Cadernos de História e Filosofia da Ciência*, CLE - UNICAMP, v. 13, n.2, pp. 233-243.
- Bardsley, N.; Mehta, J.; Starmer, C.; Sugden, Robert. (2006). “The Nature of Saliency Revisited: Cognitive Hierarchy Theory versus Team Reasoning”. Discussion Papers 2006-17, The Centre for Decision Research and Experimental Economics, School of Economics, University of Nottingham. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.111.1882&rep=rep1&type=pdf>>
- Barros, G. (2010). “Herbert A. Simon and the concept of rationality: boundaries and procedures”. *Brazilian Journal of Political Economy*, 30(3), 455-472. doi:10.1590/S0101-31572010000300006
- Becker, G. S. (1993). “Nobel Lecture: The Economic Way of Looking at Behavior”. *Journal of Political Economy*, 101(3), 385-409. doi:10.1086/261880
- Bernoulli, D. (1954). “Exposition of a New Theory on the Measurement of Risk”. *Econometrica*, 22(1): 23-36, Jan/1954. Disponível em: <[jstor.org/stable/1909829](http://www.jstor.org/stable/1909829)>
- Bertram, C. (1997). “Political Justification, Theoretical Complexity, and Democratic Community”. *Ethics*, Vol. 107, No. 4. (Jul.; 1997), pp. 563-583. Disponível em: <<http://www.jstor.org/stable/2382295>>
- Betz, G. (2006). *Prediction or prophecy? The boundaries of economic foreknowledge and their socio-political consequences*. Dordrecht: Springer.
- Binmore, K. G. (1997). “Rationality and backward induction”. *Journal of Economic Methodology*, v. 4, n° 1, pp. 23-41. Disponível em: <<http://www.tandfonline.com/doi/citedby/10.1080/135017897000000002>>.
- Binmore, K. G. (2001). “Evolutionary Social Theory: Reply to Robert Sugden”. *The Economic Journal*, 111(469), 244-248. doi:10.1111/1468-0297.00605
- Binmore, K. G. (2005). *Natural Justice*. Princeton: Princeton University Press.
- Binmore, K. G. (2007a). *Game theory: A very short introduction*. New York: Oxford University Press.

- Binmore, K. G. (2007b). *Playing for real: A text on game theory*. Oxford: Oxford University Press.
- Binmore, K. G. (2009). *Rational Decisions*. Princeton: Princeton University Press, 2009.
- Binmore, K.; Stewart, L.; Voorhoeve, A. (2012). “How much ambiguity aversion? Finding indifference between Ellsberg's risky and ambiguous bets”. *Journal of Risk and Uncertainty*, 45:215–238. doi:10.1007/s11166-012-9155-3
- Blackburn S. (2010). *Practical Tortoise Raising*. Oxford: Oxford University Press.
- Blackburn, S. (2006). *A República de Platão: uma biografia*. Trad. Roberto Franco Valente. Rio de Janeiro: Jorge Zahar.
- Blumenson, E. (2015). “The Limits of Moral Argument: Reason and Conviction in Tadros’ Philosophy of Punishment”. *Law Ethics & Philosophy* [online], (3), 30-55. Disponível em: [www.raco.cat/index.php/LEAP/article/view/313497/403617](http://www.raco.cat/index.php/LEAP/article/view/313497/403617)
- Borak, Donna (2017). "House votes to kill Dodd-Frank. Now what?". CNN Money, 8 de junho de 2017. Disponível em: <https://web.archive.org/web/20171129194402/http://money.cnn.com/2017/06/08/news/economy/house-dodd-frank-repeal/index.html>
- Bostrom, N. (2006). “Sleeping Beauty and Self-location: A Hybrid Model”. *Synthese*, 157(1), 59-78. doi:10.1007/s11229-006-9010-7
- Bostrom, N. “Infinite Ethics”. *Analysis and Physics*, 10 (2011): 9–59. Disponível em: <https://nickbostrom.com/ethics/infinite.pdf>
- Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford: Oxford University Press.
- Bratman, M. E. (2015). “Précis of Shared Agency: A Planning Theory of Acting Together”. *Journal of Social Ontology*, 1(1): 1-5. doi:10.1515/jso-2014-0042.
- Bremmer, I.; Keat, P. (2010). *The fat tail: The power of political knowledge for strategic investing*. New York: Oxford University Press.
- Britto, C. A. (2001). “O Regime Constitucional dos Tribunais de Contas”. *Revista Diálogo Jurídico*, Salvador, CAJ - Centro de Atualização Jurídica, v.I, n.º.9, dezembro/2001. Disponível em: <https://bdjur.stj.jus.br/jspui/handle/2011/22287>.
- Bruin, B. P. (2015). *Ethics and the global financial crisis: Why incompetence is worse than greed*. Cambridge, UK: Cambridge University Press.
- Buchak, L. (2013). *Risk and Rationality*. Oxford: Oxford University Press.

- Buchak, L. (2017). "Taking Risks behind the Veil of Ignorance". *Ethics*, 127(3), 610-644. doi:10.1086/690070
- Bun, M.; Sarafidis, V.; Kelaher, R. (2016). "Crime, Deterrence and Punishment Revisited". *UvA-Econometrics Working Papers* N. 16-02, Universiteit van Amsterdam, Dept. of Econometrics. Disponível em: <<https://ideas.repec.org/p/ame/wpaper/1602.html>>
- Burgos, A. V.; Oliveira Filho, J. D.; Soares, M. V. S.; Almeida, R. S. (2017). "Distributed ledger technical research in Central Bank of Brazil". *Positioning report* do Banco central do Brasil. Disponível em: <[https://www.bcb.gov.br/htms/public/microcredito/Distributed\\_ledger\\_technical\\_research\\_in\\_Central\\_Bank\\_of\\_Brazil.pdf](https://www.bcb.gov.br/htms/public/microcredito/Distributed_ledger_technical_research_in_Central_Bank_of_Brazil.pdf)>
- Carroll, L. (pseudônimo de Charles L. Dodgson). (1895). "What the Tortoise said to Achilles". *Mind*, IV(14), 278-280. doi:10.1093/mind/iv.14.278
- Chalmers, I. (2003). "Trying to do more Good than Harm in Policy and Practice: The Role of Rigorous, Transparent, Up-to-Date Evaluations". *The Annals of the American Academy of Political and Social Science*, 589(1), 22-40. doi:10.1177/0002716203254762
- Chang, R. (2017). "Hard Choices". *Journal of the American Philosophical Association*, 3(1), 1-21. doi:10.1017/apa.2017.7
- Ćirković, M. M.; Sandberg, A.; Bostrom, N. (2010). "Anthropic Shadow: Observation Selection Effects and Human Extinction Risks". *Risk Analysis*, 30(10), 1495-1506. doi:10.1111/j.1539-6924.2010.01460.x
- Coase, R. (1960). "The Problem of Social Cost." *The Journal of Law & Economics*, 3, 1-44. Disponível em : <<http://www.jstor.org/stable/724810>>
- Coase, R. H. (1991). "The Institutional Structure of Production". *Journal des Économistes et des Études Humaines*, 2(4): 431-440. doi:10.1515/jeeh-1991-0401
- Coblenz, M. (2015). "The Fight Goes on Forever: 'Limited Government' and the First Bank of the United States". *Southern Illinois University Law Journal*. 39: 391-451.
- Cohen, G. A. (2001) *If You're an Egalitarian, How Come You're so Rich?*. Cambridge, MA: Harvard University Press.
- Cohen, G. A. (2008) *Rescuing Justice and Equality*. Cambridge, MA: Harvard University Press.
- Cohen, G. A. (2011). *On the Currency of Egalitarian Justice, and Other Essays in Political Philosophy*. Princeton: Princeton University Press.
- Coleman, J. L.; Shapiro, S. (eds). (2001). *Hart's Postscript: Essays On The Postscript to the Concept of Law*, Oxford: Oxford University Press.
- Coleman, J. L. (2003). *Markets, morals, and the law*. Oxford: Oxford University Press.

- Combe, E. (2017). *Précis d'économie*. Paris: Presses universitaires de France.
- Cooke, R. (1986). "Conceptual fallacies in subjective probability". *Topoi* 5 (1):21-27. Doi: 10.1007/BF00137826
- Da Silva, S.; Matsushita, R. (2016). "The St. Petersburg paradox: An experimental solution". *Physica A: Statistical Mechanics and its Applications*, 445, 66-74. doi:10.1016/j.physa.2015.10.045
- Dagher, J. (2018). "Regulatory Cycles: Revisiting the Political Economy of Financial Crises". *IMF Working Papers*, 18(8), 1. doi:10.5089/9781484337745.001
- Dedeo, S. (2013) "The Coin Toss and the Love Triangle". *Nautilus*, Issue 2: "Uncertainty", 13 de junho de 2013. Disponível em: <<http://nautil.us/issue/2/uncertainty/the-coin-toss-and-the-love-triangle>>
- Dedeo, S. (2017). "Is Tribalism a Natural Malfunction?". *Nautilus*, Issue 52: "The Hive", 13 de junho de 2013. Disponível em: <<http://nautil.us/issue/52/the-hive/is-tribalism-a-natural-malfunction>>
- Del Río, C. A. (2017). "Uso de la tecnología de contabilidad distribuida por los bancos centrales: Una revisión". *Enfoque UTE*, 8(5), pp. 1 - 13. doi:10.29019/enfoqueute.v8n5.175
- Delfim Netto, A. (2016). "O espírito democrático da tecnocracia". *Valor Econômico*, São Paulo, 19 de julho de 2016. Caderno A02.
- Dennett, D. C. (1984). "I Could Not Have Done Otherwise-So What?". *Journal of Philosophy*, 81(10), 553-565. doi:10.5840/jphil1984811022
- Dennett, D. C. (1986). "The Moral First Aid Manual". In: The Tanner Lectures on Human Values, na Universidade de Michigan em 7 e 8 de Novembro de 1986. Disponível em: <[https://tannerlectures.utah.edu/\\_documents/a-to-z/d/dennett88.pdf](https://tannerlectures.utah.edu/_documents/a-to-z/d/dennett88.pdf)>
- Dennett, D. C. (2013). *Intuition Pumps and Other Tools for Thinking*. New York: W. W. Norton & Company.
- Dennett, D. C. (2018). *From bacteria to Bach and back: the evolution of minds*. New York: Norton & Co.
- Dietrich, F., & List, C. (2012). "Where do preferences come from?". *International Journal of Game Theory*, 42(3), 613-637. doi:10.1007/s00182-012-0333-y
- Dummett, M. (1991). *The Logical Basis of Metaphysics*. Cambridge, MA: Harvard University Press, 1991.
- Dworkin, R. (1967) "The Model of Rules" (1967). Faculty Scholarship Series. Paper 3609. Disponível em: [http://digitalcommons.law.yale.edu/fss\\_papers/3609](http://digitalcommons.law.yale.edu/fss_papers/3609)

- Dworkin, R. (1999). *O Império do Direito*. Trad. Jefferson Luis Camargo. São Paulo: Martins Fontes.
- Dworkin, R. (2000). *Sovereign Virtue: The Theory and Practice of Equality*. Cambridge, Mass: Harvard University Press.
- Dworkin, R. (2002). *Levando os direitos a sério*. Trad. Nelson Boeira. São Paulo: Martins Fontes.
- Dworkin, R. (2006). *Justice in Robes*. Cambridge, Massachusetts: Belknap Harvard University Press.
- Dworkin, R. (2011). *Justice for Hedgehogs*. Cambridge, Massachusetts: Belknap Press (Harvard University Press).
- Eisenbach, T. M.; Haughwout, A.; Hirtle, B.; Kovner, A.; Lucca, D. O.; Plosser, M. C. (2015). "Supervising Large, Complex Financial Institutions: What Do Supervisors Do?". *SSRN Electronic Journal. Federal Reserve Bank of New York Staff Reports*, n. 729. doi:10.2139/ssrn.2612020
- Elga, A. (2000). "Self-locating belief and the Sleeping Beauty problem". *Analysis*, 60(2), 143-147. doi:10.1093/analys/60.2.143
- Ellsberg, D. (1961). "Risk, Ambiguity, and the Savage Axioms". *Quarterly Journal of Economics*, 75 (4): 643–669. doi:10.2307/1884324
- Elster, J. (1977). "Ulysses and the Sirens: A theory of imperfect rationality". *Social Science Information*, 16(5), 469-526. doi:10.1177/053901847701600501
- Elster, J. (1988). "Taming chance: randomization in individual and social decisions". In *The Tanner Lecture on Human Values*, em 6 e 7 de maio de 1987, Brasenose College, Oxford University. Disponível em: <tannerlectures.utah.edu/\_documents/a-to-z/e/Elster88.pdf>
- Elster, J. (1993). "Some Unresolved Problems in the Theory of Rational Behavior". *Acta Sociologica*, 36(3), 179-189. doi:10.1177/000169939303600303
- Elster, J. (2016). *Sour grapes: Studies in the subversion of rationality*. Cambridge, UK: Cambridge University Press.
- Emon, A. M. (2009) "To Most Likely Know the Law: Objectivity, Authority, and Interpretation in Islamic Law". *Hebraic Political Studies*, n.4, set/2009, pp. 415-40. Disponível em: <<http://ssrn.com/abstract=1474746>>.
- Ésquilo. (1996). *Eumenides*. Trad. E. D. Morshead. Cambridge, Mass.: Internet Classics Archive. Disponível em : <[classics.mit.edu/Aeschylus/eumendides.html](http://classics.mit.edu/Aeschylus/eumendides.html)>
- Esquilo. (1986). *Choephoroi*. Trad. A. F. Garvie. Oxford, UK: Clarendon Press.

- Fama, E. F. (1965). "Random Walks in Stock Market Prices". *Financial Analysts Journal*, 21(5), 55-59. doi:10.2469/faj.v21.n5.55
- Fama, E. F. (1991). "Efficient Capital Markets: II". *The Journal of Finance*, 46(5), 1575. doi:10.2307/2328565
- Fama, E. F.; French, K. R. (1992). "The Cross-Section of Expected Stock Returns". *The Journal of Finance*, 47(2), 427. doi:10.2307/2329112
- Fama, E. F.; French, K. R. (2004). "The Capital Asset Pricing Model: Theory and Evidence". *Journal of Economic Perspectives*, 18 (3): 25-46. doi: 10.1257/0895330042162430
- Fama, E. F. (2014) "Two Pillars of Asset Pricing". *American Economic Review*, 104(6), 2014, 1467-1485. doi:10.1257/aer.104.6.1467
- Ferreira, I. S. M. (2012). "Lei determina atuação complementar entre BC e FGC". Artigo publicado no site *Consultor Jurídico (ConJur)*. Disponível em: <<http://www.conjur.com.br/2012-jun11/isaac-ferreira-lei-determina-atuacao-complementar-entre-bc-fgc>>.
- Ferreira, I. S. M. (2013). "Petição 1.467/2013-BCB/PGBC: Manifestação jurídica nas Arguições por Descumprimento de Preceito Fundamental (ADPF) nº 77 e nº 165, nas quais se discutem a constitucionalidade e a legitimidade dos planos econômicos de estabilização monetária editados nas décadas de oitenta e noventa do século XX". *Revista da Procuradoria-Geral do Banco Central*. 7(2): 273-292. Disponível em: <[bc.gov.br/pgbcb/201407/revista\\_pgbc\\_v7\\_n2\\_dez\\_2013.pdf](http://bc.gov.br/pgbcb/201407/revista_pgbc_v7_n2_dez_2013.pdf)>
- Fielding, E.; Lo, A. W.; Yang, J. H. (2011). "The National Transportation Safety Board: A Model for Systemic Risk Management". *Journal of Investment Management (JOIM)*, Abril/2011. Disponível em: <https://ssrn.com/abstract=1804613>
- Finlayson, J. G. (2005). *Habermas: a very short introduction*. Oxford: Oxford University Press.
- Finley, F. (2016) "A \$50 Million Hack Just Showed That The Dao Was All Too Human". In *Wired*, June 18, 2016. Disponível em: <<https://www.wired.com/2016/06/50-million-hack-just-showed-dao-human/>>
- Finnis, J. (1980). *Natural Law and Natural Rights*. Oxford, UK: Oxford University Press.
- Floridi, L. (2010). *Information: A very short introduction*. Oxford: Oxford University Press.
- Flyvbjerg, B.; Sunstein, C. R. (2016). "The Principle of the Malevolent Hiding Hand; or, the Planning Fallacy Writ Large". *Social Research*, vol. 83, no. 4, Winter, pp. 979-1004. Disponível em: <https://ssrn.com/abstract=2654423>



Ford, J.; Marriage, M. (2018). "The big flaw: auditing in crisis". In *Financial Times*, 1º de agosto de 2018. London. Disponível em: <<https://www.ft.com/content/29ccd60a-85c8-11e8-a29d-73e3d454535d>>. Republicado como "Escândalos mostram necessidade de rever regras de auditoria". Trad. Sabino Ahumada. In *Valor Econômico*, 7 de agosto de 2018. Disponível em: <[www.valor.com.br/empresas/5716671/escandalos-mostram-necessidade-de-rever-regras-de-auditoria](http://www.valor.com.br/empresas/5716671/escandalos-mostram-necessidade-de-rever-regras-de-auditoria)>

Forrester, J. W. (1998). "Designing the Future". Paper apresentado na Universidade de Sevilha em 15 de dezembro de 1998. Disponível em: <[http://static.clexchange.org/ftp/documents/whyk12sd/Y\\_1999-03DesigningTheFuture.pdf](http://static.clexchange.org/ftp/documents/whyk12sd/Y_1999-03DesigningTheFuture.pdf)>

Fox, J. (2011). *The myth of the rational market: A history of risk, reward, and delusion on Wall Street*. New York: Harper Business.

Fox, J. (2013). "What We've Learned from the Financial Crisis". *Harvard Business Review* online, edição de novembro de 2013. Disponível em: <<https://hbr.org/2013/11/what-weve-learned-from-the-financial-crisis>>

Fox, J. (2014). "What Alan Greenspan Has Learned Since 2008". *Harvard Business Review* online, 7 de janeiro de 2014. Disponível em: <<https://hbr.org/2014/01/what-alan-greenspan-has-learned-since-2008>>

Fox, J.; Lorsch, J. W. (2012). "What good are shareholders?". *Harvard Business Review*, Issue August-July. Disponível em : <[hbr.org/2012/07/what-goood-are-shareholders](http://hbr.org/2012/07/what-goood-are-shareholders)>

Freeman, S. R. (2003). *The Cambridge Companion to Rawls*. Cambridge, U.K.: Cambridge University Press.

Freeman, S. R. (2004). "Public Reason and Political Justifications". *Fordham L. Rev.* 72, pp. 2021-72. Disponível em: <<http://ir.lawnet.fordham.edu/flr/vol72/iss5/29>>

Freeman, S. (2016). "Original Position". *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), <<https://plato.stanford.edu/archives/win2016/entries/original-position/>>.

FSB - Financial Stability Board. (2018). *Stocktake of efforts to strengthen governance frameworks to mitigate misconduct risks*. Progress Report, publicado em 23 de maio de 2017. Disponível em: <<http://www.fsb.org/2017/05/stocktake-of-efforts-to-strengthen-governance-frameworks-to-mitigate-misconduct-risks/>>

- FSB - Financial Stability Board. (2018). *Strengthening Governance Frameworks to Mitigate Misconduct Risk: a toolkit for firms and supervisors*. Report ao G20, publicado em 20 de abril de 2018. Disponível em: <<http://www.fsb.org/wp-content/uploads/P200418.pdf>>
- Fukuyama, F. (2011). *The origins of political order*. New York: Farrar, Straus and Giroux.
- Gailmard, S.; Patty, J. W. (2013). *Learning while governing: Expertise and accountability in the executive branch*. Chicago: University of Chicago Press.
- Gaus, G. F. (2007). "Hayek on the evolution of society and mind". *The 3evaCambridge Companion to Hayek*, 232-258. doi:10.1017/ccol0521849772.013
- Gaus, G., & Thrasher, J. (2015). "Rational choice and the original position: The (many) models of Rawls and Harsanyi". In T. Hinton (Ed.), *The Original Position (Classic Philosophical Arguments)*, pp. 39-58. Cambridge: Cambridge University Press. doi:10.1017/CBO9781107375321.003
- Gauthier, D. (1983). "The Incomplete Egoist". *The Tanner Lectures on Human Values*, naStanford University, em 5 de outubro de 1983. Disponível em :<[http://tannerlectures.utah.edu/\\_documents/a-to-z/g/Gauthier84.pdf](http://tannerlectures.utah.edu/_documents/a-to-z/g/Gauthier84.pdf)>
- Gauthier, D. (1987). *Morals by Agreement*. Oxford: Oxford University Press, 1987.
- Gibbard, A. (2007). "Thinking How to Live with Each Other". *The Tanner Lectures on Human Values*, na Universidade da California, Berkeley, de 28 de fevereiro a 2 de março de 2006.
- Gilboa, I.; Schmeidler, D. (2003). *A theory of case-based decisions*. Cambridge: Cambridge University Press.
- Gintis, H. (2006). "Behavioral ethics meets natural justice". *Politics, Philosophy & Economics*, 5(1), 5-32. doi:10.1177/1470594x06060617
- Gintis, H. (2009). *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences*. Princeton, N.J.: Princeton University Press.
- Gintis, H. (2010). "Review 2. Rationality and its Discontents". *The Economic Journal*, 120(542), F162-F180. doi:10.1111/j.1468-0297.2009.02342.x
- Giocoli, N. (2011). "From Wald to Savage: Homo Economicus Becomes a Bayesian Statistician". SSRN Electronic Journal. doi:10.2139/ssrn.1944218
- Gneezy, U.; Rustichini, A. (2000). "A Fine is a Price". *The Journal of Legal Studies*, 29(1), 1-17. doi:10.1086/468061
- Graeber, D. (2015). *Debt: The first 5,000 years*. London: Melville.
- Grauwe, P. (2008). "Animal Spirits and Monetary Policy". CESifo Working Paper Series 2418, CESifo Group Munich.

- Gray, J. (2015). "The Friedrich Hayek I knew, and what he got right - and wrong". *Revista Newstatesman* online, 31 de julho de 2015. Disponível em: <<https://www.newstatesman.com/politics/2015/07/john-gray-friedrich-hayek-i-knew-and-what-he-got-right-and-wrong>>
- Greenberg, M. (2004). "How Facts Make Law". *Legal Theory*, 10(03): 157-198. doi:10.1017/s1352325204040212
- Greenspan, A. (1996) "The Challenge of Central Banking in a Democratic Society". Remarks by Alan Greenspan, Chairman, Board of Governors of the Federal Reserve System at the Annual Dinner and Francis Boyer Lecture of the American Enterprise Institute for Public Policy Research, em 5 de dezembro de 1996. Washington, D.C.: FED. Disponível em: <<https://www.federalreserve.gov/boarddocs/speeches/1996/19961205.htm>>
- Greenspan, A. (2013). *O mapa e o território: Risco, natureza humana e o futuro das previsões*. Trad. André Fontenelle e Otacílio Nunes. São Paulo: Penguin, Schwarcz.
- Greenwood, R.; Shleifer, A.; You, Y. (2017). "Bubbles for Fama". NBER Working Paper No. 23191. Aceito em *Journal of Financial Economics*. doi:10.3386/w23191
- Grossman, S. J.; Stiglitz, J. E. (1980). "On the Impossibility of Informationally Efficient Markets". *American Economic Review*, 70(3): 393–408. Disponível em: <<https://www.jstor.org/stable/1805228>>
- Guerrien, B. (2005). *Dictionnaire d'analyse économique: Microéconomie, macroéconomie, théorie des jeux, etc.* Paris: La Découverte.
- Habermas, J. (2012). *The lure of technocracy*. Trad. Ciaran Cronin. Cambridge: Polity Press.
- Hájek, A. (2017). "Pascal's Wager". *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition). Edward N. Zalta (ed.). Disponível em: <<https://plato.stanford.edu/archives/fall2017/entries/pascal-wager/>>.
- Haldane, A. G.; Madouros, V. (2013). "Complexity in Financial Regulation". *The Social Value of the Financial Sector*, 455-463. doi:10.1142/9789814520294\_0024
- Hand, D. J. (2008). *Statistics: A very short introduction*. Johanneshov: MTM.
- Hanson, S. G. ; Kashyap, A. K. ; Stein, J. C. (2011). "A Macroprudential Approach to Financial Regulation". *Journal of Economic Perspectives*, 1(25): 3-28. Também in *Chicago Booth Research Paper* No. 10-29. doi :10.2139/ssrn.1708173
- Hanson, R. (2013). "Shall We Vote on Values, But Bet on Beliefs?". *Journal of Political Philosophy*, 21(2), 151-178. doi:10.1111/jopp.12008

- Hansson, I. ; Skogh, G. (1987). "Moral Hazard and Safety Regulation". *The Geneva Papers on Risk and Insurance*, 12(43), 132-144. Disponível em: <<http://www.jstor.org/stable/41950219>>
- Hansson, S. O. (2014). "Risk". *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), Edward N. Zalta (ed.). Disponível em: <<https://plato.stanford.edu/archives/spr2014/entries/risk/>>
- Hansson, Sven Ove. *The Ethics of Risk: Ethical Analysis in an Uncertain World*. 2013.
- Harari, Y. N. (2015). *Sapiens: a brief history of humankind*. New York: Harper
- Hardie, W. F. (1968). "Aristotle and the Freewill Problem". *Philosophy* 43, no. 165, 274-278. doi:10.1017/s0031819100009244.
- Harsanyi, J. C. (1953) "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking". *Journal of Political Economy* 61(5): 434–35
- Harsanyi, J. C. (1975). "Can the Maximin Principle Serve as the Basis for Morality? A Critique of John Rawls's Theory". *American Political Science Review* 69: 594–606. Disponível em: <<https://www.jstor.org/stable/i306770>>
- Harsanyi, J. C. (1985). "Does Reason Tell Us What Moral Code to Follow and, Indeed, to Follow Any Moral Code at All?". *Ethics*, 96(1), 42-55. Disponível em: <<http://www.jstor.org/stable/2381321>>
- Harsanyi, J. C. (1995). "Games with Incomplete Information". *American Economic Review*, American Economic Association, vol. 85(3), pages 291-303, Junho/1995. Disponível em: <[nobelprize.org/nobel\\_prizes/economic-sciences/laureates/1994/harsanyi-lecture.pdf](http://nobelprize.org/nobel_prizes/economic-sciences/laureates/1994/harsanyi-lecture.pdf) >
- Hart, H. L. A. (1994). *The Concept of Law*. 2ª Ed. New York: Oxford University Press, 1994.
- Hausman, D. M.; McPherson, M. S. *Economic analysis, moral philosophy, and public policy*. New York: Cambridge University Press, 2006.
- Hayden, B. Y.; Platt, M. L. (2009). "The mean, the median, and the St. Petersburg paradox". *Judgm Decis Mak*. Jun 1; 4(4): 256–272.
- Hayek, F. A. (1973). *Law, legislation and liberty: A new statement of the liberal principles of justice and political economy*. Chicago, IL: University of Chicago Press.
- Hayek, F. A. (1996). "The Use of Knowledge in Society". *Knowledge Management and Organizational Design*, 7-15. doi:10.1016/b978-0-7506-9749-1.50005-3
- Heijden, M. V. D.; Beetsma, R.; Romp, W. (2017). "‘Whatever it takes’ and the role of Eurozone news". *Applied Economics Letters*, 25(16), 1166-1169. doi: 10.1080/13504851.2017.1403555

- Hershovitz, S. (2010). "The Role of Authority". *Philosophers Imprint* 11(7), maio/2010. Disponível em: <<http://ssrn.com/abstract=1615675>>, acesso em 24.7.2016.
- Herzog, L. (2017). "Markets". *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition). Edward N. Zalta (ed.). Disponível em: <<https://plato.stanford.edu/archives/fall2017/entries/markets/>>.
- Herzog, L. (2017). *Just financial markets? Finance in a just society*. Oxford: Oxford University Press.
- Himma, K. E. (2001a). "The Instantiation Thesis and Raz's Critique of Inclusive Positivism". in *Law and Philosophy*, Vol. 20, No. 1, Pp. 61-79, January 2001a. Disponível em: <http://ssrn.com/abstract=316869>
- Himma, K. E. (2001b). "Law's Claim to Legitimate Authority". In COLEMAN, Jules L. (ed.). *Hart's Postscript: Essays On The Postscript to the Concept of Law*, Oxford: Oxford University Press. Disponível em: <http://ssrn.com/abstract=928095>.
- Himma, K. E. (2002). "Inclusive Legal Positivism". Coleman, J. L.; Shapiro, S. (eds). *Oxford Handbook of Jurisprudence and Legal Philosophy*. Oxford: Oxford University Press, 2002. Disponível em: <http://ssrn.com/abstract=928098>
- Himma, K. E. (2007). "Revisiting Raz: Inclusive Legal Positivism and Our Concept of Authority". In *American Philosophical Association Newsletter on Philosophy and Law*, Vol. 6, No. 2, Spring 2007. Disponível em: <http://dx.doi.org/10.2139/ssrn.953392>
- Hirschman, A. O. (2004). *The rhetoric of reaction: perversity, futility, jeopardy*. Cambridge, MA: Belknap Press.
- Hirshleifer, J. (1971). "The Private and Social Value of Information and the Reward to Inventive Activity". *American Economic Review*, 61(4): 561-74. Disponível em: <[EconPapers.repec.org/RePEc:aea:aecrev:v:61:y:1971:i:4:p:561-74](http://EconPapers.repec.org/RePEc:aea:aecrev:v:61:y:1971:i:4:p:561-74)>
- Hobbes, T. (2002). *Os Elementos da Lei Natural e Política*. Trad.: Fernando Dias de Andrade. São Paulo: Ícone.
- Hofstadter, D. R. (2008). *I am a strange loop*. New York: Basic Books.
- Hofstadter, D. R.; Dennett, D. C. (1981). *The Mind's I: Fantasies and Reflections on Self and Soul*. New York: Basic Books.
- Hofstadter, D. R.; Sander, E.; Pratt, S. (2013). *Surfaces and essences: Analogy as the fuel and fire of thinking*. Rego Park: Gildan Media Corp.
- Holton, R. (2011). *Willing, wanting, waiting*. Oxford: Clarendon Press.

- Huemer, M. (2012). "In Praise of Passivity". *Studia Humana*, 1(2):12-28. Disponível em: <<http://studiahumana.com/pliki/wydania/In%20Praise%20of%20Passivity.pdf>>
- Huemer, M. (2013). *The Problem of Political Authority: An Examination of the Right to Coerce and the Duty to Obey*. Springer eBooks.
- Jackall, R. (1988). *Moral Mazes: the world of corporate managers*. New York: Oxford University Press.
- Jeffrey, R. (1965). *The Logic of Decision*. New York: McGraw-Hill.
- Jesus, S. S.; Souza, M. M. (2016). "Impacto do reconhecimento dos passivos contingentes na situação econômica das empresas brasileiras auditadas pelas *big four*". *Revista de Contabilidade da UFBA*, 10(2), 43. doi:10.9771/rc-ufba.v10i2.15287
- Kagan, S. (2018). "The paradox of methods". *Politics, Philosophy and Economics*, 17 (2):148-168. doi: 10.1177/1470594X17717737
- Kahneman, D. (2012). *Rápido e devagar: duas formas de pensar*. Trad. Cássio de Arantes Leite. Rio de Janeiro: Objetiva.
- Kahneman, D., & Tversky, A. (1979). "Prospect Theory: An Analysis of Decision under Risk". *Econometrica*, 47(2), 263. doi:10.2307/1914185
- Kavka, G. (1983) "The Toxin Puzzle". *Analysis*, v. 43, n° 1, pp. 33–36. Disponível em: <<http://analysis.oxfordjournals.org/content/43/1/33.extract>>, acesso em 24.7.2016.
- Kirilenko, A. A.; Lo, A. W.; (2013). "Moore's Law vs. Murphy's Law: Algorithmic Trading and Its Discontents". *Journal of Economic Perspectives*, 27(2). doi:10.2139/ssrn.2235963
- Kling, A. (2010) "The Financial Crisis: Moral Failure or Cognitive Failure?". *Harvard Journal of Law & Public Policy*, Volume 33, n° 2, Spring 2010, pp. 507-18.
- Knight, F. (1921). *Risk, Uncertainty, and Profit*. Boston, MA: Houghton-Mifflin.
- Knobe, J. (2003). "Intentional action in folk psychology: An experimental investigation". *Philosophical Psychology*, 16: 309–324. doi:10.1080/09515080307771.
- Knudsen, C. (1993). "Equilibrium, Perfect Rationality and the Problem of Self-Reference in Economics". *Rationality, Institutions and Economic Methodology*. doi:10.4324/9780203392805.ch6
- Korsgaard, C. (1986). "The right to lie: Kant on dealing with evil". *Philosophy and Public Affairs* 15(4): 325-349. Disponível em <<http://nrs.harvard.edu/urn-3:HUL.InstRepos:3200670>>.
- Kotlikoff, L. J. (2013). *Jimmy Stewart is dead: Ending the world's ongoing financial plague with limited purpose banking*. Hoboken, NJ: Wiley.

- Kronbauer, C. A.; Marquezan, L. H.; Barbosa, M. A.; Diehl, C. A. (2017). "Analysis of the effects of conservatism in accounting information after the 2011 change in the basic conceptual pronouncement". *Review of Business Management*, 19(65), 453-468. doi:10.7819/rbgn.v19i65.2742
- Krugman, P. R. (2009). *The return of depression economics and the crisis of 2008*. New York: W. Norton & Company.
- Kydland, F. (1977). "Equilibrium solutions in dynamic dominant-player models". *Journal of Economic Theory*, 15(2), 307-324. doi:10.1016/0022-0531(77)90105-3
- Kydland, F. E.; Prescott, E. C. (1977). "Rules Rather than Discretion: The Inconsistency of Optimal Plans". *Journal of Political Economy*, 85(3), 473-491. doi:10.1086/260580
- Lanchester, J. (2017). "After the Fall". *London Review of Books*, 40(3) : 3-8. Versão online. Disponível em : <<https://www.lrb.co.uk/v40/n13/john-lanchester/after-the-fall>>
- Lempert, Robert J.; Collins, Myles T. (August 2007). "Managing the risk of uncertain threshold responses: comparison of robust, optimum, and precautionary approaches". *Risk Analysis*. 27 (4): 1009–1026. doi:10.1111/j.1539-6924.2007.00940
- List, C.; Pettit, P. (2011) *Group agency: The possibility, design, and status of corporate agents*. Oxford: Oxford University Press.
- Lo, A. W. (2009). Regulatory reform in the wake of the financial crisis of 2007-2008. *Journal of Financial Economic Policy*, 1(1), 4-43. doi:10.1108/17576380910962376
- Lo, A. W. (2013). "Fear, Greed, and Financial Crises: A Cognitive Neurosciences Perspective". In Fouque, J. & Langsam, J. (Eds.). *Handbook on Systemic Risk*. Cambridge: Cambridge University Press, pp. 622-662. doi:10.1017/cbo9781139151184.032
- Lo, A. W. (2015). "The Gordon Gekko Effect: The Role of Culture in the Financial Industry". *SSRN Electronic Journal*. doi:10.2139/ssrn.2615625
- Lomasky, L. E. (2011). "Liberty after Lehman Brothers." *Social Philosophy and Policy*, 28(2), 135–65. Doi : 10.1017/S0265052510000245
- Lo, A. W.; Mueller, M. T. (2010). "Warning: Physics Envy May be Hazardous to Your Wealth!". *SSRN Electronic Journal*. doi:10.2139/ssrn.1563882
- Luban, D. (2004). "The Coiled Serpent of Argument: Reason, Authority, and Law in a Talmudic Tale. Epilogue: Law and Fable". *Chicago-Kent Law Review*, 79 (3): 1253-1288. Disponível em: <<https://scholarship.kentlaw.iit.edu/cklawreview/vol79/iss3/33>>.
- Luce, R.; H. Raiffa. (1957). *Games and Decisions*. New York: Wiley

- Łukasiewicz, Jan. (1970). *Selected Works*. Edited by Ludwik Borkowski. Amsterdam: North-Holland Pub. Co.
- Lundberg, E. (1999). “Saneamento do Sistema Financeiro – a experiência brasileira dos últimos 25 anos”. In Saddi, J. (org.). *Intervenção e Liquidação Extrajudicial no Sistema Financeiro Nacional – 25 anos da Lei 6.024/74*. São Paulo: Texto Novo. Disponível em: <<https://www.bcb.gov.br/ftp/saneamento.pdf>>.
- MacAskill, W.; Ord, T. (2018). “Why Maximize Expected Choice-Worthiness?”. *Nous*(forthcoming). doi:10.1111/nous.12264
- Macfarlane, J. (2003). “Future Contingents and Relative Truth”. *The Philosophical Quarterly*, 53(212):321-336. doi:10.1111/1467-9213.00315
- Mackay, D. M. (1960) “On The Logical Indeterminacy of a Free Choice”. *Mind* LXIX, n. 273, 31-40. doi:10.1093/mind/lxix.273.31.
- Mandelbrot, B.; Hudson, R. L. (2008). *The Misbehavior of Markets: A Fractal View of Financial Turbulence*. New York: Basic Books.
- Margin Call. (2011). Direção: J. C. Chandar. Produção: Joe Jenckes et al. Montréal, Alliance Vivafilm.
- Martin, R. (2017). "The St. Petersburg Paradox". *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition). Edward N. Zalta (ed.). Disponível em: <<https://plato.stanford.edu/archives/fall2017/entries/paradox-stpetersburg/>>.
- Marušić, B. (2013). “Promising against the Evidence”. *Ethics*, 123(2), 292-317. doi:10.1086/668704
- Marušić, B.; Schwenkler, J. (2017). “Intending Is Believing: A Defense of Strong Cognitivism”. Forthcoming in *Analytic Philosophy*. Vencedor do Marc Sanders Philosophy of Mind Prize 2017. Disponível em: <<http://www.marcsandersfoundation.org/wp-content/uploads/Intending-is-Believing-final.pdf>>.
- Maskin, E. “Decision-making under ignorance with implications for social choice”. *Theory and Decision* (1979) 11: 319. doi:10.1007/BF00126383
- McMillan, J. (pseudônimo). (2014). *The End of Banking: Money, Credit, and the Digital Revolution*. Zurich: Zero/One Economics.
- Mercier, H.; Sperber, D. (2011). “Why do humans reason? Arguments for an argumentative theory”. *Behavioral and Brain Sciences*, 34(02), 57-74. doi:10.1017/s0140525x10000968
- Michelon, C. (2006). *Being apart from reasons: The role of reasons in public and private moral decision-making*. Dordrecht: Springer.



- Migotti, M. (2015). "Paying a Price, Facing a Fine, Counting the Cost: The Differences that Make the Difference". *Ratio Juris*, 28(3), 372-391. doi:10.1111/raju.12088
- Miller, E. (1977). "Risk, Uncertainty, and Divergence of Opinion". *The Journal of Finance*, 32(4), 1151-1168. doi:10.2307/2326520.
- Milnor, J.W. (1954). "Games against nature". In: Thrall R.M.; Coombs C.H.; Davis R.L. (eds). *Decision Processes*. New York: Wiley, 49-60.
- Minsky, M. (1986). *The society of mind*. New York, NY: Simon and Schuster.
- Mises, L. von (1949). *Human action: A treatise on economics*. Yale University Press.
- Moehler, M. (2015). "The Rawls-Harsanyi Dispute: A Moral Point of View". *Pacific Philosophical Quarterly*, 99(1), 82-99. doi:10.1111/papq.12140
- Mongin, P.; Pivato, M. (2016). "Social Evaluation under Risk and Uncertainty". in Adler, M.; Fleurbaey, M. (eds). *Oxford Handbook of Well-Being and Public Policy*. New York: Oxford University Press, pp. 711-745. Disponível em: <[https://studies2.hec.fr/jahia/webdav/site/hec/shared/sites/mongin/acces\\_anonyme/MonginPivato%20Ch24%20OxfordHandbook.pdf](https://studies2.hec.fr/jahia/webdav/site/hec/shared/sites/mongin/acces_anonyme/MonginPivato%20Ch24%20OxfordHandbook.pdf)>
- Morgenstern, O. (1925). "Perfect Foresight and Economic Equilibrium". Trad. Frank W. Knight. In *Selected Economic Writings of Oskar Morgenstern*. New York: New York University Press, 1976, pp. 169-183.
- Morgenstern, O. (1972). "Descriptive, Predictive and Normative Theory". *Kyklos* 25, no. 4 (1972), 699-714. doi:10.1111/j.1467-6435.1972.tb01077.x.
- Morgenstern, O.; Neumann, J. von. (1953). *Theory of Games and Economic Behavior*. Princeton, Princeton University Press.
- Mouk, Y. B. (2015). *The Age of Responsibility: On the Role of Choice, Luck and Personal Responsibility in Contemporary Politics and Philosophy*. Tese de doutorado em Filosofia apresentada à Universidade de Harvard, Graduate School of Arts & Sciences, em dezembro de 2014. Disponível em: <<http://nrs.harvard.edu/urn-3:HUL.InstRepos:14226053>>
- Mueller, A.P. (2005) "The Misery of Central Banking". *The Free Market* 26, no. 5 (May 2005). Disponível em: <<https://mises.org/library/misery-central-banking>>
- Murray, N. ; Manrai, A. K. ; Manrai, L. A. (2017). "The financial services industry and society: The role of incentives/punishments, moral hazard, and conflicts of interests in the 2008 financial crisis". *Journal of Economics, Finance and Administrative Science*, 22(43), 168-190. Disponível em :<[http://www.scielo.org.pe/scielo.php?script=sci\\_arttext&pid=S2077-18862017000200004&lng=es&tlng=en](http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S2077-18862017000200004&lng=es&tlng=en)>.

- Myerson, R. B. (1979). "Incentive Compatibility and the Bargaining Problem". *Econometrica*, 47(1), 61. doi:10.2307/1912346
- Myerson, R. B. (2008). "Perspectives on Mechanism Design in Economic Theory". *American Economic Review*, 98(3), 586-603. Nobel Lecture Prize do dia 8 de dezembro de 2007. doi:10.1257/aer.98.3.586
- Myerson, Roger. (2009). "Learning from Schelling's Strategy of Conflict". *Journal of Economic Literature*, 47(4): 1109-25. Doi:10.1257/jel.47.4.1109.
- Nagel, T. (1979). *Mortal Questions*. New York: Cambridge University Press. doi: 10.1017/CBO9781107341050.011
- Nino, Carlos S. (1988). "Constructivismo epistemológico: entre Rawls y Habermas." *DOXA*, no. 05 (1988), 87-105. doi:10.14198/doxa1988.5.05.
- Nozick, R. (1990). *Anarchy, state, and utopia*. Oxford: Basil Blackwell.
- Nozick, Robert (1993). *The Nature of Rationality*. New jersey, Princeton University Press.
- Nussbaum, M. (2017). "Powerlessness and the Politics of Blame ». Jefferson Lecture in Humanities no John. F. Kennedy Center for the Performing Arts, em 1º de maio de 2017. Disponível em: < <https://www.law.uchicago.edu/news/martha-c-nussbaums-jefferson-lecture-powerlessness-and-politics-blame>>.
- Oaksford, M.; Chater, N. (2009). "Précis of Bayesian Rationality: The Probabilistic Approach to Human Reasoning". *Behavioral and Brain Sciences* 32, n. 01, 69-84. doi:10.1017/s0140525x09000284.
- Oliveira, F. (2003) "Democratização e Republicanização do Estado". *Teoria e Debate*, São Paulo, 16(54): 52-57. Disponível em: <<http://www.teoriaedebate.org.br/?q=materias/economia/democratizacao-e-republicanizacao-do-estado&page=0,3>>.
- Ord, T ; Hillerbrand, R.; Sandberg, A. (2008). "Probing the Improbable: Methodological Challenges for Risks with Low Probabilities and High Stakes". *Journal of Risk Research*. 13(2), out/2008. doi: 10.1080/13669870903126267.
- Ord, T. ; Wiblin, R. (2016). "Should we discount future health benefits when considering cost-effectiveness?". Disponível em:<<https://www.givingwhatwecan.org/sites/givingwhatwecan.org/files/attachments/discounting-health2.pdf>>

- Otsuka, M., & Voorhoeve, A. (2009). "Why It Matters That Some Are Worse Off Than Others: An Argument against the Priority View". *Philosophy & Public Affairs*, 37(2), 171-199. doi:10.1111/j.1088-4963.2009.01154.
- Pagin, P. (2008). "Informativeness and Moore's Paradox". *Analysis*, 68: 46-57. doi:10.1111/j.1467-8284.2007.00712.x
- Pagliari, F. (2016). "Social choice for one: On the rationality of intertemporal decisions". *Behavioural Processes*, 127, 97-108. doi:10.1016/j.beproc.2016.04.011
- Palazzo, G.; Krings, F.; Hoffrage, U. (2011). "Ethical Blindness". *Journal of Business Ethics*, 109(3), 323-338. doi:10.1007/s10551-011-1130-4
- Parfit, D. (1984). *Reasons and Persons*. Oxford: Clarendon Press.
- Parfit, D. (1997). "Equality and priority". *Ratio*, 10 (3): 202–221. doi:10.1111/1467-9329.00041
- Parfit, D. (2011). *On What Matters. Vol. .1* Oxford: Oxford University Press.
- Parker, D. C. (2011). *Closing a failed bank: Resolution practices and procedures*. Washington, DC: International Monetary Fund (IMF).
- Passas, N.; Goodwin, N. (2005). *It's Legal but It Ain't Right: Harmful Social Consequences of Legal Industries*. Michigan : University of Michigan Press. doi: 10.3998/mpub.11472
- Pedroso, Marcel. (2014). "Racionalidade limitada e uso de informações técnicas em modelos de análise de políticas públicas: proposições sobre a perspectiva integradora da Análise Multicritério de Decisão Espacial Construtivista". *RP3 - Revista de Pesquisa em Políticas Públicas*. 10.18829/rp3.v0i2.10156.
- Peres, R. A. (2016). "A Crise Moral: a responsabilidade de administradores de instituições financeiras e o argumento da inevitabilidade". *Revista da Procuradoria-Geral do Banco Central*. 10(2): 15-36, dez/2016, Brasília: BCB. Disponível em: <[http://www.bcb.gov.br/pgbcb/122016/revista\\_pgbc\\_v10\\_n2\\_dez\\_2016.pdf](http://www.bcb.gov.br/pgbcb/122016/revista_pgbc_v10_n2_dez_2016.pdf)>
- Peres, R. A. (2017). "What the Tortoise will say to Achilles – or 'taking the traditional interpretation of the sea battle argument seriously'". *Unisinos Journal of Philosophy*, 18(1):63-68, jan/apr 2017. doi: 10.4013/fsu.2017.181.08
- Peres, R. A. (2018). "Autoridade e normatividade num 'mundo amplo': considerações sobre decisões sociais em contextos de grande incerteza". Santos, F. R. L.; Ruivo, J. L.; Silva Filho, W. J. (orgs). *Anais da V Conferência da Sociedade Brasileira de Filosofia Analítica*. Pelotas, RS: Dissertatio (UFPel), pp. 247-50. Disponível em: <<https://wp.ufpel.edu.br/nepfil/files/2018/09/SBFA-VFinal.pdf>>.

- Peters, O. (2011). "The time resolution of the St Petersburg paradox". *Philosophical Transactions of Royal Society A*, 369(1956): 4913-4931. doi:10.1098/rsta.2011.0065
- Peters, O.; Mauboussin, M. (2012). "Shaking the foundation: Revisiting Basic Assumptions about Risk, Reward, and Optimal Portfolios". Entrevista com Ole Peters, em 23 de março de 2012. Legg Mason Capital Management, pp. 1-14. Disponível em: <<http://contenta.mkt1710.com/lp/26966/115068/MauboussinOnStrategyShakingtheFoundationMIPX014449.pdf>>.
- Petrosino, A.; Turpin-Petrosino, C.; Hollis-Peel, M. E.; Lavenberg, J. G. (2013). "'Scared Straight' and other juvenile awareness programs for preventing juvenile delinquency". Cochrane Database of Systematic Reviews. doi:10.1002/14651858.cd002796.pub2
- Pettit, P. (2012). *On the people's terms: A republican theory and model of democracy*. Cambridge: Cambridge University Press.
- Picavet, E. (1996) *Choix Rationnel et Vie Publique: Pensée formelle et raison pratique*. Paris: Presses Universitaires de France.
- Picavet, Emmanuel; Lafaye, Caroline Guibet. (2012). "La précaution, l'éthique et la structure de l'action", *Revue de métaphysique et de morale* 2012/4 (N° 76), p. 593-609. Doi 10.3917/rmm.124.0593
- Pistor, K. (2017). "Moneys' legal hierarchy". In Herzog, L. (ed.). *Just financial markets?: Finance in a just society*. Oxford: Oxford University Press. doi:10.1093/oso/9780198755661.003.0008
- Pitkin, H. F. (1972). *The concept of representation*. Berkeley: University of California Press.
- Platão. (1965). *República*. Trad. J. Ginsburg. São Paulo: Difusão Europeia do Livro.
- Posen, A. S. (2006). "Why Central Banks Should Not Burst Bubbles". *International Finance*, 9(1), 109-124. doi:10.1111/j.1468-2362.2006.00028.x
- Posner, R. A. (2006). *Catastrophe: Risk and Response*. New York: Oxford University Press.
- Posner, R. A. (2010) *The Crisis of Capitalist Democracy*. Cambridge: Harvard University Press.
- Postema, G. J. (2008). "Salience Reasoning". *Topoi*, 27(1-2), 2008, 41-55. doi:10.1007/s11245-008-9031-6
- Prasad, R. (1996). *The Bhagavad-Gītā - The Song of God: with Introduction, Original Sanskrit Text and Roman Transliteration*. Delhi: Motilal Banarsidass Publishers. Disponível em <[http://www.gita-society.com/bhagavadgita\\_english.html](http://www.gita-society.com/bhagavadgita_english.html)>, acesso em 18 de agosto de 2016.

- Press, W. H.; Dyson, F. J. (2012). "Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent". *Proceedings of the National Academy of Sciences*, 109(26), 10409-10413. doi:10.1073/pnas.1206569109.
- Priel, D. (2010). "Jurisprudence and Psychology". In: Mar, M.D. (ed). (2011). *New Waves in Philosophy of Law. New Waves in Philosophy*. London: Palgrave Macmillan. doi: 10.1057/9780230316645\_4
- Prior, A. N. (1967). *Past, Present and Future*. Oxford: Clarendon.
- Pritchard, D. (2015). "Risk". *Metaphilosophy*, 46(3): 436-461. doi:10.1111/meta.12142
- Putnam, H. (1980). "Models and reality". *The Journal of Symbolic Logic*, 45(03), 464-482. doi:10.2307/2273415
- Quiggin, J. (2018). *Economics in two lessons* (forthcoming). Manuscrito disponível em: <<https://johnquiggin.com/category/books-and-culture/economics-in-two-lessons/>>.
- Quiggin, J. (2010) *Zombie economics: How dead ideas still walk among us*. Princeton: Princeton University Press.
- Rajan, R. (2005). "Has Financial Development Made the World Riskier?". National Bureau of Economic Research, Working Paper No. 11728. doi:10.3386/w11728
- Rajan, R. G. (2006). "Has Finance Made the World Riskier?". *European Financial Management*, 12(4), 499-533. doi:10.1111/j.1468-036x.2006.00330.x
- Rawls, J. (1955). "Two Concepts of Rules". *The Philosophical Review* 64(1): 3. doi:10.2307/2182230.
- Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Belknap Press of Harvard University Press.
- Rawls, J. (2005) *Political Liberalism*. 3ª Ed. New York: Columbia University Press.
- Raz, J. (1979) *The Authority of Law: Essays on law and morality*. Oxford: Clarendon Press.
- Raz, J. (1986). *The Morality of Freedom*. Oxford: Clarendon Press, 1986.
- Raz, J. (1996). *Ethics in the public domain: Essays in the morality of law and politics*. Oxford: Clarendon Press.
- Raz, J. (2002). *Practical Reason and Norms*. 3ª Ed. New York: Oxford University Press.
- Raz, J. (2009). *Between Authority and Interpretation: on the theory of law and practical reason*. Oxford: Oxford University Press.
- Recamonde-Mendoza, M.; Bazzan, A. L. (2016). "Social choice in distributed classification tasks: Dealing with vertically partitioned data". *Information Sciences*, 332, 56-71. doi:10.1016/j.ins.2015.11.001

- Regan, D. (1985). "On Preferences and Promises: A Response to Harsanyi". *Ethics*, 96(1), 56-67. doi:10.1086/292718
- Reinhart, C. M.; Rogoff, K. S. (2011). *This time is different: Eight centuries of financial folly*. Princeton: Princeton University Press.
- Robinson, M.; De Souza, J. G.; Maezumi, S. Y.; Cárdenas, M.; et al. (2018). "Uncoupling human and climate drivers of late Holocene vegetation change in southern Brazil". *Scientific Reports*, 8(1). doi:10.1038/s41598-018-24429-5
- Rodrik, D. (1999). "Where Did All The Growth Go? External Shocks, Social Conflict, and Growth Collapses". *Journal of Economic Growth*, 4: 358-412. doi:10.3386/w6350
- Rodrik, D. (2012). *The globalization paradox: Democracy and the future of the world economy*. New York: W.W. Norton & Company.
- Roelfs, D. J.; Shor, E.; Davidson, K. W.; Schwartz, J. E. (2011). "Losing Life and Livelihood: A Systematic Review and Meta-Analysis of Unemployment and All-Cause Mortality". *Social Science & Medicine* (1982), 72(6), 840–854. doi:10.1016/j.socscimed.2011.01.005
- Roodman, D. (2017). *The impacts of incarceration on crime*. San Francisco : Open Philanthropy. Disponível em : < [https://www.openphilanthropy.org/files/Focus\\_Areas/Criminal\\_Justice\\_Reform/The\\_impacts\\_of\\_incarceration\\_on\\_crime\\_10.pdf](https://www.openphilanthropy.org/files/Focus_Areas/Criminal_Justice_Reform/The_impacts_of_incarceration_on_crime_10.pdf)>
- Roth, A. E. (2007). "Repugnance as a Constraint on Markets". *Journal of Economic Perspectives*, 21(3), 37-58. doi:10.1257/jep.21.3.37
- Roubini, N. (2006). "Why Central Banks Should Burst Bubbles". *International Finance*, 9(1), 87-107. doi:10.1111/j.1468-2362.2006.00032.x
- Russell, S. J.; Norvig, P. (2010). *Artificial intelligence a modern approach*. New Jersey: Prentice Hall.
- Sandri, S. (2009). *Reflexivity in Economics: an experimental examination on the self-referentiality in economic theories*. Heidelberg: Physica-Verlag.
- Sandel, M. J. (2012). *What money can't buy: the moral limits of markets*. New York: Farrar, Straus & Giroux.
- Sandel, M. (2012). *O que o dinheiro não compra: os limites morais do mercado*. Trad. Clóvis Marques. Rio de Janeiro: Civilização Brasileira.
- Sandel, M. (2013). "The Moral Economy of Speculation: Gambling, Finance, and the Common Good". *Tanner Lecture on Human Values*, Universidade de Utah, 27 de fevereiro de 2013. Disponível em: < <https://tannerlectures.utah.edu/Sandel%20Lecture.pdf>>

- Savage, L. (1951). *The Foundations of Statistics*. New York: Wiley.
- Scanlon Jr.; T. M. (1988). “The Significance of Choice”. Tanner Lecture on Human Values, em Brasenose College, Oxford University, maio de 1986. Disponível em: <[https://tannerlectures.utah.edu/\\_documents/a-to-z/s/scanlon88.pdf](https://tannerlectures.utah.edu/_documents/a-to-z/s/scanlon88.pdf)>.
- Schefczyk, M. (2012). “The Financial Crisis, the Exemption View and the Problem of the Harmless Torturer”. *Philosophy of Management* 11 (1):25-38.
- Schefczyk, M. (2016) “Financial Markets: Applying Argument Analysis to the Stabilisation Task”. In: Hansson S.; Hirsch Hadorn G. (eds). *The Argumentative Turn in Policy Analysis*. Logic, Argumentation & Reasoning (Interdisciplinary Perspectives from the Humanities and Social Sciences), vol 10. Springer, Cham.
- Schelling, T. C. (1983). “Ethics, Law, and the Exercise of Self-Command”. Tanner Lecture on Human Values, na University of Michigan, em 19 e 21 de março de 1982. Disponível em: <[https://tannerlectures.utah.edu/\\_documents/a-to-z/s/schelling83.pdf](https://tannerlectures.utah.edu/_documents/a-to-z/s/schelling83.pdf)>
- Schelling, T. C. (2015). *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.
- Schmidhuber, J. (2009). "Ultimate Cognition à la Gödel". *Cognitive Computation*, 1, 177-193. Doi: 10.1007/s12559-009-9014-y.
- Schmidt, A. R. (2009). *Contradição e Determinismo: um estudo sobre o problema dos futuros contingentes em Tomás de Aquino*. Dissertação apresentada como requisito parcial para a obtenção do título de Mestre em Filosofia ao Programa de Pós-Graduação em Filosofia da Universidade Federal do Rio Grande do Sul. Porto Alegre: UFRGS, 2009.
- Schoenfield, M. (2015). “The Accuracy and Rationality of Imprecise Credences”. *Noûs*, 51(4), 667-685. doi:10.1111/nous.12105
- Schwab, G. G. (2014). “SEC changes course: some companies must now admit wrongdoing”. *Journal of Investment Compliance*, 15(1), 33-34. doi:10.1108/joic-01-2014-0005
- Scott, A. D.; Scott, M. (1999). “The paradox of the question”. *Analysis*, 59(4), 331-334. doi:10.1093/analys/59.4.331
- Sen, A. K. (1977). “Rational Fools: A Critique of the Behavioral Foundations of Economic Theory”. *Philosophy & Public Affairs*, 6(4): 317-44. Disponível em: <<http://www.jstor.org/stable/2264946>>
- Sen, A. K. (1979). “Equality of What?”. Tanner Lecture on Human Values, na Stanford University, 22 de maio de 1979. Disponível em: <[http://www.ophi.org.uk/wp-content/uploads/Sen-1979\\_Equality-of-What.pdf](http://www.ophi.org.uk/wp-content/uploads/Sen-1979_Equality-of-What.pdf)>, pp. 197-220.

- Sen, A. K. (1986). "Information and invariance in normative choice". In W. Heller, R. Starr, & D. Starrett (Eds.), *Essays in Honor of Kenneth J. Arrow*, pp. 29-56. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511983535.006
- Sen, A. K. (1999). "The Possibility of Social Choice". *American Economic Review*, 89(3), 349-378. doi:10.1257/aer.89.3.349
- Sen, A. K. (2009) *The Idea of Justice*. Cambridge, MA: Belknap Press, 2009.
- Sewell, M. (2012). "The Efficient Market Hypothesis: Empirical Evidence". *International Journal of Statistics and Probability*, Vol. 1, n° 2, 165-78.
- Shakespeare, W. *MacBeth*. Disponível em: <<http://shakespeare.mit.edu/macbeth/>>, acesso em 24.7.2016.
- Shapiro, S. J. (2000). "Authority". Stanford/Yale Jr. Faculty Forum Research Paper 00-05;
- Shapiro, S. J. (2011). *Legality*. Cambridge, MA: Harvard University Press.
- Cardozo Law School, Public Law Research Paper No. 24, 2000. doi:10.2139/ssrn.233830.
- Shrader-Frechette, K. *Tainted: How philosophy of science can expose bad science*. New York: OUP, 2014.
- Simon, H. A. (1956). "Rational choice and the structure of the environment". *Psychological Review*, 63(2), 129-138. doi:10.1037/h0042769
- Simon, H. A. (1976) "From substantive to procedural rationality". In: Latsis, S. J.; *Method and Appraisal in Economics*, Cambridge: Cambridge University Press, pp. 129-148.
- Skyrms, B. (1987) "Dynamic Coherence". In: MacNeill I.B.; Umphrey G.J.; Safiul Haq M.; Harper W.L.; Provost S.B. (eds). *Advances in the Statistical Sciences: Foundations of Statistical Inference*. The University of Western Ontario Series in Philosophy of Science (A Series of Books in Philosophy of Science, Methodology, Epistemology, Logic, History of Science and Related Fields), vol 35. Springer, Dordrecht.
- Skyrms, B. (1999). "Précis of Evolution of the Social Contract". *Philosophy and Phenomenological Research*, Vol. 59, No. 1 (Mar.; 1999), pp. 217-220, Disponível em: <<http://www.jstor.org/stable/2653471>>
- Skyrms, B. (2007). "Evolution and the Social Contract". *The Tanner Lecture on Human Values*, na University of Michigan, 2 de Novembro de 2007. Disponível em: <[https://tannerlectures.utah.edu/\\_documents/a-to-z/s/Skyrms\\_07.pdf](https://tannerlectures.utah.edu/_documents/a-to-z/s/Skyrms_07.pdf)>
- Skyrms, B. (2014). *Evolution of the Social Contract*. 2ª Ed. Cambridge University Press.
- Smith, J. E.; Winkler, R. L. (2006). "The Optimizer's Curse: Skepticism and Postdecision Surprise in Decision Analysis". *Management Science*. 52. 311-322. 10.1287/mnsc.1050.0451.



- Soares, N.; Fallenstein, B. (2015a). "Questions of Reasoning under Logical Uncertainty". Machine Intelligence Research Institute (MIRI), technical report 2015-1. Disponível em: <[intelligence.org/files/QuestionsLogicalUncertainty.pdf](http://intelligence.org/files/QuestionsLogicalUncertainty.pdf)>
- Soares, N.; Fallenstein, B. (2015a). "Vingean Reflection: Reliable Reasoning for Self-Improving Agents." Machine Intelligence Research Institute (MIRI), technical report 2015-2. Disponível em: <<http://intelligence.org/files/VingeanReflection.pdf>>
- Soares, N.; Levinstein, B. (2017). "Cheating Death in Damascus". Paper presented at the 14th Annual Formal Epistemology (FEW), University of Washington, Seattle, USA. May 26-28, 2017. Disponível em: <<https://intelligence.org/files/DeathInDamascus.pdf>>
- Solomon, S. M. D.; Morrison, A. D.; Wilhelm JR.; W. J. (2012) "The SEC v. Goldman Sachs: Reputation, Trust, and Fiduciary Duties in Investment Banking". *The Journal of Corporation Law*, 37(3), 530-553. Disponível em <<http://scholarship.law.berkeley.edu/facpubs/2307>>
- Sorensen, R. A. (1988). *Blindspots*. Oxford England: Clarendon Press.
- Sorensen, R. A. (1995). "Unknowable Obligations". *Utilitas*, 7(02), 247-271. doi:10.1017/s0953820800002077
- Sorensen, R. A. (1999). "Infinite 'Backward' Induction Arguments". *Pacific Philosophical Quarterly*, 80(3), 278-283. doi:10.1111/1468-0114.00084
- Sorensen, R. A. (2001). *Vagueness and Contradiction*. Oxford, UK: Clarendon Press.
- Sorensen, R. A. (2003) *A Brief History of the Paradox: Philosophy and the Labyrinths of the Mind*. Oxford: Oxford University Press.
- Sorensen, R. A. (2006) "Sharp Edges from Hedges: Fatalism, Vagueness and Epistemic Possibility". *Philosophical Studies* 131, no. 3, 607-626. doi:10.1007/s11098-005-7701-4.
- Sorensen, R. A. (2006). "Originless Sin: Rational Dilemmas for Satisficers". *The Philosophical Quarterly*, 56(223), 213-223. doi: 10.1111/j.1467-9213.2006.00438.x
- Sorensen, R. A. (2014). "Epistemic Paradoxes". *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), Edward N. Zalta (ed.). Disponível em: <<https://plato.stanford.edu/archives/spr2014/entries/epistemic-paradoxes/>>.
- Soros G. (2000). *Open Society: reforming global capitalism*. New York: Public Affairs, 2000.
- Soros, G. (2008). *The New Paradigm for Financial Markets: The Credit Crisis of 2008 and What It Means*. New York: Public Affairs.
- Souza, A. C. G. (2009). *Ensaio sobre regulação monetária financeira*. Tese apresentada em 10 de fevereiro de 2009 ao Programa de Pós-Graduação em Desenvolvimento Econômico,

- Setor de Ciências Sociais Aplicadas, da Universidade Federal do Paraná (UFPR). Orientador: Prof. Dr. J. G. P. Meirelles. Disponível em: <<http://hdl.handle.net/1884/24012>>
- Stiglitz, J. E. (2010a). *The Stiglitz Report: Reforming the International Monetary and Financial Systems in the Wake of the Global Crisis*. New York, N.Y.: The New Press, 2010a. Disponível em: <[http://www.un.org/ga/econcrisissummit/docs/FinalReport\\_CoE.pdf](http://www.un.org/ga/econcrisissummit/docs/FinalReport_CoE.pdf)>, acesso em 18 de agosto de 2016.
- Stiglitz, J. E. (2010b). *Freefall: America, free markets, and the sinking of the world economy*. New York: W.W. Norton & Co, 2010b.
- Stiglitz, J. E. (2012). *Le prix de l'inégalité*. Trad. Françoise Chemla & Paul Chemla. Paris: Le Grand Livre du Mois.
- Straffin, P. D. (2010). *Game theory and strategy*. Washington, DC: Mathematical Association of America.
- Sugden, R. (2001). "Ken Binmore's Evolutionary Social Theory". *The Economic Journal*, 111(469), 213-243. doi:10.1111/1468-0297.00604
- Sunstein, C. R. (1994). "Political Conflict and Legal Agreement". *The Tanner Lectures on Human Values*, na Universidade de Harvard de 29 de novembro a 1º de dezembro de 1994. Disponível em: <<http://dx.doi.org/10.2139/ssrn.2544359>>
- Sunstein, C. R. (2006). *Infotopia: how many minds produce knowledge*. New York: Oxford University Press.
- Sunstein, C. R. (2013). "The Value of a Statistical Life: Some Clarifications and Puzzles". *SSRN Electronic Journal*. doi:10.2139/ssrn.2278402
- Sunstein, C. R. (2014). "Cost-Benefit Analysis and the Knowledge Problem". *Regulatory Policy Program Working Paper RPP-2015-03*. Cambridge, MA: Mossavar-Rahmani Center for Business and Government, Harvard Kennedy School, Harvard University. Disponível em: <<http://dx.doi.org/10.2139/ssrn.2508965>>
- Tadros, V. (2013). *The ends of harm: The moral foundations of criminal law*. New York : Oxford University Press.
- Tadros, V. (2015). "Response to Blumenson." *Law, Ethics and Philosophy* [online], (3), 56-73. Disponível em: [www.raco.cat/index.php/LEAP/article/view/313504](http://www.raco.cat/index.php/LEAP/article/view/313504).
- Talwalkar, P. (2014). *The joy of game theory: An introduction to strategic thinking*. Chicago: P. Talwalkar (e-book).
- Taylor, M. (1969). "Critique and comment: Proof of a theorem on majority rule". *Behavioral Science*, 14(3), 228-231. doi:10.1002/bs.3830140307

- Tetlock, P.; Gardner, D. (2016). *Superforecasting: art and science of prediction*. New York: Crown Publishers.
- Thomas, K. A.; DeScioli, P.; Haque, O.S.; Pinker, S. (2014). “The psychology of coordination and common knowledge”. *Journal of Pers Soc Psychol*, Oct/2014, 107(4):657-76. doi: 10.1037/a0037037
- Toh, K. (2008). “An Argument Against The Social Fact Thesis (And Some Additional Preliminary Steps Towards A New Conception Of Legal Positivism)”. *Law And Philosophy*, set/2008, 27(5): 445-504. Disponível em: <http://ssrn.com/abstract=1650749>.
- Tsetsos, K.; Moran, R.; Moreland, J.; Chater, N.; Usher, M.; Summerfield, C. (2016). “Economic irrationality is optimal during noisy decision making”. *Proceedings of the National Academy of Sciences of the United States of America*, 113(11), 3102–3107. doi:10.1073/pnas.1519157113
- Tucídides. (1874). *The Peloponnesian War*. Trad. Richard Crawley. Disponível em: <http://classics.mit.edu/thucydides/pelopwar.5.fifth.html>
- Tucker, P. (2018). *Unelected Power: The Quest for Legitimacy in Central Banking and the Regulatory State*. Princeton: Princeton University Press.
- Tversky, A.; Kahneman, D. (1982). “Judgment under uncertainty: Heuristics and biases”. *Science* 185(4157):38-55, Jan/1982. doi:10.1017/cbo9780511809477.002
- United States. (2011). *Wall Street and the Financial Crisis: Anatomy of a Financial Collapse*. Majority and Minority Staff Report. Final Report of the National Commission on the Causes of the Financial and Economic Crisis in the United States. Washington, DC: Financial Crisis Inquiry Commission. Disponível em: <https://www.gpo.gov/fdsys/pkg/GPO-FCIC/pdf/GPO-FCIC.pdf>
- Vineberg, S. (2016). "Dutch Book Arguments". *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition), Edward N. Zalta (ed.). Disponível em: <https://plato.stanford.edu/archives/spr2016/entries/dutch-book/>.
- Vivalt, E. (2017b). “How Much Can Impact Evaluations Inform Policy Decisions?”.(unpublished) Disponível em: <http://evavivalt.com/wp-content/uploads/How-Much-Can-We-Generalize.pdf>
- Wald, A. (1945). “Statistical decision functions which minimize the maximum risk”. *The Annals of Mathematics*, 46(2), 265-280. Disponível em: <https://www.jstor.org/stable/1969022>

- Waldron, J. (2009). “Dignity, Rank, and Rights”. The 2009 Tanner Lectures at UC Berkeley (September 1, 2009). NYU School of Law, Public Law Research Paper No. 09-50. doi:10.2139/ssrn.1461220
- Walter, C. (2012). “Éthique et finance: le tournant performatif”, *Transversalités* 2012/4 (N° 124), pp. 29-42. doi:10.3917/trans.124.0029
- Walter, C. (2015). “Les deux quantifications de la théorie financière. Contribution à une histoire critique des modèles financiers”. FMSH-WP-2015-89. 2015. <halshs-01118147>
- Weissmann, J. (2016). “The GOP Platform Calls Up Big Banks. Seriously!”. Artigo publicado em 18 de julho de 2016. Disponível em: <[http://www.slate.com/blogs/moneybox/2016/07/18/trump\\_campaign\\_says\\_gop\\_platform\\_will\\_call\\_for\\_breaking\\_up\\_big\\_banks.html](http://www.slate.com/blogs/moneybox/2016/07/18/trump_campaign_says_gop_platform_will_call_for_breaking_up_big_banks.html)>
- Williams, B. (1993). *Shame and Necessity*. Berkeley: University of California Press.
- Williams, B. (2005). *Moral: uma introdução à ética*. Trad. Remmo Manarino Filho. São Paulo: Martins Fontes.
- Wittgenstein, L. (1953). *Philosophical Investigations*. Trad. G.E.M. Anscombe, Oxford: Blackwell.
- Wittgenstein, L. (1977). *On Certainty*. Trad. e ed. G.E.M. Anscombe e G.H. von Wright. Oxford: Blackwell.
- Yudkowsky, E. (2008). “Cognitive Biases Potentially Affecting Judgment of Global Risks.” In Bostrom, N., & Čirković, M. M. *Global catastrophic risks*. New York: Oxford University Press, 91–119. Disponível em: <[intelligence.org/files/CognitiveBiases.pdf](http://intelligence.org/files/CognitiveBiases.pdf)>
- Yudkowsky, E. (2015). *Rationality: from AI to Zombies*. Berkeley: Machine Intelligence Research Institute (MIRI). Disponível em: <[intelligence.org/rationality-ai-zombies/](http://intelligence.org/rationality-ai-zombies/)>
- Yudkowsky, E.; Soares, N. (2017). “Functional Decision Theory: A New Theory of Instrumental Rationality”. arXiv:1710.05060, versão: [cs.AI].
- Zagzebski, L. T. (2012). *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief*. Oxford University Press.
- Zawidzki, T. W. (2015). “As close to the definitive Dennett as we're going to get”. *Studies in History and Philosophy of Science, Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 50, 98-102. doi:10.1016/j.shpsc.2015.01.012
- Zhang, R.; Brennan, T. J.; Lo, A. W. (2014). “The origin of risk aversion”. *Proceedings of the National Academy of Sciences*, 111(50), 17777-17782. doi:10.1073/pnas.1406755111