

Raquel Bavaresco Cipriani Xavier

**UM ESTUDO SOBRE OS ARGUMENTOS EM FAVOR DA  
ESTABILIDADE EM *UMA TEORIA DA JUSTIÇA***

Dissertação submetida ao Programa de Pós-graduação em Filosofia da Universidade Federal de Santa Catarina como requisito para a obtenção de grau de mestre em Filosofia, sob a orientação do Prof. Dr. Denilson Luís Werle.

Florianópolis  
2017

Ficha de identificação da obra elaborada pelo autor,  
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

CIPRIANI XAVIER, Raquel Bavaresco

Um estudo sobre os argumentos em favor da estabilidade  
em Uma Teoria da Justiça / Raquel Bavaresco CIPRIANI  
XAVIER ; orientador, Denilson Luis Werle - Florianópolis,  
SC, 2017.

128 p.

Dissertação (mestrado) - Universidade Federal de Santa  
Catarina, Centro de Ciências Físicas e Matemáticas.  
Programa de Pós-Graduação em Filosofia.

Inclui referências

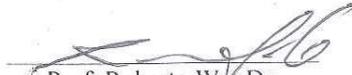
1. Filosofia. 2. Justiça como Equidade. 3. Estabilidade.  
4. Congruência entre Correto e Bem. 5. Psicologia Moral .  
I. Werle, Denilson Luis. II. Universidade Federal de Santa  
Catarina. Programa de Pós-Graduação em Filosofia. III. Título.

Raquel Bavaresco Cipriani Xavier

**“UM ESTUDO SOBRE OS ARGUMENTOS EM FAVOR  
DA ESTABILIDADE EM *UMA TEORIA DA JUSTIÇA*”**

Esta dissertação foi julgada adequada para obtenção do Título de “Mestre em Filosofia”, e aprovada em sua forma final pelo Programa de Pós-Graduação em Filosofia.

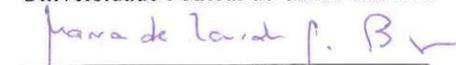
Florianópolis, 08 de março de 2017.

  
Prof. Roberto Wyl, Dr.  
Coordenador do Curso

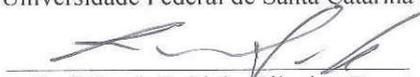
**Banca Examinadora:**

  
Prof. Denilson Luis Werle, Dr.  
Orientador

Universidade Federal de Santa Catarina

  
Prof.<sup>a</sup> Maria de Lourdes Alves Borges, Dr.<sup>a</sup>  
Universidade Federal de Santa Catarina

  
Prof. Alessandro Pinzani, Dr.  
Universidade Federal de Santa Catarina

  
Prof. Denis Coitinho Silveira, Dr.  
Universidade do Vale do Rio dos Sinos



## **AGRADECIMENTOS**

Apesar do esforço individual e dos momentos de pesquisa e redação solitária que a produção de um trabalho como este envolve, acredito que o mérito não é individual, mas sim social e institucionalmente construído e que, do mesmo modo, o conhecimento se constrói em um processo coletivo. Por este motivo, quero agradecer e manifestar meu reconhecimento ao Estado brasileiro, que por meio da universidade pública e órgãos de fomento incentiva o ensino, a pesquisa e a extensão.

Agradeço também ao Programa de Pós-Graduação em Filosofia da Universidade Federal de Santa Catarina, aos funcionários e a todos meus professores, em especial a meu professor orientador Dr. Denilson Luís Werle por todo apoio, incentivo e liberdade durante o processo de pesquisa.

Agradeço aos professores Dr. Denis Coitinho Silveira, Dra. Maria de Lourdes Alves Borges e Dr. Alessandro Pinzani pela participação na banca examinadora.

Agradeço às amigas e amigos que estiveram ao meu lado nesta trajetória, em especial a Ilze Zirbel, Jorge Sell, Nunzio Ali, Eduardo de Borba, Marina dos Santos e Nikolay Steffens Martins.

A meus pais, Jaime e Ilda, a Dona Gilda e ao meu marido Gabriel por todo amor, carinho e suporte afetivo.



*Se uma sociedade razoavelmente justa, que subordina o poder a seus objetivos, não é possível e se as pessoas são em grande medida amorais, quando não incorrigivelmente cínicas e autocentradas, talvez devêssemos indagar, com Kant, se vale a pena para os seres humanos viverem na Terra.*

John Rawls, Introdução à edição de 1996 do  
Liberalismo Político, p. LXIX



## RESUMO

O objetivo dessa dissertação é apresentar um estudo sobre os argumentos em favor da estabilidade desenvolvidos na Parte III do livro *Uma Teoria da Justiça* de John Rawls, com foco especial no argumento da congruência entre correto e bem compreendido a partir do ponto de vista da estrutura da teoria moral. Nossa estratégia para alcançar esse objetivo consiste em evidenciar o modo como a teoria da justiça como equidade define e articula as noções morais de correto e de bem em sua estrutura moral a partir de uma deontologia não-rigorista. O passo seguinte é apontar os âmbitos da especificação do correto (normativo) e da descrição do bem (descritivo) na estrutura da teoria e mostrar a implicação destes dois âmbitos nas duas estratégias de justificação dos princípios – posição original e equilíbrio reflexivo. Ao fazer isso, evidenciamos que ambas as estratégias possuem são estruturadas em dois estágios e conseguimos localizar o argumento da estabilidade como fazendo parte da segunda etapa argumentativa da posição original. Uma vez tendo localizado o argumento da estabilidade nas estratégias justificatórias, buscamos mostrar porque Rawls supõe que os princípios de justiça como equidade, escolhidos provisoriamente na primeira etapa da posição original, seriam relativamente mais estáveis que o princípio da utilidade.

**Palavras-chave:** Justiça como Equidade. Congruência entre o Correto e o Bem. Psicologia Moral.



## ABSTRACT

This dissertation aims to present a study on the arguments for stability developed in Part III of *A Theory of Justice*, with special focus on the argument of the congruence between the correct and the good understood from the point of view of the structure of moral theory. Our strategy to achieve this goal is to highlight how the theory of justice as fairness defines and articulates the moral notions of right and of good in its moral structure from a non-rigorous deontology. The next step is to point out the scopes of the specification of the correct and the description of the good in the structure of the theory and show the implication of these two scopes in the two strategies of justification of the principles - original position and reflective equilibrium. In doing so, we have shown that both strategies possess are structured in two stages and we can locate the argument of stability as being part of the second argumentative step of the original position. Once we have located the argument of stability in the justificatory strategies, we seek to show why Rawls supposes that the principles of justice as fairness, provisionally chosen in the first stage of the original position, would be relatively more stable than the principle of utility.

**Keywords:** Justice as Fairness. Stability. Congruence Between the right and the Good. Moral Psychology.



## **SOBRE AS CITAÇÕES**

Neste trabalho, demos preferência ao uso das obras de John Rawls traduzidas para a língua portuguesa publicadas no Brasil. Nos casos em que tradução brasileira, por ventura, não acompanhar o texto original optaremos por fazer uma tradução livre do texto em inglês acompanhada da devida referência.

Todas as traduções do inglês para o português dos textos do *Collected Papers* utilizados e dos comentários à obra de John Rawls citados no presente trabalho são de minha autoria, sendo minha, portanto, a responsabilidade por eventuais falhas e imprecisões.

Ao citar os escritos de Rawls, optamos por informar na referência a sigla do livro utilizado, pois acreditamos que a identificação imediata da obra a qual estamos nos referindo facilita o processo de leitura. Ademais, ao adotar esse modelo da citação da abreviação da obra, seguimos o padrão empregado pelos principais comentadores de John Rawls.



## LISTA DE ABREVIATURAS

Siglas utilizadas dos textos de John Rawls citados.

### *Obras traduzidas para o português*

- TJ *Uma Teoria da Justiça*. Tradução de Jussara Simões. Revisão técnica e da tradução Álvaro de Vita. 3 ed. São Paulo: Martins Fontes, 2008.
- JF *Justiça como equidade: uma reformulação*. Tradução de Claudia Berliner. Revisão técnica e da tradução Álvaro de Vita. São Paulo: Martins Fontes, 2003.
- PL *Liberalismo Político*. Tradução de Álvaro de Vita. 2 ed. São Paulo: Martins Fontes, 2011.
- DP *Direito dos povos*. Tradução de Luís C. Borges. São Paulo: Martins Fontes, 2004.
- CHFP *Conferências sobre a história da filosofia política*. Tradução de Fabio M. Said. São Paulo: WMF Martins Fontes, 2012.
- JE:PM *Justiça como equidade: uma concepção política, não metafísica*. Tradução de Álvaro de Vita. *Lua Nova*. São Paulo. n. 25, 1992. pp. 25-59.
- HFM *História da Filosofia Moral*. Tradução de Ana Aguiar Cotrim. São Paulo: Martins Fontes, 2005.

### *Textos originais*

- IMT Independence of Moral Theory (1975). In: *Collected Papers*. Samuel Freeman (Org.). Cambridge: Harvard University Press, 1999, pp. 286-302.
- FG Fairness to Goodness (1975). In: *Collected Papers*. Samuel Freeman (Org.). Cambridge: Harvard University Press, 1999. pp. 267-285.

- KCE Kantian Conception of Equality (1975). In: *Collected Papers*. Samuel Freeman (Org.). Cambridge: Harvard University Press, 1999, pp. 254-266
- KCMT Kantian Constructivism in Moral Theory (1980). In: *Collected Papers*. Samuel Freeman (Org.). Cambridge: Harvard University Press, 1999. pp. 303-258.
- PRIG The Priority of Right and Ideas of the Good. In: *Collected Papers*. Samuel Freeman (Org.). Cambridge: Harvard University Press, 1999. pp. 449-472.
- RAM Reply to Alexander and Musgrave (1974). In: *Collected Papers*. Samuel Freeman (Org.). Cambridge: Harvard University Press, 1999. pp. 232-253.
- SJ The Sense of Justice (1963). In: *Collected Papers*. Samuel Freeman (Org.). Cambridge: Harvard University Press, 1999. pp. 96-116.

#### *Abreviatura de termos*

- PO Posição Original  
SBO Sociedade Bem-Ordenada

## SUMÁRIO

<b>INTRODUÇÃO .....</b>	<b>19</b>
<b>1. APRESENTAÇÃO GERAL DA TEORIA .....</b>	<b>25</b>
1.1 Delineando o contexto e principais antagonistas teóricos .....	26
<b>1.1.1 Positivismo Lógico .....</b>	<b>27</b>
<b>1.1.2 Utilitarismo .....</b>	<b>28</b>
<b>1.1.3 Intuicionismo.....</b>	<b>32</b>
1.2 Contratualismo e a primazia do justo sobre o bem.....	35
<b>2. A ESTABILIDADE EM UMA TEORIA DA JUSTIÇA: INTRODUÇÃO AO PROBLEMA .....</b>	<b>41</b>
2.1 Que tipo de estabilidade?.....	45
2.2 Ameaças à estabilidade .....	53
<b>3. O PROBLEMA DA CONGRUÊNCIA ENTRE CORRETO E O BEM DO PONTO DE VISTA DA ESTRUTURA DA TEORIA MORAL .....</b>	<b>57</b>
3.1 A justiça como equidade enquanto parte da Teoria Moral.....	57
3.2 Distinções gerais entre as noções de correto e de bem .....	65
3.3 Sobre o modo como são articuladas as noções morais na estrutura da teoria utilitarismo e da justiça como equidade.....	69
3.4 A estrutura do Deontologismo rawlsiano .....	82
3.5 Localizando o problema da estabilidade na estrutura justificatória da teoria moral.....	88
3.6 As duas etapas argumentativas da Posição Original.....	96
<b>4. O PAPEL DA PSICOLOGIA MORAL.....</b>	<b>105</b>
4.1 O papel da psicologia moral na escolha dos princípios na Posição Original.....	109
4.2 Senso de Justiça e Desenvolvimento Moral.....	112
<b>CONSIDERAÇÕES FINAIS .....</b>	<b>117</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>119</b>
<b>DIAGRAMAS.....</b>	<b>125</b>



## INTRODUÇÃO

O objetivo do presente trabalho é apresentar e discutir a explicação da estabilidade apresentada na Parte III de *Uma Teoria da Justiça* (TJ), com ênfase no argumento da congruência entre o correto e o bem. Ao fazer esse caminho, pretendemos evidenciar para o leitor como Rawls respondeu em seu primeiro livro a questão que foi uma das motivações subjacentes também em suas outras duas obras (*Liberalismo Político* e *Direito dos Povos*), a saber, o que a justiça requer de nós, e dado o que ela requer, se somos, nós seres humanos, capazes disso. Nas palavras de Samuel Freeman, “a questão é saber se a humanidade é capaz de justiça, ou no caso de nossa natureza ser tão egoísta, apavorada e corrupta, como colocar a justiça entre as capacidades (*capabilities*) humanas.” (FREEMAN, 2007a, p. 323) Esperamos que ao final dessa jornada tenhamos deixado claro porque alguns dos pressupostos de TJ implicaram em uma exigência demasiada da justiça para uma sociedade democrática e que, por isso, conduziram Rawls a reformular sua teoria para torná-la compatível com a condição da cidadania democrática.

Desde sua publicação em 1971 até hoje, as ideias apresentadas no livro *Uma Teoria da Justiça* do filósofo americano John Rawls têm sido intensamente estudadas e debatidas no mundo todo, mobilizando pesquisadores de várias áreas do conhecimento. Nesses quase 50 anos de debates, a produção bibliográfica em torno das obras de Rawls gerou uma infinidade de traduções, comentários, compêndios, resenhas e artigos<sup>1</sup> (até mesmo um

---

<sup>1</sup> *A Theory of Justice* foi traduzida para 27 idiomas (FREEMAN, 2007a, p. 323), e foram impressas mais de 300.000 cópias em inglês (MAFFETTONE, 2010, p. vii). Só no Brasil, já estamos na terceira tradução de TJ: a primeira foi feita por Vamireh Chacon, publicada em 1981 pela Editora Universidade de Brasília; a segunda por Almiro Pisetta e Lenita M.R. Esteves, publicada em 1997, pela Livraria Martins Fontes Editora, com segunda edição em 2002; a terceira edição, publicada em 2008, contou com nova tradução, feita por Jussara Simões e revisada por Álvaro de Vita, teve uma segunda tiragem em 2010; em 2016 foi publicada a quarta edição da TJ, também pela Martins Fontes. Não é apenas na venda de livros que o impacto da *Theory* pode ser percebido, mas também pela quantidade de estudos que influenciou. Quanto à produção acadêmica em torno dos textos de Rawls, podemos mencionar que na base de dados JSTOR (um sistema online de arquivamento de periódicos acadêmicos), por exemplo, em pesquisa de acervo realizada para a categoria 'John Rawls' encontramos 21.400 artigos acadêmicos de algum modo relacionados ao filósofo e sua teoria.

musical já foi produzido sobre *Uma Teoria da Justiça*)<sup>2</sup> que mudaram a agenda e vocabulário da filosofia política contemporânea. Não obstante, apesar de todo esse florescimento, apenas recentemente um dos temas tratados por Rawls passou a receber mais atenção<sup>3</sup>, a saber, o argumento da estabilidade da concepção de justiça desenvolvido na Parte III de TJ. Tal questão depende da descrição da aquisição e desenvolvimento do senso de justiça (Capítulo VIII da TJ) e justificar como a concepção de justiça pode fazer parte do bem do indivíduo (o argumento da congruência entre o correto e o bem) (Capítulo XIX da TJ).

Sabemos que uma obra de vigor filosófico nunca é completamente exaurida, sempre são possíveis novas interpretações que dão mais importância a um ou outro aspecto que outrora não fora notado ou correlacionado (prova é que Platão e Aristóteles são até hoje intensamente debatidos). No presente caso, contudo, é pertinente buscar alguma explicação para o fato de que as quase 200 páginas (182 páginas, para ser exata) que compõem a Parte III de um livro amplamente divulgado, debatido e traduzido terem entrado na agenda de discussões apenas 25 anos depois da publicação de TJ, mesmo apesar de o filósofo ter alertado ainda no prefácio de 1971 que sem a análise dos argumentos da última parte do livro corria-se o risco de a teoria da justiça não ser compreendida corretamente (TJ, p. XLV). Das duas etapas do argumento da estabilidade da concepção de justiça desenvolvidas na Parte III, o argumento da congruência (que trata mais especificamente da motivação moral), elaborado no Capítulo XIX, foi ainda mais negligenciado pelos comentadores do que a questão da aquisição e desenvolvimento do senso de justiça. Em uma conversa

---

<sup>2</sup> Em 2013, alunos da Universidade de Oxford escreveram e produziram o musical *A Theory of Justice: The Musical!* ([www.demproductions.org/#!atojtm/c9dh](http://www.demproductions.org/#!atojtm/c9dh))

<sup>3</sup> Até a publicação do *Liberalismo Político*, o tema da estabilidade recebeu pouca atenção dos comentadores, e os comentários que surgiram lidavam apenas com a primeira parte do argumento: o senso de justiça e psicologia moral. Nos comentários e compêndios sobre a obra de Rawls, o tema da estabilidade, ainda ligado apenas ao senso de justiça, começa a aparecer nos sumários a partir do ano 1999, com RICHARDSON, WEITHMAN. *Moral Psychology and Community* (Philosophy of Rawls-Volume 4). De acordo com Thomas Baldwin, em 2002, Chandran Kukhatas editou 4 volumes com artigos críticos e comentários à obra de Rawls (*John Rawls: Critical Assessments of Leading Political Philosophers*. London; New York: Routledge, 2002), entretanto, não há nenhum artigo que trate diretamente sobre a questão da psicologia moral (Cf. BALDWIN, 2008, p. 247). Apenas a partir de 2003, com o *Cambridge Companion to Rawls* surgiram publicações investigando também o argumento da congruência.

pessoal com Samuel Freeman, Rawls revelou ter ficado perplexo com a pouca repercussão do argumento da congruência, pois acreditava que esta tinha sido uma das contribuições mais originais que havia feito em TJ (Cf. FREEMAN, 2007a, p. 143, n.2). Em outra ocasião, declarou que esta foi a parte do livro que mais gostou, tanto que após concluí-lo planejava seguir sua pesquisa acadêmica em temas conectados com psicologia moral, no entanto, em razão da grande repercussão da publicação teve de mudar de planos e concentrar sua atenção em explicar a justiça como equidade de modo mais convincente.<sup>4</sup> Rawls dedicou mais de 40 anos de sua vida acadêmica ao desenvolvimento de sua teoria da justiça como equidade; foram pelo menos 20 anos<sup>5</sup> para concluir *Uma Teoria da Justiça*, e nos 22 anos que se seguiram até a publicação do *Liberalismo Político*, praticamente todos seus textos publicados têm relação direta com o processo de elaboração da teoria. Foi tamanha sua dedicação que podemos afirmar, junto com Miguel Rodilla, que Rawls “é um caso infrequente de autor de uma só obra, (...) Sem incorrer em exagero, pode-se dizer que empenhou todo seu trabalho na composição de um só texto” (RODILLA, 2013, p. 291). Talvez o próprio Rawls concordasse com a assertiva de Rodilla, pois, em evento realizado em outubro de 1995 na Universidade de Santa Clara na Califórnia para celebrar os 25 anos da publicação de TJ, disse, num misto de ironia e modéstia, que “a principal razão para seu livro ter sido tão amplamente lido e citado foi que em sua vida ele havia tido apenas uma ideia e escrito apenas um livro” (Cf. MAFFETTONE, 2010, p. 211). Essa fala de Rawls indica que todos seus escritos podem ser entendidos como fazendo parte de “um projeto unitário, no qual não há justiça moral sem estabilidade institucional” (MAFFETTONE, 2010, p. 19).

Algumas características da própria TJ podem nos ajudar a entender o que levou ao lapso exegético anteriormente mencionado. Na Parte III estão as páginas que dão o fechamento a um conjunto de ideias complexas que foram desenvolvidas e aperfeiçoadas ao longo de 20 anos até culminar no livro de 587 páginas publicado em

---

<sup>4</sup> John Rawls: For the Record. *The Harvard Review of Philosophy* (Spring 1991), p. 44

<sup>5</sup> Em uma entrevista, ao ser questionado sobre quando iniciou o pensamento que o conduziu à *Uma Teoria da Justiça*, Rawls respondeu que começou a coletar notas depois de ter acabado sua tese, em 1950. E que foi mais ou menos no ano de 1951 que ele teve a ideia que veio se transformar na ‘posição original’. John Rawls: For the Record. *The Harvard Review of Philosophy* (Spring 1991), p. 39.

1971. Some-se a isso o fato de que nessas 182 páginas finais o argumento da estabilidade foi desenvolvido de modo intermitente e com uma falta de clareza que não é habitual a Rawls. Ele chegou a admitir, treze anos mais tarde, que em relação à parte final do livro “um número excessivamente grande de conexões foram omitidas, esperando-se que o leitor as fizesse, a tal ponto que se pode ficar em dúvida sobre qual é o tema de boa parte dos capítulos 8 e 9” (JE:PM, p. 58, n.33). Talvez não seja insensato concordar com Samuel Freeman quando sugere que essa parte acabou sendo negligenciada nas discussões por pura exaustão (FREEMAN, 2007a, p. 143).

Do ponto de vista da estrutura e do conteúdo da teoria, a complexidade e dificuldade da Parte III também se explica pelo fato de ali Rawls executar aquele que considera ser o principal esforço teórico da Teoria Moral: o estudo comparativo da estrutura de cada doutrina moral a partir do modo como cada uma delas especifica e relaciona as noções básicas de correto, bem e valor moral e como tal arranjo se relaciona com nossos sentimentos morais. Depois de ter apresentado sua concepção do correto (como equidade) e de ter formulado o procedimento para a escolha dos princípios na Parte I, Rawls desenvolve na última parte do livro uma concepção de bem (como racionalidade) para, ao fim, mostrar como a articulação entre os conceitos de correto e de bem por ele apresentados explicam o terceiro principal conceito da ética, o valor moral (os motivos pelos quais agimos). É na Parte III que Rawls faz a conexão entre teoria moral e psicologia moral, onde relaciona o conceito de correto da justiça como equidade com os sentimentos morais e, ao mesmo tempo, para evidenciar a força de sua teoria, compara com o modo como o utilitarismo faz tal relação. Em termos gerais, o argumento da congruência consiste em mostrar que a concepção de justiça especificada na primeira etapa argumentativa da posição original é congruente, ou pelo menos, não incompatível, com a concepção de bem apresentada na segunda parte. A suposição de Rawls é que os dois princípios da justiça como equidade têm uma probabilidade muito maior, se comparado com o princípio da utilidade em uma SBO, de ser visto como parte do bem das pessoas.

O fato é que a Parte III entrou definitivamente na agenda de discussões depois que Rawls chamou a atenção para a questão na Introdução de seu novo livro, *Liberalismo Político* (PL) publicado em 1993. Ele afirma que para entender a natureza e a extensão das

mudanças que o liberalismo político introduziu na teoria da justiça como equidade é preciso compreendê-las como um esforço para resolver o que considera um grave problema interno de TJ: a explicação da estabilidade de uma sociedade bem-ordenada na Parte III “não é coerente com a visão de um todo” e torna irrealista a ideia de sociedade bem-ordenada, tal como aparece em *Teoria*, por isso a necessidade de reformulá-la (PL, p. XVII). No entanto, não foi toda a explicação da estabilidade em TJ que Rawls tomou como errônea ou ambígua, mas apenas uma parte dela, tanto que é parte dos argumentos foram mantidos no PL e reafirmados em JF. Neste trabalho acompanharei a interpretação dos dois principais comentadores de Rawls no que se refere à questão da estabilidade em TJ, Samuel Freeman e Paul Weithman, que defendem, por razões distintas, que a falha que tornou a SBO irrealizável e, por isso, motivou Rawls à rever sua teoria, foi o argumento da congruência. (FREEMAN, 2007a, p. 143; WEITHMAN, 2010, p.43)

O propósito do presente trabalho não é fazer uma análise da transição da TJ para o PL, nem comparar o modo como a estabilidade e seus argumentos são tratados em cada um dos livros ou mesmo fazer uma análise sobre o desenvolvimento da questão estabilidade ao longo dos textos de Rawls, mas antes, se resume a compreender a estrutura, a argumentação e o papel justificatório da estabilidade apenas em TJ, a partir dos objetivos do filósofo ao escrever o livro, das principais teorias e problemas a que estava respondendo. Por isso, esclarecemos ao leitor que as referências aos *Collected Papers*, *Liberalismo Político* e *Justiça como Equidade: uma reafirmação*<sup>6</sup> são feitas apenas na medida em que nos auxiliam a compreender os problemas e argumentos apresentados em *Uma Teoria da Justiça*.

---

<sup>6</sup> No Brasil, o título do livro *Justice as Fairness: a Restatement* foi traduzido como *Justiça como Equidade: uma Reformulação*. O termo *restatement*, de acordo com o dicionário Oxford, significa afirmar algo de novo ou diferentemente, especialmente de modo mais claro e convincente. É um termo que admite a tradução tanto por ‘reformulação’, quanto por ‘reafirmação’. Se, por um lado, a tradução atual do título do livro seja cabível e preserva uma das propostas do livro, por outro, não contempla um sentido importante do termo que é a ‘reafirmação’. Em *Justice as Fairness*, mais do que estar reformulando a sua teoria da justiça como equidade, Rawls a está reafirmando: apresentando uma versão sintética da teoria que desenvolveu, reformulou e aperfeiçoou ao longo de 50 anos. Por esta razão, adotaremos ‘reafirmação’ como tradução mais adequada de *restatement* uma vez que este termo melhor preserva o sentido do livro de Rawls. Agradeço ao prof. Dr. Denilson Werle por estes esclarecimentos.



## 1. APRESENTAÇÃO GERAL DA TEORIA

*Uma Teoria da Justiça*, publicado em 1971, é uma das obras que mais impactou a filosofia moral e política no século passado. O livro suscitou muitas questões, críticas e interpretações não apenas no âmbito da Filosofia, como também na Economia, Psicologia, Direito, dentre outras ciências sociais. Tratar de todos, ou da maioria dos pontos discutidos não é o objetivo deste trabalho, que pretende manter o foco na questão da estabilidade, em especial na argumentação em favor da congruência entre o correto e o bem apresentada na Parte III de TJ. No entanto, antes de tratar desse tema, convém fazer uma apresentação introdutória da principal obra de Rawls.

Rawls parte do pressuposto filosófico, e que já faz parte do senso comum democrático, de que existe, de um lado, “uma distinção entre as exigências da liberdade e do correto (*right*), e o desejo de aumentar o bem-estar agregado, de outro” (TJ, §6, p. 33) e que nós, cidadãos democráticos temos a tendência de dar prioridade, no mais das vezes, às exigências da liberdade e do correto. Por isso, busca defender uma doutrina moral cuja estrutura reivindique a primazia do correto sobre o bem, o que implica uma sociedade em que as instituições da estrutura básica tenham compromisso primeiro com a justiça, e não com alguma concepção de bem específica. Como então fundamentar a prioridade do justo sobre uma concepção de bem? Rawls parece ter diagnosticado uma situação complexa em que estavam os estudos em língua inglesa sobre a relação entre os dois principais conceitos da filosofia moral – correto e bem. A partir dos anos 1930, com a virada linguística, o pensamento anglo-saxão se desenvolveu eminentemente em torno do Positivismo Lógico. No campo da filosofia moral, o positivismo desenvolveu-se na forma do Expressivismo moral, corrente teórica que não reconhece qualquer objetividade nos juízos morais, pois considera que um juízo de valor é apenas a expressão de um sentimento, seja ele de aprovação ou desaprovação. Dessa perspectiva, a discussão sobre as concepções de bem e de correto ficam restritas à manifestação de sentimentos e impossibilitadas de falar objetivamente sobre eles.

Por outro lado, a tradição moral dominante até a virada linguística era o utilitarismo, que se caracteriza por reduzir todos os

juízos de valor a um único princípio: a utilidade, ou satisfação de um desejo racional. A estrutura geral da doutrina utilitarista clássica define o bem de modo independente anterior ao correto, e subordina este à maximização do bem. Os intuicionistas, por sua vez, a despeito de terem feito críticas bem sucedidas ao princípio da utilidade, não tiveram êxito em elaborar uma teoria tão sistemática e vigorosa quanto o utilitarismo para substituí-lo. De um modo geral, os intuicionistas negam a possibilidade de “uma solução explícita e útil para o problema da prioridade” (TJ, §7, p. 49).

Enfim, este era o panorama que Rawls tinha a sua frente e que desejava superar ao propor sua TJ. Do ponto de vista do método, não poderia se apoiar inteiramente nos paradigmas do Positivismo Lógico, que sequer considera a Filosofia Moral e Política como um *locus* de investigação filosófica; e do ponto de vista substantivo, recusava as consequências do utilitarismo, e o intuicionismo, por sua vez, não apresentava soluções de modo que pudesse ser um substituto à altura. Ainda assim, Rawls encontra alguma inspiração nessas teorias antagonistas: pretende desenvolver uma teoria moral da justiça cujos juízos tenham objetividade, e deseja que sua teoria seja tão sistemática e vigorosa como o utilitarismo. Para dar conta de seu projeto, retoma a tradição do contrato social.

O empenho de Rawls em discutir o papel do correto e do bem tem a ver com sua aspiração de desenvolver uma concepção filosófica de justiça que possa servir como base à uma democracia constitucional. A doutrina utilitarista, que dominou por cerca de três séculos as reflexões morais na filosofia anglo-saxã, é considerada uma base muito frágil para sustentar instituições democráticas, pois não consegue oferecer “uma teoria convincente dos direitos e das liberdades fundamentais e de sua prioridade” (TJ, p.XXXVI).

### **1.1 Delineando o contexto e principais antagonistas teóricos**

É apropriado fazer um breve panorama da filosofia na primeira metade do século XX, com especial ênfase à filosofia moral e política anglo-saxã, pois foi nessa atmosfera filosófica que John Rawls fez sua formação e a partir da qual formulou suas proposições. Nesse período, o pensamento político e moral encontrava-se preso nos paradigmas e limites do positivismo lógico,

por um lado, e por outro, pela doutrina utilitarista clássica, sendo esta o principal adversário teórico de Rawls.

### **1.1.1 Positivismo Lógico**

Embora Rawls não fale diretamente do positivismo lógico, é importante incluí-lo aqui para compreender a situação com a qual se deparava um pensador da época que desejasse estudar filosofia moral (Cf. VITA, 2008).

O positivismo lógico foi um movimento filosófico que se desenvolveu com vigor entre as décadas de 1920 e 1940 em Viena. Os positivistas consideram que é através da linguagem que conhecemos o mundo e, por isso, defendem que o método mais adequado para investigações filosóficas é a análise lógica da linguagem. Aos positivistas interessa apenas a linguagem que é capaz de trazer alguma informação objetiva sobre o mundo, por isso sua análise lógica só pode ser feita em proposições que possuem algum conteúdo cognitivo, ou seja, nas que descrevem um estado de coisas existente que pode ser verificado, ou não, na realidade. As proposições prescritivas e valorativas, características da Ética e Política, não são consideradas como possuindo conteúdo cognitivo, sob a justificativa de que não há como verificar objetivamente na realidade comportamentos e valores. Desta perspectiva, chamada Emotivismo (ou expressivismo) moral, os juízos valorativos não possuem nenhum conteúdo logicamente analisável, uma vez que não descrevem uma característica de um objeto, mas somente expressam ou descrevem um estado emocional do sujeito. Como o emotivismo não reconhece qualquer valor objetivo nos juízos morais, proposições como 'Matar é errado' e 'Mentir é imoral', são vistas apenas como a expressão do descontentamento do locutor com tais atos, e que não há qualquer objetividade sobre a qual se possa justificar que é moralmente errado matar ou mentir. No caso específico da justiça social, o emotivismo moral pode ser verificado nas assertivas dos filósofos conservadores que refutam as reivindicações por igualdade e distribuição de bens alegando que não possuem nenhum conteúdo com força normativa, pois estariam apenas descrevendo o estado emocional do indivíduo, isto é, tais

reivindicações são tidas como expressão da inveja, ressentimento, indignação, despeito.<sup>7</sup>

O Positivismo Lógico teve o mérito de imprimir mais rigor na análise da linguagem, mas também tornou árido o desenvolvimento de áreas da filosofia ao negar a possibilidade de haver discussão racional acerca de juízos de valor. Se, de um lado, a filosofia moral anglo-saxã estava presa ao ceticismo do positivismo lógico quanto à objetividade dos juízos de valor, por outro, podia contar com a longa tradição utilitarista que fornece um critério objetivo para correção de todos os juízos de valor: a utilidade.

Rawls defende que há um conteúdo objetivo que pode ser racionalmente conhecido nos juízos morais sobre a justiça igualitária; busca mostrar que seu igualitarismo não é produto da racionalização dos sentimentos de inveja ou ressentimento, mas sim de um sentimento de justiça cujo conteúdo é definido a partir de um ponto de vista equitativo que pode razoavelmente ser aceito por todos.

---

<sup>7</sup> Para Frederick von Hayek, a justiça social se fundamenta na inveja, no ódio aos que desfrutam uma posição confortável. De acordo com o economista austríaco “o evangelho da ‘justiça social’ visa a sentimentos muito mais sórdidos: a aversão aos que estão em condições melhores, ou simplesmente a inveja, (...) essa animosidade para com a grande fortuna, que considera um ‘escândalo’ que alguns desfrutem da riqueza enquanto outros têm necessidades básicas insatisfeitas, e esconde sob o nome de justiça o que nada tem a ver com ela” (HAYEK, 1985, p. 120). Seguindo esta linha, o espanhol Fernandez de La Mora defende que não há nenhuma justificativa com fundamento na justiça para defender a taxaço progressiva, nem a taxaço sobre grandes fortunas – questões caras à justiça social – pois tais reivindicações seriam apenas a expressão, nas palavras de De La Mora “do desejo de punir ou eliminar os ricos e as pessoas de maior capacidade” (DE LA MORA, p. 186), desejo esse que teria sua origem na inveja. Para ele, a justiça social nada mais é que “o disfarce contemporâneo da inveja coletiva” (DE LA MORA, p. 93). O alemão Max Scheler, citado por Helmut Schoeck, defende que o ressentimento se esconde na demanda por igualdade (*equality*). “Na realidade, [aquele que demanda por igualdade] deseja nada menos do que a destruição de todos aqueles que incorporam os valores superiores que despertam sua raiva” (SCHOECK, 1969, p. 282). Para tais pensadores, parece haver uma conexão necessária entre o igualitarismo e a inveja e ressentimento, e, por isso, defendem que insistir na reivindicação de igualdade e justiça social apenas serve para estimular comparações interpessoais desfavoráveis, promovendo, deste modo, sentimentos que são pessoal e socialmente destrutivos. Assim, o igualitarista que persistir nesse campo “está condenado a uma vida não só de resmungos e inveja eterna, mas de decepção interminável e inevitável” (LUCAS, 1965, p. 306).

### 1.1.2 Utilitarismo

Rawls deixa claro desde o prefácio e ao longo de *Uma Teoria da Justiça* que seu objetivo principal é elaborar uma concepção de justiça sistemática e razoável para servir como base das instituições de uma democracia constitucional, e é importante que tal concepção seja uma alternativa viável ao utilitarismo em geral. Essa afirmação significa que boa parte dos problemas levantados e respondidos em TJ são uma resposta aos limites e falhas incontornáveis do princípio da utilidade.

Mesmo antes da crítica rawlsiana à doutrina utilitarista, outros filósofos e correntes teóricas se dedicaram a questionar os pressupostos e consequências sociais da aplicação do princípio da utilidade. Embora várias das críticas feitas pelos intuicionistas e idealistas fossem pertinentes, não se conseguiu formular uma doutrina alternativa com o mesmo grau de refinamento e sistematicidade que pudesse ser uma substituta à altura do utilitarismo (Cf. CHFP, p. 427). Na falta de uma teoria melhor, os pensadores acabam por ter que aquiescer com o utilitarismo dominante.

Rawls reconhece que a doutrina utilitarista tem uma tradição longa e vigorosa na filosofia moral de língua inglesa e que possui muitos méritos. Ele aponta que definir a satisfação como critério objetivo para correção moral representou um avanço em relação à tradição moral que lhe era anterior, pois, com isso, secularizou a moral. Com adoção de tal critério, os fundamentos para a obrigação moral não são mais buscados fora da experiência, o que permite à ética emancipar-se de pressupostos teológicos e transcendentais. O utilitarismo foi, de acordo com Rawls, a primeira tradição com condições de fornecer uma base moral filosófica sistemática, não metafísica e dissociada de discursos religiosos, para as instituições políticas, tornando-as compatíveis “com os pressupostos seculares e com as tendências do mundo moderno”. Por isso Rawls afirma que “o utilitarismo clássico aceitou como condição do estado da cultura moderna o fato de que uma concepção moral e política deve obrigatoriamente ser secular” (CHFP, p. 429).

A despeito de seus méritos para o mundo moderno, os limites do princípio da utilidade acabam por tornar a concepção utilitarista de justiça inadequada para orientar as principais instituições sociais

em democracias contemporâneas. O tipo de utilitarismo que importa para Rawls não é enquanto uma teoria moral da ação individual, mas sim quando aplicado à distribuição dos bens na sociedade<sup>8</sup>.

De acordo com Rawls, a estrutura de uma doutrina ética depende de como ela relaciona os conceitos de correto e de bem e de como define as diferenças entre ambos (Cf. TJ, §5, p. 29; §68, 552). A estrutura geral do utilitarismo se caracteriza por definir o bem de modo anterior e logicamente independente do correto e, só posteriormente define o que é o correto, como sendo aquilo que eleva o bem ao máximo. Esta estrutura de argumentação é o foco principal da crítica de Rawls e sua proposta representa uma inversão, pois defenderá que o correto seja definido antes do bem, e limitará as concepções de bem à concepção de justiça.

Vejamos agora com mais detalhes como o utilitarismo define e relaciona as ideias de bem e de correto. O que é descrito como 'bem' pode mudar de uma teoria teleológica para outra: o bem do hedonismo é o prazer, o do perfeccionismo pode incluir certas formas de excelência humana (perfeições humanas e outras) ou certos valores (a beleza das coisas, o conhecimento do mundo) (Cf. CHFP, p. 431), do eudaimonismo é a felicidade. O bem do utilitarismo (ou melhor, da versão utilitarista que Rawls está interpretando) é a satisfação do desejo (ou preferências) racional, e, neste sentido, "age muito bem a pessoa que procura alcançar o máximo de seu próprio bem e promover tanto quanto possível seus objetivos racionais" (TJ, §5, p. 31). Nessa formulação, o princípio da utilidade é um princípio de escolha para uma única pessoa e funciona para avaliar a ação individual. Se é racional um indivíduo agir de modo a promover e aumentar seu próprio bem-estar através da busca da satisfação de seus desejos, parece sensato que as instituições sociais também dirijam seus fins com vistas a elevar ao

---

<sup>8</sup> É o que Rawls chama de doutrina clássica estrita, cuja ideia principal é a de que "a sociedade está ordenada de forma correta, e, portanto, justa, quando suas principais instituições estão organizadas de modo a alcançar o maior saldo líquido de satisfação, calculado com base na satisfação de todos os indivíduos que a ela pertencem", e tem em Henry Sidgwick seu principal representante (TJ, §5, p. 27). Rawls considera que a obra de Sidgwick, *The Methods of Ethics* (1907) representa a formulação mais clara e acessível da doutrina utilitarista clássica, "segundo a qual o fim moral último de todo ato social ou individual é obter a maior soma líquida possível de felicidade para todos os seres sensíveis" (CHFP, p. 410).

mais alto grau o saldo de realização dos interesses coletivos. “Assim, chega-se ao princípio da utilidade de modo natural: a sociedade está ordenada adequadamente quando as instituições maximizam o saldo líquido de satisfações” (TJ, §4, p. 26). Para fazer essa passagem problemática de princípio individual para coletivo, o utilitarismo insere a figura do Observador Imparcial. Como o bem de uma coletividade é difícil de ser determinado em razão de existir uma pluralidade de concepções de bem, adota-se a concepção daquilo que é bom a partir dos princípios de escolha racional para um único indivíduo: o observador imparcial e solidário. Esse observador idealmente racional e imparcial se identificaria com os desejos dos outros e os experimentaria como se fossem de fato seus, “encararia cada membro da sociedade exatamente como cada um de nós individualmente encara suas próprias necessidades” (ESTEVES, 2002, p. 5). Depois de avaliar sua intensidade, organizaria todos os desejos de todas as pessoas em um único sistema coerente de desejos, e aplicaria este sistema a toda a sociedade. Uma sociedade ordenada de maneira justa seria então aquela que recebe a aprovação deste observador imparcial e solidário, capaz de colocar-se na situação dos outros.

Quando estende para toda a sociedade o princípio de escolha racional para um único ser humano, o utilitarismo desconsidera as diferenças entre os indivíduos e funde todas as pessoas numa só.<sup>9</sup> Adotando o princípio utilitarista em escala social, a distribuição correta da satisfação entre os indivíduos será tanto mais justa quanto maior for a tendência a produzir satisfação, mesmo que isso implique que uma parcela da sociedade tenha seus desejos racionais sacrificados (Cf. TJ, §6, p. 31-2).

Uma das implicações sociais da teoria utilitarista é o grande esforço que exige das pessoas que acabariam entre os menos favorecidos. Essas pessoas teriam que cooperar com instituições que promovem o bem-estar dos outros ao custo do sacrifício de seus

---

<sup>9</sup> Cf. TJ, §6, p. 32-35. Veremos que a justiça como equidade opera no sentido oposto do utilitarismo. Enquanto este opera segundo a lógica daquilo “que é racional para um homem, é justo para uma associação de homens” (TJ, §5, p. 28), a concepção rawlsiana define primeiro o que é o justo para uma “associação de homens”, no caso, as partes representativas na Posição Original, para, então na Parte III da TJ, através dos argumentos da congruência mostrar como o justo pode ser racional para um indivíduo, e aí sim, “(...) e sendo racional para qualquer pessoa, é racional para todos nós” (TJ, §86, p. 700).

próprios desejos racionais e isso seria muito prejudicial não só para a autoestima dos indivíduos, como também para o desenvolvimento do senso de justiça e para a cooperação social. Num mundo social em que o princípio da utilidade fosse publicamente reconhecido, “os vínculos de fraternidade e reciprocidade são fracos, dificilmente uniriam toda a sociedade. Alguns grupos talvez tenham pouco desejo de agir com justiça” (TJ, § 76, p. 616-7). De fato é muito difícil imaginar como, em sociedades que estimam os valores da liberdade e igualdade, as pessoas menos favorecidas apoiariam espontaneamente - sem serem coagidas, sem terem sua opinião manipulada e sem serem submetidas à doutrinação forçada - uma concepção de justiça que considera justo lhe ceifar a autoestima, liberdades individuais e oportunidades para garantir um maior bem-estar a outras pessoas - que geralmente não estão entre os menos favorecidos.

Diferente de outras teorias políticas, como a de Kant, em que os interesses dos menos favorecidos pelo sistema de cooperação - como mulheres e pessoas não proprietárias - não são relevantes na definição dos termos de cooperação, na teoria da justiça formulada por Rawls, os menos favorecidos não apenas são levados em conta como têm, por assim dizer, poder de veto. Se os princípios escolhidos forem tais que, ao serem institucionalmente aplicados, exigirem dos menos favorecidos um compromisso maior do que a capacidade humana suporta, a concepção de justiça não será validada pelas partes na posição original. A estabilidade de uma sociedade passa a depender também dos menos favorecidos. Aqueles que já são mais favorecidos - seja por dons naturais ou pela aleatoriedade da classe econômica e social em que nasceram - possuem motivos para continuar a cooperar, pois se beneficiam do sistema. O desafio é formular termos em que aqueles menos favorecidos sintam-se motivados a cooperar, sem que a motivação seja baseada em falsas crenças, doutrinação forçada ou coerção.

Os utilitaristas, apesar de ardorosos defensores da liberdade individual não levam a sério as diferenças entre os indivíduos. Uma concepção de justiça que aceita sacrificar o bem-estar de alguns indivíduos em prol de outros não é compatível com valores da liberdade e igualdade, que são essenciais em uma sociedade democrática.

### 1.1.3 Intuicionismo

Um dos propósitos da teoria da justiça como equidade é fornecer uma justificação adequada da prioridade do correto sobre o bem, ao que as doutrinas morais utilitaristas não lograram êxito. Rawls não foi o primeiro filósofo a atacar as falhas do utilitarismo, muitos outros antes dele, dentre os quais os intuicionistas, direcionaram críticas consistentes ao princípio da utilidade, entretanto, não tiveram êxito em formular uma teoria sistemática forte o suficiente para substituir a doutrina utilitarista (Cf. CHFP, p. 427). Por entender que um único princípio moral não é suficiente para dar conta da complexidade dos fatos morais, os intuicionistas reconhecem uma pluralidade de princípios que muitas vezes são concorrentes entre si, e admitem como válidos inclusive juízos que não sejam de natureza ética nas considerações sobre a justiça social (Cf. TJ, §7, p. 48).

Existem muitos tipos de intuicionismo, entretanto Rawls irá se referir a essa doutrina de modo mais amplo, tomando-o como uma forma de pluralismo, cujas características essenciais são: (i) reconhecer uma pluralidade de princípios fundamentais, os quais podem, dependendo da situação, entrar em conflito; (ii) não estabelecer nenhum método de comparação ou regra de prioridade entre tais princípios para oferecer soluções no caso de princípios conflitantes. Nesses casos, a solução que o intuicionismo oferece para lidar com problemas morais é recorrer à intuição daquilo “que nos parece aproximar-se do que é mais justo” (TJ, §7, p. 41). O intuicionismo oferece uma pluralidade de princípios sem hierarquia alguma entre eles e não fornece um critério racional que ajude a equilibrá-los ou a compará-los para ajudar a resolver problemas morais. Nesse sentido uma das afirmações marcantes do intuicionismo é a de que

em nossos juízos da justiça social só o que é possível é chegar por fim a uma pluralidade de princípios fundamentais a respeito dos quais só podemos dizer que nos parece mais correto equilibrá-los de certa maneira e não de outra (TJ, §7, p. 47). (ênfase adicionada)

Nas situações em que se tem que escolher entre um princípio e outro (inclusive se a escolha for entre um princípio ético e um não ético, como o da eficiência) o intuicionismo não oferece razões

necessárias que indiquem o porquê o indivíduo (ou as instituições) deve escolher 'x' e não 'y', recomenda apenas um equilíbrio de razões circunstancial, que implica escolher o princípio ou que pareça mais adequado no momento, como uma mera questão de preferência.

As dificuldades geradas pela da ausência de razões objetivas para justificar a prioridade ou equilibrar uns princípios sobre outros ficam ainda mais evidentes quando refletimos sobre a questão da congruência colocada por Rawls. Considerando a pluralidade de fins humanos, por que um cidadão de uma sociedade bem-ordenada escolheria um plano de vida em que o desejo de apoiar instituições justas e agir com justiça (de acordo com os princípios escolhidos na posição original) é mais forte e regulativo sobre outros desejos, como maximizar a felicidade ou de satisfazer um desejo interesse particular contrário ao senso de justiça, isto é, por quais motivos o indivíduo faria seu desejo de ser justo ter primazia sobre outros objetivos?

Do ponto de vista da motivação moral, a dificuldade adicional do intuicionismo (em especial o intuicionismo racional formulado por David Ross, tal como interpretado por Rawls) é que não oferece razões suficientes para dizer o que nos motiva a agir moralmente. De acordo com a doutrina do ato puramente consciencioso, defendida por Ross, o que nos motiva a agir moralmente é “o desejo de fazer o que é certo e justo, simplesmente por ser certo e justo, não sendo apropriada nenhuma outra *descrição* de tal motivação” (TJ, §72, p. 589). Ou seja, a única descrição da motivação moral oferecida por tal doutrina é afirmar que quando agimos moralmente o fazemos pela mera convicção de que isso é o correto a ser feito. Rawls não descarta esse tipo de motivação, mas considera que ela não é suficiente para sustentar o desejo forte de agir com justiça. A consequência de conceber o desejo de agir com justiça sem oferecer outras descrições, faz com que o desejo de agir com justiça se assemelhe a uma questão de mera preferência, como escolher entre chá ou café, e “embora tal preferência possa existir, torná-la reguladora da estrutura básica da sociedade é mera extravagância” (TJ, §72, p. 590).

A dificuldade das teorias intuicionistas do ponto de vista da motivação moral é que não apresentam nenhuma descrição para o senso de justiça. A vantagem da teoria do contrato é que ela oferece

uma diversidade de descrições que o senso de justiça pode desejar (Cf. Weithman, p. 266; TJ, §86, p. 702), como por exemplo, o desejo de expressar a natureza livre e igual, de manter laços de amizade, amor e confiança mútua. São disposições afetivas que adquirimos durante o desenvolvimento moral e que valorizamos, por isso Rawls as conecta com o senso de justiça (cf. Weithman, p. 286). A variedade de descrições do contrato permite vincular o desejo pela justiça a vários objetos (e não a um único fim predominante, como no caso do utilitarismo), sendo, então, necessário chegar a um “equilíbrio de motivos”.

O intuicionismo não proporciona um procedimento ou regra para equilibrar objetivos sociais e econômicos, motivos morais, nem para lidar com conflitos entre princípios e auxiliar os indivíduos a resolver problemas morais; seu maior defeito é ser impraticável (Cf. O’NEILL, 2003, p. 350). Rawls não refuta inteiramente e nem considera irracionais as doutrinas intuicionistas, mas contesta seu uso para justificar os objetivos da justiça social e o considera insuficiente para, sozinho, dar conta de um critério sistemático e seguro de justiça capaz de mostrar como se justifica a primazia do correto sobre o bem. É necessário ir além dele e fazer uma teoria construtiva (cuja parte principal é a posição original).

## **1.2 Contratualismo e a primazia do correto sobre o bem**

Nesta seção pretendemos apresentar algumas das razões que levaram Rawls a retomar uma tradição da filosofia política que já estava, parcialmente, desacreditada e qual o papel dela na teoria da justiça como equidade, a saber, o contratualismo. Antes, no entanto, procuraremos esclarecer a qual vertente de interpretação contratualista Rawls se filia. De um modo geral, podemos dizer que o contrato foi utilizado como uma estratégia de justificação capaz de fornecer uma fundamentação sistemática da primazia do justo sobre o bem, e assim inverter o modo como o utilitarismo estruturava tais conceitos. O papel do contratualismo pode ser encontrado pelo menos em três momentos da teoria da justiça como equidade: (i) na concepção de pessoa compatível com o modo como se compreendem os cidadãos das democracias constitucionais contemporâneas, (ii) é através do contrato que as características da sociedade bem-ordenada são incorporadas na descrição da posição

original; e (iii) o contrato introduz as restrições para as partes na PO, condição que permite que o acordo resultante seja equitativo (RAM, p. 249).

Uma das principais críticas ao utilitarismo é o modo como articula os conceitos de bem e de correto (*right*). Como dito anteriormente, tal doutrina primeiro define de forma única o que é o bem para um indivíduo - a satisfação do desejo racional -, e por ser um bem, considera-se, então, que é correto. No nível da justiça social, o princípio de escolha do que é o bem para um único indivíduo passa a ser, automaticamente, o bem de uma associação de indivíduos (Cf. TJ, §6, p. 36). A proposta de Rawls consiste em inverter essa estrutura utilitarista definindo primeiro o que é o correto para uma associação de indivíduos, ou melhor, do ponto de vista de partes representativas em uma condição de igualdade (posição original sob o véu da ignorância), para apenas num segundo momento verificar se é racional para uma pessoa desejar aquilo que foi definido como correto pela “associação de indivíduos”. Em resumo, na teoria rawlsiana, a escolha na posição original primeiro define o que é o correto e só depois procura mostrar que o correto corresponde (é congruente) ao bem de um indivíduo, e, nas palavras de Rawls, “sendo racional para qualquer pessoa, é racional para todos” (TJ, §86, p. 700).

Para fazer essa inversão no modo como os dois principais conceitos da moral - correto e bem - se articulam, Rawls buscou inspiração na teoria do contrato social, que, nas palavras de Freeman, estava já agonizante (Cf. FREEMAN, 2007b, p. 14). Não obstante, é justamente no modelo do contrato que Rawls encontra um modo de estruturar a relação entre o conceito de correto e de bem para que a concepção de justiça resultante seja compatível com os valores das democracias contemporâneas. Essa tradição é importante para o pensamento político liberal, pois foi ela que forneceu “a justificação primária para as revoluções democráticas e republicanas do século dezoito, incluindo a Declaração Americana de Independência e a Declaração Francesa dos Direitos do Homem” (FREEMAN, 2007b, p. 14). A teoria clássica do contrato social utiliza o contrato como uma estratégia para justificar a autoridade política como sendo uma criação dos próprios indivíduos, e não fundamentada em entidades divinas. As teorias contratualistas oferecem uma explicação desvinculada de fundamentos religiosos

sobre porque as pessoas são obrigadas a cumprir as leis, que normas são legítimas e quem as pode propor. São os próprios indivíduos, através do contrato, que definem os termos de cooperação para criar a sociedade política. As leis e os sistemas de instituições que as fazem valer passam a ser compreendidas como resultado de um acordo coletivo.

Rawls encontra estrutura de justificação do contrato social elementos para fundamentar uma das marcas de sua teoria, a primazia do correto sobre o bem. A ideia de contrato guarda em si a noção de acordo entre indivíduos que funciona “como uma estrutura para a justificação na ética” (FREEMAN, 2007a, p. 17). Nesse acordo os indivíduos se empenham em justificar porquê e em quais limites são obrigados a cumprir normas e a se submeterem à dominação política. Como na tradição contratualista em geral os indivíduos se compreendem mutuamente como livres e iguais, a subordinação a uma autoridade política passa a ser concebida como resultado de um acordo feito pelos próprios indivíduos e com razões dadas por si mesmos, e não podendo mais ser justificada por entidades metafísicas. A partir do contrato, as pessoas podem responder à questão ‘Por que razão obedecemos a autoridade?’ de um modo bem simples: “obedecemos a autoridade porque nos comprometemos a isso” (GARGARELLA, 2008, p. 14). Esta afirmação nos leva a uma das restrições formais ao conceito de correto (*right*) impostas pelo contrato que é essencial para entendermos posteriormente os argumentos da estabilidade da teoria: as exigências do comprometimento (*the strains of commitment*). Ao avaliar o contrato em questão, pressupõe-se que cada uma das partes, com boa fé, se comprometa apenas com termos cujas consequências de sua aplicação possam honrar.

Importa acrescentar, no entanto, que há uma divergência no modo como os principais nomes da tradição contratualista interpretam o contrato. Hobbes, Locke, Rousseau e Kant, em comum, tomam como fundamental a ideia de que a cooperação social deve resultar em vantagem mútua dos membros da sociedade, no entanto, o contrato hobbesiano independe de exigências morais, são puramente baseados no interesse<sup>10</sup> (*purely*

---

<sup>10</sup> Esse termo, de autoria de Joshua Cohen, foi utilizado por Rawls nas *Conferências Sobre a História da Filosofia Política* para se referir às concepções de contrato tal qual o

*interest-based*). Este tipo de contrato se caracteriza por descrever as pessoas como fundamentalmente autocentradas, individualistas e abstraídas das relações sociais, o que implica no modo como se inserem na cooperação social: com vistas a atender apenas seus interesses e desejos (FREEMAN, 2007a, p. 18-20). Como os interesses básicos das pessoas não estão definidos em termos morais ou sociais, importando apenas a concepção do bem racional de cada indivíduo, esse modelo de contratualismo necessita de um Estado forte para fazê-las cumprirem com os termos de cooperação. A própria lei não engendra nos cidadãos o desejo de agir de acordo com ela por outro móbil que não o medo da punição.

Outra parte da tradição, representada por Locke, Rousseau e Kant, defende o que Freeman designou como ‘concepção de contrato baseada no direito’<sup>11</sup> (*right-based contract view*), que desenvolvem uma concepção de pessoa, de razão prática, e por isso do próprio contrato bem diferente da versão hobbesiana. As pessoas são descritas não como interessadas apenas na satisfação de seus interesses e desejos individuais, mas sim como possuindo “um interesse social fundamental (seu senso de justiça) em cooperar uns com os outros sob termos publicamente justificáveis que expressam sua concepção de si próprios como livres e iguais.” (FREEMAN, 2007a, p. 33). O contrato social deve, então, incorporar elementos morais das partes e ser feito em termos em que todos os cidadãos possam consentir. Ter em mente que é nessa vertente do contrato social a qual Rawls se filia, e não à hobbesiana, ajuda a compreender melhor o modo como concebe as pessoas e afastar as críticas de que as concebe como indivíduo isolado ou atomizado, além de ser essencial para compreender, mais adiante, o tipo de estabilidade que a teoria da justiça como equidade pretende alcançar.

A principal diferença entre a posição específica de Rawls e a tradição contratualista é o objeto do contrato social. Na justiça como equidade, o dispositivo do contrato não é precedido de um estado

---

formulado por Hobbes, que se fundamentam em uma consideração não moral do bem de um indivíduo. (FREEMAN, 2007a, p.18, n2)

<sup>11</sup> Freeman explica que usa o termo “right-based” não no sentido de direitos individuais, mas no sentido de princípios do direito (*right*) de Rawls. “Uma concepção moral que é baseada em uma concepção de direitos naturais ou individuais poderia ser baseada no direito (*right-based*), também, é uma posição como a de Rawls, a qual é baseada em outros princípios do direito.” (FREEMAN, 2007, p. 18, n.2)

de natureza pré-político e nem pretende chegar ao consenso sobre qual a forma de governo legítima, mas sobre “o conteúdo de uma concepção pública de justiça – os princípios e ideais da justiça e do bem comum – a ser aplicado à estrutura básica da sociedade como um sistema unificado de cooperação social” (CHFP, p. 18). Os princípios de justiça são contratados a partir da Posição Original, uma situação inicial de escolha puramente hipotética. A escolha por uma situação inicial hipotética tem consequências relevantes para a teoria da justiça como equidade. Significa que os princípios da justiça não podem estar subordinados as nossas circunstâncias particulares, nem a nossa desigual capacidade de negociação. Nos contratos do tipo hobbesianos, que se baseiam em indivíduos concretos concebidos como apenas auto interessados, os acordos “vão depender da capacidade de negociação – da força - de cada um dos participantes: os mais fortes, os mais talentosos, os mais poderosos obterão mais vantagens, e os mais desfavorecidos são os que ficarão pior” (GARGARELLA, 2008, p. 17). Por ser um acordo não empírico e altamente idealizado, as partes não são descritas como cidadãos empíricos, mas representantes de cada um dos cidadãos de uma sociedade bem-ordenada. Rawls adota uma concepção normativa de pessoa, ou seja, um ideal do que poderíamos realizar. São consideradas como “livres e iguais e possuem duas competências morais: (1) capacidade de ter senso de justiça e (2) capacidade de ter uma concepção de bem.

Rawls explica que a razão para utilizar o mecanismo do contrato na posição original repousa em sua correspondência com as características formais de uma sociedade bem-ordenada. (Cf. RAM, p. 250). Ou seja, é o contrato que incorpora na PO as características da SBO. Por isso, Freeman afirma que o modo como Rawls utiliza a teoria do contrato social tem muito mais a ver com a ideia de sociedade bem-ordenada do que a posição original:

É o acordo geral entre os membros de uma SBO que principalmente conduz o elemento contratualista na visão de Rawls, e não simplesmente a posição original. (...) O acordo na posição original é concebido para espelhar o acordo dos membros de uma sociedade bem-ordenada sobre princípios públicos de justiça. A escolha hipotética da posição original faz sentido apenas no contexto de pessoas livres e

iguais tentando chegar a um acordo social em termos de cooperação e padrões para uma justificação pública que todos possam aceitar, seja qual for situação (FREEMAN, 2007a, pp. 4; 35).

O contrato incorpora a concepção kantiana de pessoa, concebidas como, por natureza, livres iguais e racionais e que em uma SBO elas publicamente se concebem desse modo. As pessoas se concebem reciprocamente como racionais, livres e iguais e se consideram com direito de fazer reivindicações umas às outras.

## 2. A ESTABILIDADE EM UMA TEORIA DA JUSTIÇA: introdução ao problema.

O leitor mais familiarizado com as obras de Rawls pode questionar a pertinência de investigarmos o argumento da estabilidade em *Uma Teoria da Justiça* (1971) ao invés de focar a pesquisa no *Liberalismo Político* (1993), uma vez que é neste último livro que a questão assume o primeiro plano da reflexão filosófica e se torna verdadeiramente fundamental para a filosofia política, e também é nele que Rawls faz uma autocrítica apontando que parte do argumento da estabilidade desenvolvido na Parte III, da TJ, é, em certa medida, inconsistente com o restante da teoria.

Se a argumentação em favor da estabilidade na TJ é algo que fez com que Rawls ficasse tão insatisfeito a ponto de reformulá-la em obra posterior, por que insistir em entendê-la? Por que insistir em investigar o capítulo IX, que trata do argumento da congruência se é justamente este que o filósofo considerou inconsistente? Considerando todo o trabalho intelectual que Rawls empenhou para desenvolver sua teoria, não supomos razoável concordar com Brian Barry quando afirma – talvez até em tom provocativo – “que a única coisa a fazer [a respeito do Cap. IX] é seguir o curso seguido praticamente por todos os comentadores de *Uma Teoria da Justiça* e esquecê-lo” (BARRY, 1995, p. 915n54). Acreditamos que tal postura iria contra o próprio método interpretativo que Rawls utilizava para estudar os textos da história da filosofia e apresentá-los a seus alunos, o qual julgamos razoável seguir:

Sempre pressupus que os escritores que estudávamos eram muito mais espertos que eu. Se não fossem, por que eu desperdiçava meu tempo e o tempo de meus alunos estudando-os? Se via um erro em seus argumentos, supunha que estes escritores também o teriam visto e teriam por certo se ocupado dele. Mas onde? Eu procurava por sua saída, não pela minha. (...) Assim, aprendemos filosofia moral e política – ou, em verdade, qualquer parte da filosofia – estudando os modelos, aquelas individualidades insígnies que fizeram estimadas tentativas de filosofar; e, se temos sorte, encontramos um caminho para ir além deles (HFM, 2005, p. XVII-XVIII).

No caso, o próprio Rawls identificou qual seria o erro em seu argumento da estabilidade tal como formulado na obra de 1971 e se ocupou de resolvê-lo nos escritos posteriores. No presente trabalho, nossa intenção é, antes de tudo, compreender quais problemas Rawls estava tentando resolver quando propôs o argumento da estabilidade, e em especial, o argumento da congruência. Compreender o percurso percorrido para formular a pergunta é tão importante quanto compreender o trajeto da resposta. E aqui, novamente buscamos nos guiar por um conselho de Rawls a seus alunos, em 1964: não se pode perder de vista a intuição filosófica que conduziu o filósofo em sua investigação, pois

É fácil destruir um assunto (*subject*) demandando muito dele cedo demais; uma questão precisa ser guiada por grandes ideias intuitivas, particularmente no começo... É uma ilusão pensar que tenha grande valor uma análise rigorosa em uma pequena área que não seja guiada por uma grande ideia. Não se entende nem mesmo uma pequena coisa desta maneira. John Rawls, em 1964, a estudantes de seu curso de filosofia moral (REIDY, 2014, p. 9).

Nesse sentido, consideramos apropriado colocar a seguinte indagação: por que para Rawls foi necessário – no sentido mais preciso do termo – escrever a Parte III, e assim, aprofundar seu já extenso e denso livro? Caso Rawls tivesse encerrado a TJ com a Parte II, ainda assim teria feito uma extraordinária contribuição à Filosofia Política, no entanto, para ele, as Partes I e II não eram suficientes para dar conta do problema da justiça política em uma democracia constitucional. Isso faz com que nos questionemos qual a intuição, ou melhor, a ambição filosófica que o compeliu a escrever essa última parte do livro, a qual, por vezes, parece destoar das outras duas partes.<sup>12</sup> Um dos objetivos que esperamos cumprir com este trabalho é tornar tais intuições um pouco mais claras.

Importante ressaltar ainda que não nos aprofundaremos aqui nas diferenças entre *Uma Teoria da Justiça* e *Liberalismo Político*, entretanto, realçar algumas das semelhanças e diferenças entre tais livros nos auxilia a compreender melhor algumas das intuições

---

<sup>12</sup> “ (...) a intrincada estrutura final da teoria, com suas complicações labirínticas, obedece a tentativas desesperadas de salvar uma ideia deslumbrante – uma das mais belas ideias da teoria social e política (...)” (RODILLA, 2013, p. 46).

filosóficas que guiaram o autor. Apesar de possuírem objetivo primário diferente, essas obras fazem parte de um propósito comum – ainda que alcançado de maneiras diferentes. Ambos têm o interesse em desenvolver uma teoria da justiça para uma sociedade democrática, cuja “concepção de justiça basicamente dependa da prática institucional” (MAFFETTONE, 2010, p. 18). A despeito das mudanças entre um livro e outro, a estrutura do pensamento rawlsiano não foi fundamentalmente modificada e pode-se notar que o fio condutor presente desde os primeiros textos de Rawls é a prioridade do correto e complexa relação entre o correto e o bem (Cf. MAFFETTONE, 2004, p. 545-546), ou, melhor dizendo, entre o correto e as diversas ideias de bem. Há uma continuidade<sup>13</sup>, ainda que o problema primário de cada um dos dois livros tenha se modificado. Enquanto que TJ está definindo o conteúdo da justiça social a partir da teoria contratualista em contraposição ao utilitarismo o PL trata do significado e valor da justificação política que abrange limites e possibilidades da democracia liberal (Cf. MAFFETTONE, 2004, p. 541).

Rawls concebe a justificação como sendo “uma argumentação dirigida àqueles que discordam de nós, ou a nós mesmos, quando estamos indecisos” (TJ, §87, p. 716). Tendo essa definição em vista, outra diferença que podemos levantar entre TJ e PL diz respeito a qual audiência cada um desses livros está oferecendo razões. No livro de 1971, em que pese se esteja elaborando uma teoria da justiça para uma sociedade democrática, há um importante sentido no qual sua audiência não é – em primeira instância – de cidadãos em geral, mas sim de filósofos, economistas e matemáticos<sup>14</sup> atraídos pelo sistema formal do utilitarismo. É este público especializado que Rawls tem de primeiramente persuadir para que

---

<sup>13</sup> De acordo com Maffettone, há três pontos relevantes em que se percebe a continuidade entre TJ e LP: “[a] a justificação de uma concepção política não pode coincidir com o conteúdo de uma teoria moral; [b] a validade de uma concepção política, ao contrário, depende profundamente da [c] legitimidade de seu suporte (*background*) institucional (principal tema do LP). Mas, por outro lado, a legitimidade não é independente de uma teoria política liberal específica tal como a justiça como equidade (o conteúdo da TJ). TJ repousa no LP e vice-versa, em um espírito que pode ser chamado de institucionalismo normativo” (MAFFETTONE, 2010, p.17).

<sup>14</sup> “TJ está preocupada em oferecer uma alternativa sistemática ao utilitarismo, em especial ao aparato formal de uma longa linhagem de utilitaristas matemáticos e economistas, bem como filósofos morais a desenvolveram” (LADEN, 2014, p. 67).

sua alternativa seja levada a sério, por isto, busca mostrar que sua teoria pode ser articulada no mesmo nível de precisão formal e sistematicidade que o é a teoria preferida de sua audiência. Em razão disso Rawls, adota uma linguagem mais técnica no tratamento dos seus argumentos em seu primeiro livro (Cf. LADEN, 2014, p. 67).

Audiência à qual se dirige o PL são os que “podem não estar convencidos da superioridade intelectual da teoria da justiça como equidade” (MAFFETTONE, 2010, p. 220), os cidadãos de fé e os cidadãos que perderam fé na possibilidade de uma democracia constitucional (CF. LADEN, 2003, 380-383). Neste livro fica mais clara a ideia da filosofia como defesa em vez de mera teorização filosófica, a defesa da possibilidade de um regime constitucional justo (cf. PL, III, §2.2, p. 120).

Em que pese possamos fazer uma leitura que endosse a continuidade entre TJ e PL, não podemos negar que há, algumas discontinuidades específicas entre um livro e outro, sendo a estabilidade uma delas. O modo como a pergunta é reformulada no segundo livro ensejou o desenvolvimento de três conceitos ausentes no livro de 1971: consenso sobreposto, concepção política de pessoa e a ideia de Razão Pública. Não nos deteremos na explicação destes conceitos.

Um dos objetivos da justiça como equidade é estabelecer termos equitativos para a partir dos quais serão atribuídos direitos e partilhados os frutos da cooperação social. O papel distributivo de uma concepção de justiça, entretanto, não pode ser a única métrica utilizada para avaliá-la, é preciso, segundo Rawls, considerar como a aplicação da concepção influi em problemas sociais fundamentais como os da eficiência, coordenação e estabilidade (cf. TJ, §1, p. 7). Na Parte III de TJ, Rawls concentra seu esforço em mostrar que uma sociedade bem-ordenada cuja estrutura satisfaça os princípios da justiça como equidade é relativamente mais estável que uma sociedade bem-ordenada pelo princípio da utilidade.

A justificação da concepção de justiça mais apropriada para orientar os termos de cooperação de uma sociedade bem-ordenada é feita a partir da posição original, na qual a argumentação em favor da concepção de justiça a ser escolhida é feita em duas etapas. Na primeira, princípios são provisoriamente escolhidos a partir da perspectiva pública e imparcial da justiça, todas informações particulares e circunstanciais sobre as partes são abstraídas através

do expediente do véu da ignorância. Já na segunda etapa ocorre a confirmação (ou não) dos princípios previamente selecionados na etapa anterior, mas agora a partir da perspectiva individual, isto é, confere-se se a concepção adotada é viável e se é a mais estável dentre as concepções. As partes devem verificar se, aplicados em uma sociedade bem-ordenada, tais princípios são mais capazes de garantir por si sua estabilidade do que outra concepção de justiça, isto é, se as exigências dos princípios de justiça ensejam seu cumprimento espontâneo ou se há necessidade de coação. “Permanecendo constantes os demais fatores, a concepção de justiça preferida é a mais estável” (TJ, §76, p. 615). Portanto, a escolha dos princípios e o raciocínio iniciados na PO só se completam ao final desta segunda etapa, depois analisada a estabilidade. Este segundo estágio da argumentação trata da questão motivacional, isto é, porque e em que condições os membros de uma sociedade em circunstâncias favoráveis dariam sua adesão unânime e espontaneamente agiriam de acordo com a concepção de justiça.

Podemos qualificar dois tipos de questões que o foco na estabilidade em TJ procura responder. A primeira é mais geral, está relacionada com o papel da estabilidade para justificar princípios morais e consiste em saber se a SBO correspondente a uma concepção particular de justiça é estável, ou mais ou menos estável relativamente a outras concepções. A segunda questão, mais específica, quer mostrar que fazer regularmente o que a justiça como equidade requer de nós está de acordo com capacidades humanas e sentimentos morais, e que ter um senso de justiça não é autodestrutivo e não enfraquece nossa busca por bens importantes, pelo contrário, é compatível com nosso bem (Cf. FREEMAN, 2007a, p. 12). Antes de discutirmos os argumentos que Rawls apresenta em favor de sua concepção de justiça como equidade, convém que esclareçamos em que sentido é tratada a estabilidade.

## **2.1 Que tipo de estabilidade?**

Na introdução ao *Liberalismo Político*, Rawls afirma que, apesar de ser fundamental para a filosofia política, “o problema da estabilidade desempenhou um papel muito pouco relevante na história da filosofia moral” (PL, p. XVIII). Num primeiro momento,

essa assertiva pode causar estranheza ao leitor, pois se pensarmos o problema da estabilidade como o problema da ordem, tal como sugere Brian Barry, veremos que desde Platão, praticamente todos os filósofos que pensaram a política estavam preocupados em como manter a ordem da sociedade política, evitar sua dissolução e garantir sua continuidade ao longo do tempo. Esse empenho é ainda mais claro nos tratados contratualistas que concebem a manutenção da ordem social como objeto primeiro da justiça política e buscam mantê-la mesmo ao custo do sacrifício das liberdades individuais. Uma das contribuições mais originais da teoria de Rawls diz respeito justamente ao modo como concebe a estabilidade, por isso essa dificuldade inicial em compreender como tal conceito desempenhou papel irrelevante na história. Ele transformou “uma noção de estabilidade tipicamente psicológica, sociológica e institucional em uma noção fundamentalmente moral” (MAFFETTONE, 2010, p. 252). A afirmação de Rawls inicialmente mencionada passa a fazer sentido quando esclarecemos que ele compreende a estabilidade não como o objeto principal da justiça, tal como Hobbes, mas sim enquanto atributo de uma sociedade justa, ou melhor, “enquanto consequência de uma atitude moral e psicológica, que expressa a relação ética entre cidadãos e sociedades justas” (MAFFETTONE, 2010, p. 252). Como expresso na frase que inaugura *Uma Teoria da Justiça*, “a virtude primeira das instituições sociais” é a justiça, e não a estabilidade. Assim é que o foco da teoria de Rawls não é a estabilidade das sociedades em geral ou de um sistema de instituições (JF, §54, p. 258, n.1), mas a estabilidade da concepção de justiça. Nesse sentido que Marta Nussbaum afirma que a estabilidade é

(...) uma noção moral: ela envolve não apenas a persistência de um conjunto de arranjos políticos, mas a persistência de um respeito incorporado tanto nas instituições quanto nas atitudes dos cidadãos que as suportam (NUSSBAUM, 2015, p. 5).

Antes de Rawls, o termo ‘estabilidade’ na filosofia e ciência política era utilizado para indicar regimes ou sistemas estáveis como sendo aqueles “caracterizados pela lei e ordem, por instituições que funcionem regularmente” (BARRY, 1995, p. 880; KLOSKO, 1994, p. 1884), isto é, aqueles regimes em que durante um longo período de tempo não ocorre nenhuma mudança significativa

na estrutura do governo e que as leis são obedecidas com regularidade por uma ampla parcela da população. Uma outra definição genérica de estabilidade pode ser dada nos seguintes termos:

(...) é a capacidade previsível que um sistema tem de se prolongar no tempo. Deste modo, um sistema deve ser considerado estável quando, num momento dado, tendo em conta um conjunto de sinais, é razoavelmente previsível que ele continuará no tempo (BOBBIO, 1998, p. 394).

A explicação segue no sentido de que a continuidade no tempo implica capacidade do sistema em ajustar-se aos desafios e demandas que advêm de uma realidade que está em constante mudança e movimento. Como consequência, a estabilidade, entendida nestes termos, não implica nem que o sistema seja imóvel, nem estático, pois se o sistema permanece estável é “precisamente porque é capaz de se adaptar aos desafios que vêm do ambiente, de se modificar de forma autônoma” (BOBBIO, 1998, p. 395).

O principal problema filosófico desses modos de conceber a estabilidade é que ela não tem compromisso algum com a justiça, pois pode ser obtida mesmo que sejam injustas as leis que os cidadãos obedecem e as instituições que regulam a sociedade. (WEITHMAN, 2010 p. 44). Recordemos, por exemplo, de alguns regimes políticos em nossa história recente, como a Ditadura Civil-Militar imposta ao Brasil entre 1964 e 1989, o Estado Nazista ou a ex-URSS, que mantiveram a ordem social por um período de tempo, contando com a obediência regular da população, e nem por isso podemos afirmar que eram sociedades estavelmente justas. Também a literatura, com o *Admirável Mundo Novo*, de Aldous Huxley, e *1984*, de George Orwell, nos traz exemplos de sociedades que podem ser consideradas estáveis se tomarmos como critério apenas obrigação civil e a permanência e regularidade de instituições ao longo do tempo. Entretanto, não é esse tipo de estabilidade descompromissada com uma concepção de justiça que pessoas que se compreendam reciprocamente como morais, livres e iguais possam endossar que Rawls está buscando justificar. “Uma vez que uma sociedade justa tenha liberdade civil e participação democrática no governo, não pode ser mantida por métodos

Hobbesianos” (BARRY, 1995, p. 881).

Rawls passa a tratar da estabilidade de uma concepção moral de justiça aplicada a uma sociedade bem-ordenada cujas instituições são capazes de gerar o próprio apoio. A concepção de justiça é estável, no sentido rawlsiano, se ao ser aplicada na estrutura básica de uma sociedade bem-ordenada, as instituições sociais, políticas e econômicas são capazes de gerar nas pessoas que participam desses arranjos o correspondente senso de justiça e o desejo de preservar tais instituições justas. Uma sociedade justa ‘inerentemente’, ou ‘intrinsecamente’ estável (TJ, §76, p. 614; SJ, p. 106), capaz de gerar as forças que garantam espontaneamente seu próprio apoio, é relativamente mais estável, e, portanto, preferível, a que uma sociedade cuja estabilidade foi adquirida através da força ou pela doutrinação coercitiva. A estabilidade inerente implica que cada cidadão “poderia voluntariamente fazer e saber o que fazer o que fosse necessário para preservar o seu senso de justiça, de modo que a justiça de uma SBO ao longo do tempo poderia ser mantida pela atividade autônoma de seus membros” (WEITHMAN, 2010, p. 56).

Afirmar que uma concepção de justiça seja estável não quer dizer as instituições permanecerão estagnadas ao longo do tempo, refratárias às novas circunstâncias e demandas da vida social ou mesmo imunes a praticarem em graves abusos. Implica sim, que as instituições são capazes de se remodelarem de acordo com reivindicações de justiça dos cidadãos, e caso venham a se tornar injustas e abusivas, que existam forças estabilizadoras que evitam outras violações e tendam a restaurar a ordem (TJ, §1, p. 7). Se as forças estabilizadoras tiverem origem no interior do sistema de cooperação, temos uma sociedade inerentemente justa, e caso forem externas a tal sistema, a estabilidade é imposta (Cf. WEITHMAN, 2010). Num esquema de cooperação social estável, para que as normas básicas sejam cumpridas voluntariamente “as pessoas precisam ter a garantia que todos os demais têm razão suficiente para agir de acordo com as regras” (FREEMAN, 2007a, p. 145n6). Se a concepção de justiça for tal que as instituições sociais, econômicas e políticas não conseguirem formar em seus cidadãos um desejo efetivo de agir de modo justo e apoiar as instituições, é necessário acrescentar um mecanismo externo ao sistema social para garantir o cumprimento da lei, tal como a figura do soberano

de Hobbes (TJ, §76, p. 613). O senso de justiça que as instituições hobbessianas formam nos cidadãos é tal que exige um sistema punitivo cuja finalidade exclusiva é impor a obediência através do medo e da punição, sem tais expedientes, não há garantia alguma de que indivíduos cumpram as leis. Como são as próprias instituições da estrutura básica – o soberano e o sistema penal – que mantêm a estabilidade da sociedade, parece natural concluir que se trata de um caso de uma sociedade inerentemente estável, mas não é o que ocorre. Paul Weithman lembra que o soberano de Hobbes não faz parte da cooperação social entre os indivíduos, por esta razão, não está submetido nem ao esquema de recompensas, ou melhor, à distribuição das vantagens mútuas advindas da cooperação, não aos termos que regulam esse esquema, mas está acima dos sujeitos e é quem tem poder para alterar, arbitrariamente, os termos de cooperação usando ameaças e coerção, implicando que a estabilidade dos termos de cooperação social seja imposta (Cf. WEITHMAN, p. 50).

O modo hobbessiano como a concepção de justiça é estabilizada não seria aceito pelos membros de uma SBO, uma vez que esses cidadãos, por se compreenderem como iguais, não aceitariam se submeter a uma dominação política arbitrária. A preocupação de Rawls é que as ameaças à estabilidade possam ser removidas por forças que estejam no interior do sistema de cooperação, dispostas a “preservar a justiça do arranjo como um todo” (TJ, §35, p. 270).

Rawls aposta que se a concepção de justiça for capaz de gerar em seus membros um forte senso de justiça – que é o desejo de agir de acordo com a justiça e apoiar instituições justas - permanecerá estável mesmo quando o sistema não for perfeito (TJ, §69, p. 565-6). Nesse caso, os desvios da justiça seriam corrigidos por forças de dentro do sistema: os próprios cidadãos movidos pelo seu senso de justiça contribuiriam para restaurar o equilíbrio das instituições. Se possuírem apreço sincero pelas instituições justas sob as quais vivem e se espontaneamente afirmam seu desejo de agir com justiça, tais cidadãos, presume-se, manifestar-se-iam também contrários aos atos de injustiça e às instituições injustas. Dois exemplos de mecanismos inerentemente estabilizadores da sociedade são a desobediência civil e objeção de consciência, os quais, se empregados com a devida limitação e critério justo,

ajudam a manter e fortalecer as instituições justas. O que acrescenta estabilidade à SBO, nesse caso, é uma disposição geral dos cidadãos a se engajarem na desobediência civil quando ela for justificada (Cf. TJ, §59, p. 477). Nesse sentido, a desobediência civil ilustra o que ocorre quando há uma sociedade bem-ordenada com cidadãos engajados e comprometidos com instituições justas. Assim fica mais clara a razão de a estabilidade ser concebida como consequência de uma atitude moral e psicológica, e não como uma característica estrutural de sistemas políticos (Cf. MAFFETTONE, 2004, p. 552).

Outro aspecto importante é a estabilidade relativa de uma concepção moral. Lembremos que na segunda etapa da escolha dos princípios na PO, as partes precisam testar os princípios constantes da lista inicial na SBO, isto é, dadas as exigências que o compromisso com tais princípios exige das pessoas em uma sociedade bem-ordenada, as partes verificam se é possível se comprometer a agir de acordo com eles. As partes vão comparar o resultado da aplicação das concepções de justiça e escolherão aquela que for mais estável em comparação com as outras. O objetivo é conferir se a concepção já adotada é viável e se não é tão instável que qualquer outra poderia ser melhor. (TJ, §76, p. 622). Podemos afirmar que a estabilidade relativa resulta da aplicação do método que Rawls defende em teoria moral: “o estudo comparativo das sociedades bem-ordenadas” (IMT, 1975, p. 294), através do qual argumenta que a justiça como equidade é superior ao princípio da utilidade média em relação a sua capacidade de garantir a estabilidade inerente.

Com os argumentos de que uma sociedade regulada pela justiça como equidade é mais estável do que se o fosse pelo princípio da utilidade, Rawls está defendendo que o ideal de sociedade justa apresentado seja de algum modo praticável. Ele tem a ambição de que a justiça como equidade ofereça um ideal normativo, uma meta que possa orientar o curso de uma reforma social (TJ, §39, p. 304) e de uma reforma do sistema legal (*legal reform*) (TJ, §38, p. 299). Reconhece, entretanto, que esse estado idealmente justo é muito difícil de ser alcançado. E ainda que uma tal sociedade nunca se concretize, permanece como uma grande contribuição a justificativa da *possibilidade* de existir uma ordem social justa e duradoura. Não se trata de “uma mera possibilidade lógica, mas uma possibilidade que se liga às tendências e inclinações profundas do mundo social” (DP, §18, p. 168).

“Assim como Kant, Rawls pensou que uma tarefa importante, mas difícil para a filosofia é estabelecer a possibilidade de várias crenças fundamentais e esperar que elas sejam racionalmente sustentáveis” (HILL Jr., 2014, p. 201). Na Terceira Antinomia da Razão Pura, Kant não prova - e nem tem esta pretensão - que a liberdade de fato existe, mas sim que existe um fundamento racional que nos permite pensá-la como possível. Analogamente, esse parece ser o intuito de Rawls: mostrar que é possível justificar racionalmente, com base em argumentos que cidadãos democráticos possam razoavelmente concordar, a existência de uma sociedade justa e estável. É possível, até mesmo provável, que uma sociedade perfeitamente justa como a descrita por Rawls nunca se concretize, ainda assim, a mera possibilidade racional de sua existência é importante por

mostrar que há razões suficientes para qualquer cidadão razoável e racional livremente endossar a justiça como equidade, por isso, satisfaz uma exigência liberal e kantiana para a plena justificação do uso do poder coercitivo do estado. (...) [Rawls] pensa que qualquer uso coercitivo do poder de acordo com os seus princípios perdem sua justificação plena se sua estabilidade pelas razões corretas não for, pelo menos, possível nesse sentido (HILL Jr., 2014, p. 201).

Rawls divide a teoria da justiça em duas partes, ideal e não-ideal, e logo no início de TJ informa o leitor que sua investigação se restringe apenas à teoria ideal, comprometida em responder quais princípios são os ideais para uma sociedade em circunstâncias ideais, ou seja, pretende formular uma concepção de justiça a partir de e para ser aplicada em uma sociedade idealmente justa. Uma sociedade perfeitamente justa tem pelo menos três pressuposições que permitem pensar nela como idealizada: (i) obediência estrita, (ii) circunstâncias favoráveis, e (iii) ausência de limitações naturais (Cf. STEMPOWSKA & SWIFT, 2014, p. 113-116). A teoria ideal pressupõe que praticamente todos os cidadãos da sociedade perfeitamente justa irão se comprometer em agir de acordo com as exigências dos princípios de justiça, obedecer estritamente às leis e apoiar as instituições. As partes na posição original devem levar em conta a condição da obediência estrita e escolher princípios cujas

exigências os cidadãos da SBO possam cumprir espontaneamente sem que isso lhe implique alguma privação. Injustiças decorrentes da desobediência dos cidadãos,<sup>15</sup> acidentes naturais e limitações da vida humana ocorrem mesmo sob circunstâncias favoráveis, pois no mais das vezes estão ligadas a contingências históricas e sociais. Estes casos devem ser resolvidos pela teoria não-ideal, a partir da métrica fornecida pela teoria ideal.

Se a justiça não for possível nem mesmo do ponto de vista ideal, se não for possível sequer justificar racionalmente a estabilidade em uma sociedade pensada nas melhores condições possíveis, então a justiça não resiste em lugar algum. Nesse caso, “talvez devêssemos indagar, com Kant, se vale a pena para os seres humanos viverem na Terra” (PL, p. LXIX).

Ainda que uma sociedade perfeitamente justa seja um ideal irrealizável, levar a cabo, filosoficamente, a tarefa de construí-la ilumina o modo sobre como podemos agir no mundo. Enquanto acreditarmos,

(...) por boas razões, que é possível uma ordem política e social razoavelmente justa e capaz de sustentar a si mesma, dentro do país e no exterior, poderemos ter esperança razoável de que nós ou outros, algum dia, em algum lugar, a conquistaremos; podemos, então, fazer algo por essa conquista. Apenas isso, deixando de lado o sucesso ou o fracasso, é suficiente para eliminar os perigos da resignação e incredulidade. (...) ao trabalhar rumo [a um objetivo de esforço político de longo prazo] dá significado ao que podemos fazer hoje (DP, §18, 168).

Para alcançar esse fim, mesmo que nunca venha a se concretizar, a concepção moral de justiça deve justificar seus argumentos como *se fosse possível* realizá-la, o que implica considerar que qualquer que seja a concepção de justiça ela tem de poder ser alcançada por seres humanos, ou seja, não pode

---

<sup>15</sup> Quando as pessoas deixam de obedecer às leis e instituições que aplicam os princípios de justiça, estamos diante de casos de injustiça causados pela desobediência parcial. Lidar teoricamente com esse tipo de injustiça é tarefa da teoria não-ideal, a qual “abrange temas como a teoria do direito penal, a doutrina da guerra justa e a justificação das diversas modalidades de oposição a regimes injustos, da desobediência civil e da objeção de consciência à militância de resistência e à revolução civil” (TJ, §2, p. 10).

demandar o impossível da natureza humana, não pode exigir das pessoas que elas ajam para além dos limites suportáveis (Cf. FREEMAN, 2007a, p. 101), não pode exigir dos indivíduos, em nome da justiça, o auto sacrifício, motivações heroicas, nem a disposição para que sejam mártires.<sup>16</sup>

## 2.2 Ameaças à estabilidade

Vimos na seção anterior que o tipo de estabilidade que importa à teoria da justiça é a estabilidade inerente, e não imposta pela força de mecanismos coercitivos. O desafio teórico de Rawls para justificar como é possível a estabilidade inerente é mostrar como surgem e como podem ser mantidas na sociedade as “forças estabilizadoras” internas, as quais nada mais são que um senso de justiça dos cidadãos da sociedade bem-ordenada forte o suficiente para neutralizar as principais ameaças à estabilidade.

A estabilidade da concepção depende de um equilíbrio de motivações: o senso de justiça que cultiva e os objetivos que incentiva devem normalmente ter preponderância sobre às propensões à injustiça (TJ, §69, p. 561).

Tanto mais estável será uma concepção de justiça quanto mais forte for o senso de justiça gerado e a tendência para anular inclinações desestabilizadoras.

Os argumentos apresentados na TJ para justificar a estabilidade da concepção de justiça, a aquisição e desenvolvimento do senso de justiça e o argumento da congruência, além de terem por objetivo relacionar a justiça com a natureza humana e com o bem humano, são estratégias para responder à alguns tipos de ameaças à estabilidade. Mesmo que a concepção de justiça seja tal que não demande compromissos que estão para além das capacidades humanas, existem tendências dos indivíduos que podem conduzir à instabilidade.

---

<sup>16</sup> Rawls não nega que atos de misericórdia, heroísmo e auto-sacrifício (atos supererrogatórios/benevolentes) sejam valorosos. “É bom praticá-los, mas não constituem obrigação nem dever para com ninguém” (TJ, §19, p. 140). Ou seja, podem ser atos moralmente bons, mas não podem ser a base de um dever de justiça. Nesse sentido, o utilitarismo se torna bastante problemático, pois de acordo com tal teoria “parece que somos obrigados a praticar atos que promovem o bem de outrem, seja qual for o custo para nós mesmos, contanto que o total das vantagens produzidas exceda o total das vantagens produzidas por outros atos que não são acessíveis” (TJ, §19, p. 140).

“Essas tendências, que têm origem no egoísmo, conduzem à primeira espécie de instabilidade” (TJ, §51, p. 419). Aqui Rawls se referindo ao problema da perspectiva do autointeresse<sup>17</sup>. Tal problema surge quando a pessoa se beneficia dos bens públicos (frutos da cooperação social), mas se sente tentada a não dar a sua contrapartida devida (por exemplo, pagar corretamente os impostos) quando houver ocasião em que julgue que seu interesse particular seria melhor atendido. Ao agir desse modo, o indivíduo estará violando o princípio da equidade<sup>18</sup>, segundo o qual “não devemos nos beneficiar dos esforços cooperativos das outras pessoas sem fazer a parte que nos cabe” (TJ, §52, p. 428). Contanto que a instituição seja justa, a pessoa tem a obrigação de cumprir com a sua parte.

Quando o indivíduo cede à tentação de se eximir de fazer a sua parte, gera a segunda ameaça à instabilidade, que Rawls denominou como o problema da garantia. Todos querem uma garantia de que o acordo coletivo para distribuir bens de modo justo “será honrado se cada um se dispuser a fazer a sua parte” (TJ, §42, p. 334). Os indivíduos somente cumprirão “com as exigências do empreendimento cooperativo se acreditarem que os outros também farão a parte deles” (TJ, §51, p. 419). Na medida em que algumas pessoas começam a agir como um carona (*free-rider*), beneficiando-se dos bens públicos sem dar sua contribuição, aumenta o sentimento de desconfiança entre os cidadãos, e os laços de confiança mútua se tornam frágeis. Por não terem confiança que seus concidadãos irão cooperar, os demais cidadãos também deixam de fazer a sua parte.

De acordo com Rawls, uma vez que se consiga resolver o problema da perspectiva do autointeresse, o problema da garantia e

---

<sup>17</sup> Embora esta fonte de instabilidade tenha sua origem na disposição ao egoísmo, “é importante não confundir esse problema com o de justificar para um egoísta por que se deve ser uma pessoa justa” (TJ, §86, p. 700). O que Rawls buscará mostrar, com o argumento da congruência, para solucionar essa questão é que agir a partir do ponto de vista da justiça é um bem que é racional desejar. E que, então, esse bem da justiça pode, não apenas ser parte dos fins, mas é regulador dos fins. E que, enquanto sentimento regulador dos outros fins, a justiça continua sendo um bem. (Cf. TJ, §86).

<sup>18</sup> Na posição original, as partes irão escolher um sistema completo de princípios, isto é, (i) princípios aplicáveis à instituições, (ii) princípios aplicáveis à indivíduos e (iii) regras de prioridade. Na concepção da justiça como equidade, os princípios para indivíduos que fazem parte deste esquema são o princípio de equidade e o de respeito/auxílio mútuo (cf. TJ, §51).

confiança-mútua também será solucionado.

A questão que Rawls busca responder é por quais motivos os cidadãos de uma SBO adotariam e aceitariam que suas instituições fossem guiadas por aqueles princípios escolhidos do ponto de vista público da PO? Por que aceitariam agir de acordo com os princípios escolhidos por seus representantes? A estratégia de argumentação em favor da estabilidade da justiça como equidade é dividida em duas partes. A primeira delas, desenvolvida no Capítulo VIII de TJ, desenvolve uma psicologia moral concebida para mostrar como os princípios da justiça como equidade são compatíveis com a natureza humana, e, por conseguinte, como os membros de uma SBO pela justiça como equidade podem vir a adquirir o senso de justiça, definido como o desejo normalmente efetivo de aplicar e agir de acordo com as exigências dos princípios de justiça e apoiar as instituições que os aplicam.

A segunda parte, referente ao Capítulo IX, pretende mostrar como o senso de justiça desenvolvido em uma SBO equitativa não só é compatível como constitui uma parte do bem de uma pessoa.

Por fim, para que a concepção de justiça obtenha apoio e adesão dos seus cidadãos, é preciso mostrar que uma sociedade justa é um bem para cada um dos indivíduos a tal ponto que o desejo de viver sob a condição da justiça é mais forte que qualquer outro. Por isso Rawls argumenta que uma sociedade política justa é um bem para a pessoa, tanto do ponto de vista individual, pois permite expressar sua natureza livre e igual, além de que o exercício das capacidades morais é experimentado como algo bom congruente com o bem de cada indivíduo, quanto do ponto de vista coletivo.



### **3. O PROBLEMA DA CONGRUÊNCIA ENTRE CORRETO E O BEM DO PONTO DE VISTA DA ESTRUTURA DA TEORIA MORAL.**

O objetivo de *Uma Teoria da Justiça* é apresentar uma teoria moral da justiça cuja estrutura represente uma alternativa às teorias até então dominantes na tradição filosófica. O utilitarismo e o perfeccionismo estruturam seus conceitos morais no padrão teleológico, considerado por Rawls como radicalmente equivocado por relacionar o correto ao bem de maneira indevida: tornam o correto um mero instrumento da realização do bem, definindo-o sem referências a princípios ou conceitos morais anteriores. Inspirado no padrão deontológico, Rawls desenvolve uma estrutura moral oposta à teleológica: formula primeiro uma teoria do correto, esboça uma teoria do bem e justifica a prioridade da primeira sobre a segunda. Defender a prioridade do correto não implica negar a ideia de bem ou que haja incompatibilidade entre correto e o bem.<sup>19</sup> Por isso, uma vez justificada a prioridade do correto sobre o bem, Rawls procura mostrar como combinar o correto com o bem “a fim de que não fiquem separadas a justiça e a utilidade” (ROUSSEU, *Do Contrato Social*, Livro Primeiro).<sup>20</sup>

Com vistas a compreender algumas das intuições filosóficas que conduziram Rawls ao problema da estabilidade e ao argumento da congruência é relevante que os localizemos na estrutura argumentativa da teoria da justiça como equidade. Para tanto, mostraremos primeiro a importância que Rawls atribui ao esforço de se evidenciar e compreender a estrutura das concepções morais e sua conexão com a sensibilidade humana.

#### **3.1 A justiça como equidade enquanto parte da Teoria Moral.**

O projeto da justiça como equidade foi desenvolvido em TJ com o objetivo de formular as características estruturais de uma

---

<sup>19</sup> Nenhuma concepção de justiça pode se basear inteiramente no correto ou no bem. Toda concepção de justiça deve, isto sim, combinar ambos os tipos de ideias de forma específica. A prioridade do correto não nega isso (LP, V, p. 205).

<sup>20</sup> “Quero indagar se pode existir, na ordem civil, alguma regra de administração legítima e segura, tomando os homens como são e as leis como podem ser. Esforçar-me-ei sempre, nessa procura, para unir o que o direito permite ao que o interesse prescreve, a fim de que não fiquem separadas a justiça e a utilidade”(ROUSSEAU, *Contrato Social*, p.21).

concepção de justiça capaz de fornecer o fundamento moral mais apropriado para uma sociedade democrática. Foi pensada como parte da filosofia moral na esteira da tradição do contrato social, sem traçar qualquer distinção inicial com a filosofia política. Décadas depois da publicação de TJ, Rawls reposiciona seu objetivo com vistas a fornecer um fundamento estritamente político para a justiça. A separação da filosofia política de uma abordagem estritamente ética/moral começa a ser delineada nos textos rawlsianos publicados na década de 1980. É neste período que o filósofo passa a fazer o deslocamento de ênfase de sua teoria, de moral para política.

Embora em raras passagens do livro a concepção de justiça desenvolvida seja designada como política (como, por exemplo, no último parágrafo do prefácio da edição revista, datado de 1990), o fato é que a distinção entre uma concepção estritamente política da justiça e uma doutrina moral ou filosófica abrangente não é feita em TJ, o que levou Rawls a afirmar, em 1992, que “fica claro que o livro [TJ] trata a teoria da justiça como equidade e o utilitarismo como doutrinas mais ou menos abrangentes” (RAWLS, 1992, p. IX).

Não pretendemos aqui tratar da diferença entre concepção política de justiça e doutrina moral abrangente<sup>21</sup>, nem a distinção entre filosofia moral e filosofia política, mas deixar marcado que a abordagem da justiça como equidade formulada em TJ foi desenvolvida do ponto de vista da Teoria Moral concebida como parte da Filosofia Moral. Portanto, com vistas a seguir a intuição inicial do filósofo, é a partir desta perspectiva moral – uma vez que o campo do político ainda não está claramente delineado em TJ – que buscaremos compreender os argumentos, questões e dificuldades ali apresentadas.

Rawls defende que a filosofia moral deve ser compreendida como independente (*independent*) de outros ramos da filosofia.<sup>22</sup>

---

<sup>21</sup> Uma doutrina é abrangente quando “inclui concepções do que se considera valioso na vida humana, bem como ideais de virtude e caráter pessoais, que devem conformar grande parte de nossa conduta não política (no limite, nossa vida como um todo)” (p. 206). Já uma concepção política de justiça é uma “concepção moral formulada para um objeto específico, a saber, a estrutura básica de um regime constitucional”, que não é formulada com base em qualquer doutrina abrangente e não pressupõe o endosso a qualquer doutrina religiosa, filosófica ou moral específica (Cf. LP, V, §, 205-206).

<sup>22</sup> No LP, Rawls usa o termo ‘*free standing*’ para indicar o “status epistemológico” da concepção política de justiça: é uma teoria *free-standing*, isto é, que se sustenta por si própria, a partir de pressupostos próprios, não é derivada de uma doutrina abrangente e

‘Independente’ não deve ser aqui tomado num sentido muito restrito, pois não se quer dizer que a filosofia moral é isolada ou que não guarda certa relação de dependência mútua com partes da filosofia, mas sim que não está em uma relação de subordinação ou prioridade metodológica em relação aos outros campos da filosofia, tais como filosofia da mente, epistemologia e teoria da significação (Cf. IMT, p. 302). A filosofia moral, assim como outros ramos da filosofia, tem problemas e assuntos específicos, cuja investigação deve ser feita a partir de seu próprio ponto de vista, e, portanto, em muitos aspectos será independente do desenvolvimento de outras áreas da filosofia.<sup>23</sup> O filósofo não está afirmando que o domínio da Filosofia Moral é inteiramente separado de outros campos do saber – aliás, como veremos adiante, reconhece que há uma relação de dependência da filosofia moral com a psicologia e teoria social. Mas que existe um campo específico da Filosofia Moral que é reconhecido em razão, principalmente, dos problemas próprios e do modo específico de tratá-los. Rawls defende que domínio próprio da Filosofia Moral diz respeito à compreensão da estrutura das concepções morais e das suas conexões com a sensibilidade humana. (IMT, p. 302). O aprimoramento da Filosofia Moral depende, então, de que essa tarefa seja realizada, por um campo específico da própria Filosofia Moral: a Teoria Moral.<sup>24</sup>

Enquanto disciplina da filosofia moral, a Teoria Moral é responsável por realizar o estudo substantivo das doutrinas morais, o que implica em (i) fazer uma análise de como as noções morais

---

nem é parte de uma (Cf. LP, I, §1.2, p. 14). Supomos que o termo ‘*independent*’ usado por Rawls para definir o status da Filosofia Moral pode ser compreendido no sentido de ‘*free-standing*’ para definir a relação da filosofia moral em relação a outras áreas da filosofia. É independente no sentido de que possui problemas específicos e métodos específicos e cujo avanço da disciplina depende muito mais do estudo e desenvolvimento dessas questões e métodos do que do desenvolvimento de outros ramos da filosofia. Fazer comparações sistemáticas entre as concepções morais “é uma tarefa que devemos ser capazes de realizar, e a maior parte [da realização dessa tarefa] é independente do restante da filosofia” (IMT, p. 290-1).

<sup>23</sup> Rawls menciona o exemplo da Filosofia Moderna que fez da epistemologia metodologicamente anterior ao resto da filosofia, assim passou-se a acreditar que outras questões filosóficas não poderiam ser satisfatoriamente resolvidas até que os problemas da epistemologia tivessem sido estabelecidos (Cf. IMT, p. 287).

<sup>24</sup> “Assim como a teoria do sentido e a filosofia da matemática estão relacionadas com a lógica e os fundamentos da matemática, ou até mesmo a filosofia da física está relacionada à física teórica, a filosofia moral está relacionada com a teoria moral, isto é, a explicação das estruturas morais e suas bases na psicologia moral (IMT, p. 287-8).”

básicas – bem (*good*), correto (*right*) e valor moral (*moral worth*) – são definidas e arranjadas de modo a formar diferentes estruturas morais, (ii) tentar identificar as principais semelhanças e diferenças entre essas estruturas e caracterizar o modo como se relacionam com nossa sensibilidade moral e atitudes naturais, (iii) determinar as condições<sup>25</sup> que as estruturas morais devem satisfazer se quiserem desempenhar o papel esperado na vida humana (Cf. IMT, p. 286), (iv) compreender qual concepção de pessoa, de relações entre pessoas, da estrutura geral e fins da cooperação social<sup>26</sup> são expressas por cada estrutura moral (KCE, p. 254). Ao realizar essa tarefa, a teoria moral cumpre o seu objetivo de “esclarecer e organizar os nossos pensamentos” (TJ, §9, p. 63), e produzir uma descrição detalhada de nosso senso de justiça (Cf. TJ, §9, p. 62-63).

Nesse estudo substantivo, o esforço teórico central para uma compreensão profunda das teorias morais “deve ser o estudo comparativo sistemático das sociedades bem-ordenadas.” (Cf. IMT, p. 294; 301).<sup>27</sup> É na comparação das doutrinas morais aplicadas a sociedades bem-ordenadas que todos os pontos de análise acima mencionados se articulam, tornando mais evidente o conteúdo das concepções morais e suas raízes na sensibilidade humana; neste momento da teoria - do confronto entre SBOs - é “que as conexões entre a teoria moral e a teoria social e psicológica são mais evidentes.” (IMT, p. 294). Esse estudo comparativo de teorias

---

<sup>25</sup> Condições ou restrições formais “são propriedades bem gerais que parece natural impor às concepções morais por várias razões” (IMT, p. 291). São elas: generalidade, universalidade, ordenamento/hierarquização, finalidade e publicidade (Cf. TJ, §23: *As restrições formais do conceito de correto*).

<sup>26</sup> De acordo com Rawls, as várias teorias morais incorporam diferentes concepções e ideais de pessoa e de relação entre pessoas (Cf. IMT, p. 293; 296; 301). No caso das concepções de justiça, quando plenamente articuladas, expressam também uma concepção “da estrutura geral e fins da cooperação social” (KCE, p. 254). É nesse sentido que podemos afirmar que o estudo substantivo das teorias morais implica em evidenciar como cada doutrina explica e interpreta as concepções acima mencionadas.

<sup>27</sup> Rawls foi um grande estudioso do tratado *Os Métodos da Ética* (1874), do filósofo inglês Henry Sidgwick, o qual fez um estudo comparativo e sistemático das concepções morais. De acordo com Rawls, a concepção de filosofia moral de Sidgwick tem a ver com a noção de que “qualquer justificação fundamentada e satisfatória de qualquer concepção moral específica deve partir do conhecimento completo e de uma comparação sistemática das concepções morais mais significativas da tradição filosófica” (CHFP, p. 411). Podemos perceber, pela análise de TJ e de artigos posteriores como *The Independence of Moral Theory* (1975), que o método de Sidgwick teve impacto sobre Rawls e que ele buscou aplicar esse método comparativo da filosofia moral em suas próprias investigações.

morais Rawls faz, sobretudo, na parte III de TJ<sup>28</sup> quando confronta a sociedade bem-ordenadas resultante da aplicação do princípio da utilidade com a da justiça como equidade.<sup>29</sup> A concepção de justiça que depois de aplicada às instituições da estrutura básica da sociedade for capaz de receber, em tese, adesão voluntária e duradoura dos cidadãos será mais estável em relação à outra concepção.

A sociedade bem-ordenada é concebida por Rawls como “um sistema social teoreticamente definido” (IMT, p. 294), e não necessariamente a descrição de alguma sociedade real. Por isso que já no início da TJ o filósofo alerta que intencionalmente está restringindo o alcance de sua teoria: não irá examinar a justiça de instituições em geral, nem de sociedades concretas, mas tão somente “os princípios da justiça que regeriam uma sociedade bem-ordenada” (TJ, §2, p. 9-10). Um dos objetivos de Rawls ao elaborar o expediente da sociedade bem-ordenada é investigar como os principais elementos de cada teoria moral

se encaixam dentro de um sistema em curso. Não se deve esperar obter conclusões detalhadas que dêem conta de situações particulares e casos práticos. Deve-se buscar uma visão geral de como a estrutura maior opera e mantém a si própria (IMT, p. 294).

Esse sistema social em curso é teoreticamente definido como (i) um sistema de cooperação (ii) moldado para promover o bem de seus membros<sup>30</sup> e que (iii) é efetivamente regulado por uma

---

<sup>28</sup> Essa comparação entre a SBO pela concepção de justiça como equidade e a SBO pela concepção utilitarista ocorre de modo intermitente ao longo dos Cap. VIII e XIX, entretanto, é apresentada de modo mais coeso em especial no §76.

<sup>29</sup> “O estudo comparativo de sociedades bem-ordenadas é, acredito, o principal empenho teórico da teoria moral: pressupõe uma compreensão de várias estruturas morais e sua relação com nossa sensibilidade moral e inclinações naturais. (...) “é neste estudo comparativo de sociedades bem-ordenadas que as conexões entre a teoria moral e a teoria social e psicológica são mais evidentes” (IMT, p. 294).

<sup>30</sup> Consideramos relevante destacar que essa ideia da sociedade bem-ordenada entendida como um sistema de cooperação destinado a promover o bem de seus membros pode ser encontrada textualmente em três passagens da TJ (Cf. TJ, §1, p. 5; §29, p. 218 ; §69, p. 560): “Digamos agora que uma sociedade é bem-ordenada não somente quanto foi planejada para promover o bem de seus membros, mas também quando é realmente regulada por uma concepção pública de justiça.”; (TJ, §1, p. 5); “(...) quando a sociedade é concebida como um sistema de cooperação destinado a promover o bem de seus membros (...)” (TJ, §29, p. 218); “No início, caracterizei a sociedade bem-ordenada como aquela moldada para promover o bem de seus membros e regulada de forma efetiva por

concepção pública de justiça, seja qual concepção for. Ser efetivamente regulada por uma concepção de justiça significa que tanto as instituições sociais quanto os indivíduos seguem e afirmam uma mesma concepção de justiça que é pública e mutuamente reconhecida. Rawls buscará mostrar ao longo do livro, em especial na Parte III, que uma sociedade justa (ou seja, uma sociedade bem-ordenada pelos princípios da justiça como equidade) deve ter como resultado a promoção de bem para os indivíduos tanto pessoalmente quanto coletivamente. Ou, dizendo de outro modo, é na Parte III da TJ que Rawls faz a análise mais profunda da estrutura das concepções morais sob análise (utilitarismo e justiça como equidade) e do modo como tais concepções se relacionam com nossas atitudes naturais e nossos sentimentos morais.

A execução da tarefa da filosofia moral na medida em que tal teoria busca vincular as concepções morais à sensibilidade humana acaba por colocar a psicologia no coração da teoria moral (Cf. BALDWIN, 2008, p. 253). Esse tipo de estudo depende de um trabalho cooperativo entre Teoria Moral, disciplinas filosóficas e ciências mais empíricas como a psicologia e a teoria social, a tal ponto que pode parecer que a execução de tal tarefa “parece pertencer à psicologia ou teoria social e não à filosofia” (IMT, p. 302). Rawls alerta que não devemos nos afastar dessa tarefa, e, nesse caso, aconselha manter a motivação/inclinação filosófica da teoria moral na investigação, pois sem ela a investigação corre o risco de ficar sem foco e conduzida por um caminho equivocado (Cf. IMT, p. 302).

Rawls considera que os sentimentos morais são parte normal da vida humana, e por mais desagradável<sup>31</sup> que possa ser ter de tratar desse assunto, “não temos como evitar a propensão a eles sem nos desfigurar a nós mesmos” (TJ, §74, p. 603). Uma teoria da justiça que se pretenda realista, que não se afaste muito da

---

uma concepção pública de justiça.” (TJ, §69, p. 560). Tanto no LP, quanto na JF, ao explicitar a noção de SBO, Rawls exclui essa característica da definição de SBO – moldada para promover o bem – mantendo somente a ideia de sociedade efetivamente regulada por uma concepção de justiça. Optamos por adotar a explicação inicial da SBO oferecida na TJ, pois ela permite que compreendamos uma das intuições filosóficas que, supomos, conduziu Rawls ao argumento da congruência. Rawls acredita que a justiça deve promover o bem humano e tenta mostrar isso no Capítulo IX.

<sup>31</sup> Para Rawls, lidar com os sentimentos morais não é algo agradável: “os sentimentos morais são realmente desagradáveis, em algum sentido mais amplo do termo desagradável” (TJ, §74, p. 603).

realidade, não pode deixar de levar em conta como os seres humanos são. Rawls defende que existe um vínculo entre sentimentos morais e atitudes naturais que pode ser expresso do seguinte modo:

(...) esses sentimentos [morais] e essas atitudes [naturais] são ambas famílias ordenadas de disposições características, e essas famílias se superpõem de tal maneira que a ausência de certos sentimentos morais demonstra a ausência de certos laços naturais. Ou, alternativamente, a presença de certos vínculos morais dá origem à propensão a certas emoções morais depois que ocorre o desenvolvimento moral necessário (TJ, §74, p. 601).

Em decorrência desse vínculo, uma característica de nossa humanidade é que possuímos sentimentos, por essa razão que ignorá-los nos alija de nosso caráter humano. Dos laços naturais de sociabilidade surge a propensão ao amor, confiança, amizade, afeto, devoção a instituições, e com eles também a culpa, indignação ressentimento, dentre outros sentimentos morais (Cf. TJ, §74, p. 602-3). Não apenas valorizamos tais disposições afetivas, como necessitamos delas. “Supondo-se, portanto, que precisamos desses afetos” (TJ, §86, p. 704), Rawls atribui à teoria moral a tarefa de conectá-las ao senso de justiça (WEITHMAN, 2010, p. 286).

É necessário, portanto, reservar um lugar apropriado para a psicologia moral no interior da teoria da justiça, uma vez que é através dela que se faz a conexão entre a estrutura das concepções morais com a nossa sensibilidade humana. Ademais, a teoria da justiça deve levar em conta a psicologia moral gerada pela aplicação das concepções de justiça, pois

a justiça ou injustiça dos arranjos institucionais da sociedade e as convicções humanas sobre essas questões influenciam de maneira profunda os sentimentos sociais; em grande parte, definem como vemos a aceitação ou a rejeição de alguma instituição por outra pessoa, ou seu empenho em reformá-la ou defendê-la (TJ, §75, p. 607).

Os valores morais que orientam a sociedade em que crescemos e vivemos influenciam a nossa psicologia, isto é, muitas

das tendências e inclinações que desenvolvemos enquanto indivíduos são estimuladas/engendradas pelas instituições sociais sob as quais crescemos, ou dito de outro modo, o contexto social em que estamos inseridos influencia não apenas o nosso senso de justiça, como vários outros sentimentos morais, tais como ressentimento, aversão ao risco, vergonha e humilhação, tendência a dominar ou a se subjugar, etc<sup>32</sup>. Rawls formula leis da psicologia moral tendo em vista um contexto institucional que é justo e que é publicamente reconhecido como tal, e também afirma que cada concepção de justiça irá gerar diferentes princípios ou tendências da psicologia moral (Cf. TJ, §75, p. 606). É nesse sentido que Baldwin afirma que o uso que Rawls faz do termo ‘psicologia moral’ implica que nossa psicologia é “inerentemente moral, não apenas em relação ao seu conteúdo, mas também a respeito de sua dependência da justiça, e assim, da moralidade, da nossa sociedade” (BALDWIN, 2008, p. 249).

Por que razão a teoria da justiça precisa de uma psicologia moral? Para explicar porque os princípios de justiça são aceitos pelas pessoas. Não é suficiente que os princípios sejam racionais e que tenham sido escolhidos em um procedimento imparcial e equitativo, é necessário que sejam tais que as pessoas possam concordar em agir a partir deles. “ (...) uma teoria correta da política num regime constitucional justo pressupõe uma teoria da justiça que explique como sentimentos morais influem na condução dos assuntos públicos” (TJ, §75, p. 608). Refletir sobre o modo como as principais noções da filosofia moral estão ligadas à “psicologia e sociedade humana é enfrentar a mais importante questão com a qual podemos nos confrontar: como deveríamos viver as nossas vidas?” (LARMORE, 2008, p. 71).

Ao se justificar uma teoria moral, os pressupostos devem ser confrontados com os nossos sentimentos morais cotidianos sobre a justiça e passar por um processo de ajuste mútuo entre os princípios da teoria e os nossos juízos ponderados<sup>33</sup>. O objetivo

---

<sup>32</sup> “Essa explicação sugere que a psicologia moral lida com aspectos do desenvolvimento humano, o que faz com que a psicologia moral faça parte, em TJ, de uma explicação abrangente da psicologia humana” (Baldwin, 2008, p. 249).

<sup>33</sup> Juízos ponderados são aqueles juízos nos quais é mais provável que nossas capacidades morais se manifestem sem distorção. Ao contrário de juízos feitos com hesitação ou nos quais temos pouca confiança, são emitidos em condições favoráveis ao exercício do senso de justiça, e, por conseguinte, em circunstâncias nas quais são inaceitáveis as desculpas e

deste capítulo é abordar o primeiro elemento característico das teorias morais: o modo como a teoria da justiça como equidade define e relaciona os conceitos de bem e de correto na estrutura argumentativa da teoria, e a partir desta perspectiva mostrar apontar como surge o problema da congruência.

### 3.2 Distinções gerais entre as noções de correto e de bem.

A fim de podermos analisar como a concepção moral da justiça como equidade define e articula os conceitos de correto, bem e valor moral na estrutura da teoria e como os relaciona com nossos sentimentos morais, explicitaremos primeiro a distinção entre estes conceitos, para em seguida, tratar do modo como são articulados na estruturas morais.

Em comum, correto e bem são noções morais, no entanto, cada uma delas faz parte de famílias de critérios distintos. “Mais precisamente, cada conceito, com seus princípios associados, define uma perspectiva [a partir] da qual se podem avaliar as instituições, os atos e os planos de vida” (TJ, §86, p. 699).

Como vimos na seção anterior, uma das tarefas da Teoria Moral consiste em fazer uma análise de como as noções morais de bem, correto e valor moral são definidas e ordenadas na estrutura das diferentes doutrinas morais. A fim de compreender o modo como a teoria da justiça como equidade estrutura tais conceitos – e os vincula a nossos sentimentos morais – passamos a explicitar as diferenças entre as noções morais de correto e bem, uma vez que “a estrutura de uma doutrina ética depende de como ela relaciona essas duas ideias e define essas diferenças” (TJ, §68, p 552).

A primeira distinção entre tais noções, de acordo com Rawls, diz respeito ao âmbito de aplicação. O conceito de correto é aplicado a ações (enquanto tipos ou espécies) e a instituições – consideradas justas (*right*) ou injustas (*wrong*), ao passo que o conceito de bem é aplicado a estado de coisas, atividades ou coisas de vários tipos - considerados bons (*good*) ou maus (*bad*) (Cf. HFM, p. 63, n2). Já o conceito de valor moral é aplicado a

---

explicações mais comuns para o erro. Já equilíbrio reflexivo é o estado em que a pessoa chega depois de ponderar (considerar/analisar) as diversas concepções propostas e de ter ou bem reconsiderado os próprios juízos para que se adaptem a uma delas, ou bem se apegado a suas convicções iniciais (e à concepção correspondente) (Cf. TJ, §9, p. 57-59).

(...) nosso caráter como um todo, ou às várias disposições ou motivos pelos quais agimos, bem como a ações particulares que reputamos como virtuosas em vista das disposições e motivos pelos quais são realizadas – considerados moralmente valorosos ou não valorosos, ou, em casos extremos, maus (*evil*) (HFM, p. 63, n2).

Dos três principais conceitos morais, Rawls considera que as noções de correto e de bem são as mais fundamentais, uma vez que são elas que permitem explicar a terceira noção básica: o conceito de pessoa moralmente digna ou de valor moral (*moral worth*) (Cf. TJ, §5, p. 29; §68, p. 552). Assim é que o valor moral de cada doutrina depende de qual dessas noções – correto ou bem – é pensada como a mais fundamental.<sup>34</sup> Outra explicação que Rawls nos oferece é que “nenhuma doutrina moral pode existir sem uma ou mais concepções de bem, e uma concepção aceitável do correto deve guardar um espaço adequado para tais concepções” (HFM, p. 265). Em TJ estão presentes cinco ideias de bem: Bem como racionalidade, bens primários, concepções abrangentes e aceitáveis de bem, o bem representado por uma sociedade política bem-ordenada, e a ideia de virtudes políticas (Cf. RAWLS, PRIG, p. 449). Buscaremos mostrar, ao longo do trabalho, qual o lugar destas ideias e como elas estão articuladas com relação ao correto na teoria da justiça como equidade.

A primeira distinção fundamental entre correto e bem – quanto ao escopo – conduz a outras implicações: como têm aplicações diferentes, tais noções devem ser definidas a partir de pontos de vistas distintos. Uma vez que os princípios do correto serão aplicados na elaboração de constituições e arranjos sociais básicos, tais princípios devem ser escolhidos – e também suas regras de prioridade – a partir de um ponto de vista público, em termos tais que possam ser aceitos por todos; não há essa exigência quanto ao bem. Por isso, o conceito de correto deve ser expresso de um modo claro e preciso de modo que possa ser compreendido e aceito por todos. O conteúdo do conceito de correto não pode ser

---

<sup>34</sup> Na visão de Larmore, o embate entre teleologia e deontologia ocorre no padrão da ética imperativa (até mesmo por uma questão histórica.). O que tem predomínio na imperativa é a concepção de correto: “uma prescrição autoritativa, uma regra ou um ditame ao qual a conduta deve estar subordinada” (LARMORE, 1996, p. 20).

expresso de forma vaga, ambígua, imprecisa, amorfa (termos que Rawls usa ao longo do livro para caracterizar o que considera um defeito do princípio de justiça utilitarista), sob pena de o conceito do que é justo estar sempre em disputa interpretativa. É necessário que seja claro e que se chegue a um consenso acerca de seu conteúdo para que a concepção de justiça sirva como uma base pública a partir da qual os conflitos oriundos das circunstâncias da justiça possam ser mediados.

Enquanto a teoria do correto tem por escopo oferecer um padrão objetivo de justiça aplicável a todos, a teoria do bem busca um critério objetivo para o bem da pessoa, o qual “é definido em relação ao plano racional que seria escolhido com plena racionalidade deliberativa” (TJ, §64, p. 524). Dada a heterogeneidade dos indivíduos, variados podem ser os bens – ou, projetos de vida – que cada um pode buscar. Por isso, no que diz respeito ao bem, um acordo unânime tal como o em relação ao correto, não apenas não é desejável como não é possível (a menos que haja a imposição pela violência). O próprio conceito de bem é polissêmico, Rawls considera que essa pluralidade de concepções de bem não só é um fato inevitável<sup>35</sup>, como é, em si mesmo, uma coisa boa, na medida em que é

(...) racional que os membros de uma sociedade bem-ordenada queriam que seus projetos sejam diferentes. Os motivos disso são óbvios. Os seres humanos têm inúmeros talentos e capacidades cuja totalidade é inalcançável por qualquer pessoa ou grupo de pessoas. Assim, além de nos beneficiarmos da natureza complementar das nossas inclinações desenvolvidas, temos prazer nas atividades uns dos outros. É como se os outros estivessem revelando uma parte de nós mesmos que não conseguimos cultivar (TJ, §68, p. 554).

A concepção de bem deve se ajustar a situação particular do indivíduo, já a de justiça não deve ser concebida para se ajustar a isso. Estes contrastes entre correto e bem tornam mais clara a ideia de que essas noções estão vinculadas, na teoria da justiça como equidade, à duas perspectivas ideais distintas: (i) a posição original,

---

<sup>35</sup> No LP, o pluralismo de concepções abrangentes é um dos fatos gerais que caracterizam as sociedades bem-ordenadas.

que é perspectiva coletiva pública e exige acordo social unânime, o qual fornece a base para os princípios de justiça; e a (ii) racionalidade deliberativa, que é o ponto de vista do indivíduo (Juízos são feitos individualmente (*singly*), por cada indivíduo), e diz respeito à base para os juízos que consideram o bem da pessoa; neste ponto de vista, todas as informações abstraídas pelo véu da ignorância são restauradas. O conhecimento pleno das circunstâncias antes abstraídas é requerido, uma vez que os juízos de valor serão explicados relativamente aos fins particulares de indivíduos e de situações. Ambas as perspectivas são idealizações, nenhuma delas toma os indivíduos concretos. (Cf. FREEMAN, 2007a, p. 150-151). Assim é que, enquanto que os princípios do justo são escolhidos na Posição Original, os princípios da escolha racional e os critérios da racionalidade deliberativa<sup>36</sup> não o são.

Na estrutura moral da justiça como equidade, as noções de correto e de bem são especificadas a partir de dois pontos de vista distintos. Veremos que o âmbito em que é especificado o correto é logicamente anterior ao âmbito em que é especificado – ou melhor, descrito<sup>37</sup> – o bem. Significa dizer, há uma relação de prioridade do ponto de vista do correto em relação ao do bem. Tal prioridade não implica que a teoria moral tenha de ignorar o bem. A noção moral de bem continua a ter papel importante na estrutura da teoria, pois, enquanto que a justiça estabelece os limites, o bem indica a finalidade. Além da relação de prioridade entre correto e bem, tais conceitos devem ser articulados de tal modo que também estejam numa relação de complementaridade (cf. PRIG, p. 449). A proposta de Rawls é que os juízos elaborados a partir dessas duas perspectivas distintas devem ser congruentes, ou, pelo menos, não podem ser incompatíveis.

---

<sup>36</sup> Princípios de escolha racional: princípio de meios eficazes; princípio da abrangência; princípio da maior probabilidade; princípios de contagem (Cf. TJ, §63, p. 509-513).

Princípios da racionalidade deliberativa: princípio do adiamento, princípio da continuidade; princípio da vantagem em elevar as expectativas; princípio da responsabilidade consigo mesmo (Cf. TJ, §68, p. 519-523).

<sup>37</sup> Rawls assume que bom tem uso descritivo (e não prescritivo) das propriedades que é racional desejar em um determinado tipo de coisas. Isso implica que o significado do termo é explicado pelo sentido descritivo constante dado por seu uso (Cf. TJ, §62, p. 501).

### 3.3. Como as noções morais são articuladas nas estruturas morais do utilitarismo e justiça como equidade

É consenso que qualquer teoria moral que se pretenda plausível deve apresentar reflexões de duas naturezas: sobre o que nós devemos ou não fazer tanto do ponto de vista individual quanto coletivo, e sobre quais objetivos finais devemos realizar tanto direta quanto indiretamente. O primeiro tipo de reflexão diz respeito ao conceito de correto, enquanto que o segundo envolve a concepção de bem. Rawls considera em TJ três modos em que é possível relacionar os conceitos de correto e bom na estrutura interna das concepções morais: (i) relacioná-los sem estabelecer qualquer prioridade e ordenamento entre eles; (ii) definir a primazia do bem em relação ao correto e (iii) estabelecer a prioridade do correto em relação ao bem. As teorias intuicionistas são do primeiro tipo, conferem uma validade *prima facie* aos conceitos de correto e de bom, razão pela qual não consideram necessário definir uma regra de prioridade de um conceito sobre outro. Em caso de conflito de valores e princípios, ou de conflito entre princípios primeiros, o equilíbrio deve ser feito através da intuição, pelo que parece se aproximar do que é justo (cf. TJ, §7, p. 41). Mas, em se tratando da escolha de princípios de justiça esse método não parece o mais adequado, pois estamos falando de princípios que vão regular as instituições sociais, e o modo como elas são reguladas afeta diretamente a vida das pessoas, seus afetos e laços sociais, interfere no modo como os indivíduos concebem a si próprios, seus desejos e seus fins. A escolha de princípios de justiça deve se basear em razões com as quais todos possam razoavelmente concordar – ou, pelo menos, que não possam razoavelmente rejeitar –, não pode ser algo como uma mera preferência, como se estivéssemos escolhendo entre tomar chá ou café, entre morar numa casa ou num apartamento.

Já as concepções morais do tipo teleológicas, das quais utilitarismo e perfeccionismo teleológico fazem parte, têm a vantagem, em relação ao intuicionismo, de admitirem a necessidade de que seja estabelecida uma regra de prioridade entre os conceitos de correto e bem na estrutura da teoria moral. Reconhecem o bem como um fim último que é predominante sobre todos os outros, inclusive sobre o correto, por isso “a noção do bem é definida

anterior e independentemente da de correto, sendo esta aquilo que maximiza o bem” (CHFP, p. 433). Já a terceira maneira de definir e estruturar as noções morais no interior da teoria é através do padrão deontológico, o qual inverte a estrutura teleológica e passa a definir o correto com prioridade sobre o bem.

Na busca por formular um método claro e aplicável de raciocínio moral, as teorias teleológicas<sup>38</sup> parecem representar a maneira mais simples e intuitiva de definir e interligar o bem e o correto acordo com nossas crenças sobre a justiça. A estrutura destas se caracteriza por definir a concepção de bem geralmente como um único fim predominante (a felicidade ou o prazer obtido na satisfação do interesse racional, nas teorias utilitaristas, ou, a excelência humana ou certos valores, nas perfeccionistas<sup>39</sup>) “ao qual todos os outros fins estão subordinados” (TJ, §83, p. 681). Isso significa que quaisquer outros valores, fins e condutas humanas devem ser justificados e ordenados com referência ao fim último predominante. O bem é especificado de um modo não moral, isto é, “sem referência a quaisquer conceitos ou princípios morais e em termos naturais, ou que de outra forma seriam apropriados nas ciências sociais e naturais” (FREEMAN, 2007a, p. 84), e o conceito de correto é, então, definido como aquilo que produz e acumula o máximo possível do bem anteriormente especificado. A relação entre esses conceitos na estrutura da teoria teleológica ocorre do seguinte modo: o bem é definido anterior e de modo independente do correto, sendo este formulado como mero instrumento de realização do bem (um único fim último) (Cf. TJ, §5 p. 29-30). Para fixar o raciocínio, elencaremos as três principais características da estrutura das teorias morais teleológicas vistas até aqui:

- (i)** O bem é definido antes e independente do correto: significa que o conceito de correto é definido sem referência a qualquer princípio ou conceito moral, mas com termos que são usados de modo mais apropriado nas ciências sociais ou naturais.

---

<sup>38</sup>Um dos aspectos importantes da discussão e críticas que Rawls faz ao utilitarismo é que fornece “um padrão de raciocínio que pode ser aplicado a outras concepções teleológicas, como aquelas que afirmam que é correto e justo sempre agir de modo a maximizar o bem” (FREEMAN, 2007b, p 146).

<sup>39</sup> Outros exemplos de fins dominantes mencionados por Rawls: unidade religiosa, máxima excelência da cultura, poder e prestígio nacionais; felicidade (Cf. TJ, §79, p. 652; §85, p. 693).

**(ii)** O bem é definido como um fim dominante (bem supremo) ao qual todos os outros fins estão subordinados, ou, como um único bem racional. Esse bem pode ser a utilidade ou mesmo a perfeição.

**(iii)** O correto é definido de modo instrumental, compreendido como aquilo que promove a maximização do bem; a ação moralmente correta é, então, aquela que maximiza o bem -tal como definido em (ii).

O pressuposto de que há um único bem racional a ser maximizado é uma das características mais marcantes das teorias teleológicas, em especial do utilitarismo, por isso, começaremos por explicá-lo. A definição do bem como um fim predominante definido independentemente do justo tem pelo menos duas funções nas teorias teleológicas: (i) a função de “tornar razoavelmente precisa a concepção de correto” (TJ, §85, p. 698) e (ii) tornar sempre possível tomar uma decisão racional (Cf. TJ, §83, p. 681). Por definir o correto como o meio mais eficiente para maximizar o bem, é essencial que o conteúdo do fim último seja claro e preciso, pois, “qualquer ambiguidade ou imprecisão na concepção é transferida para a do correto” (TJ, §84, p. 690). Acontece que é inegável que há uma pluralidade de experiências prazerosas, bens, valores e atividades que podem ser consideradas como um fim último racional a ser alcançado. Como então, escolher um dentre estes fins para ser o predominante? Para lidar com esse dado, a teleologia precisa de “um modo comparar os diversos bens de diferentes indivíduos para que se possa maximizar o bem total” (TJ, §84, p. 690). É desta necessidade que se origina a concepção de um único bem racional que possa ser o propósito último de todas as atividades humanas. “Como não se pode maximizar mais de uma única coisa, deve haver um fim último que preencha esse padrão de comparação e, com isso, o objeto final de toda atividade correta” (FREEMAN, 2007a, p. 54).

A segunda função do pressuposto de um fim predominante é fazer com que seja sempre possível encontrar o curso de ação mais racional possível para cada caso que se apresente, garantindo que escolhas racionais sejam sempre possíveis, quaisquer que sejam as circunstâncias, em qualquer âmbito das atividades humanas. Utilitaristas clássicos, como Sigdwick e Mill, consideram que a

definição de um fim racional predominante sobre outros fins tem origem na necessidade da razão prática de conferir um ordenamento racional completo, sistemático e coerente de todas as condutas humanas (Cf. FREEMAN, 2007a, p. 78). Essa função confere ao utilitarismo a capacidade, muito admirada por Rawls, de sistematizar e ordenar<sup>40</sup> completamente as condutas e reivindicações humanas, o que implica na possibilidade de oferecer normas de justiça e moralidade para qualquer mundo social, para qualquer circunstância. O utilitarismo tem a capacidade de fornecer “um modelo para uma moralidade completamente racional, capaz de determinar para cada situação o melhor e mais racional curso de ação” (FREEMAN, 2007a, p. 77). Tomando a perspectiva de outras doutrinas morais, pode haver situações em que pode parecer ao agente que qualquer curso de ação que tomar estará moralmente errado, como no clássico dilema do Trolley. Na perspectiva do utilitarismo, entretanto, é praticamente impossível haver alguma situação para a qual não possa se oferecer uma resposta racional sobre qual é conduta correta a ser seguida, porque o correto (enquanto um conceito moral) é concebido como o meio mais eficiente para chegar a um fim último. Nesse caso, a “decisão correta é, em essência, uma mera questão de administração eficiente” (TJ, §5, p. 33) para se chegar ao fim último, no caso do utilitarismo, produzir e acumular o saldo máximo de satisfação.

O pressuposto de um único fim racional predominante ao mesmo tempo em que é responsável pela grande virtude que representam o vigor, a sistematicidade e ordenamento das teorias utilitaristas, é o ponto por onde Rawls começa a desarmar tal doutrina. Esse tipo de estrutura moral não oferece bases para lidar de modo adequado com a questão da distribuição de bens que orienta a investigação de Rawls. O filósofo reconhece que, por um lado, a definição de um bem único justificado como tendo primazia

---

<sup>40</sup>Importante notar que a condição do ordenamento é uma vantagem que as teorias teleológicas mantêm em relação às várias formas de intuícionismo. Estas últimas teorias, ao contrário do utilitarismo, reconhecem uma pluralidade de bens e desejos, no entanto, não dispõem de um procedimento para equilibrar considerações morais conflitantes, deixando esta tarefa ao cargo da intuição. Já o utilitarismo define um único fim racional e todas as escolhas devem ser feitas com vistas a alcançar esse bem (no caso, maximizar a experiência prazerosa ou a satisfação de preferências racionais). Assim, diante de reivindicações ou considerações morais conflitantes, o princípio da utilidade fornece um critério: escolher aquela que melhor realiza e maximiza o bem.

sobre quaisquer outros bens representa um mérito da doutrina utilitarista clássica que é reconhecer com clareza que o que está em questão em termos de justiça social é estabelecer a “prioridade relativa dos princípios de justiça e dos direitos derivados desses princípios” (TJ, §6, p. 40). Ao contrário as doutrinas intuicionistas, que reconhecem uma pluralidade de princípios e fins, mas não apresentam um procedimento ou regra de hierarquia para lidar com princípios conflitantes, o utilitarismo oferece um único princípio, com um único e supremo critério para lidar com as reivindicações conflitantes: em qualquer dilema sobre qual princípio escolher, sobre qual fim realizar, a resposta utilitarista é que deve prevalecer aquele que permitir a maior produção e acumulação do interesse racional dos indivíduos ou grupos. “O fim predominante tem, pelo menos lexicalmente, primazia sobre todos os outros objetivos e procurar promovê-los sempre requer precedência absoluta” (TJ, §83, p. 682). O problema é que ao estabelecer um único princípio com relação ao qual todos os demais princípios e valores devem ser ordenados e justificados, o utilitarismo acaba por reduzir a moralidade ao dever de maximizar a soma total dos bens, e continua sem oferecer uma resposta adequada sobre como esse bem é distribuído. Rawls defenderá que a solução mais adequada para este problema passa por uma deliberação complexa que implica em ordenar uma pluralidade de princípios em ordem serial.

A crítica que Rawls faz às concepções morais de fim predominante mostra que a solução dada pelo utilitarismo para o problema de “como equilibrar princípios e determinar precedências” (TJ, §83, p. 683) é equivocada, pois, dada a heterogeneidade do bem humano, é impossível definir um único objeto definitivo apropriado para ser maximizado, como é questionável também especificar um único fim para ser o propósito último de todas as atividades. Há várias finalidades que os indivíduos podem racionalmente desejar como fim último para orientar seus projetos de vida e a distribuição de bens na sociedade, que não apenas o acúmulo máximo de experiências prazerosas, como a liberdade, busca pela verdade, pelo poder, refinamento estético, a Visão de Deus, a salvação da alma, dentre tantos outros e que não são necessariamente irracionais ou excludentes entre si. A razão prática ao ser concebida como requerendo de nós a maximização de um único bem racional entra em conflito direto

com nossas crenças mais fundamentais sobre a justiça que reconhecem que há uma pluralidade de bens e valores que podemos racionalmente desejar. Dada as diferenças significativas que existem entre as diversas concepções de bem, o problema prático que recai sobre o conceito de correto é que este passa a estar sempre em disputa.

Ademais, a impossibilidade de dar uma definição precisa ao conteúdo do fim único, como os acima mencionados (liberdade, verdade, experiência prazerosa, etc.), implica que o conteúdo da concepção de justiça permaneça continuamente em disputa, coisa que não é desejável em se tratando do conteúdo de uma concepção de justiça. Por outro lado, ainda que fosse possível definir de modo claro e preciso um bem último por referência ao qual fosse possível fazermos todas as nossas escolhas racionalmente, as consequências poderiam ser terríveis tanto para pessoas, individualmente, uma vez que “o eu (*self*) fica desfigurado quando é colocado, por razões de sistema, a serviço um de seus fins”, quanto do ponto de vista coletivo, pois,

(...) certamente quando um fim predominante é claramente especificado por referência a alguma meta objetiva como o poder político ou a riqueza material, o fanatismo subjacente e a desumanidade se tornam manifestos (TJ, §83, p. 684).

Outro problema de uma concepção da justiça pensada no padrão teleológico utilitarista diz respeito a como proceder a distribuição dos frutos da cooperação social. O utilitarismo não oferece resposta razoável para pensar a justiça para instituições em termos da distribuição de bens (primários), pois seu raciocínio contempla não a distribuição de vários bens, mas a maximização, e de um único bem, um único fim dominante. Nesse sentido, uma instituição orientada pelo princípio da utilidade é considerada justa quando maximiza a utilidade, isto é, quando proporciona a maior produção e o maior acúmulo possível da satisfação do interesse racional. Não importa se nesse cálculo para a maximização algum indivíduo ou grupo receber uma soma zero de bens, não é relevante se por receber a soma zero dos bens oriundos da cooperação social a autoestima e o senso do próprio valor dos indivíduos sejam acachapados, importa sim é que a soma final de satisfação para toda a sociedade seja maximizada. Na teoria da justiça como equidade,

entretanto, o princípio de elevação ao máximo não é usado, e “a questão de se produzir o mais alto saldo líquido de satisfação nunca se apresenta” (TJ, §6, p. 37).

Cabe ainda acrescentar que do ponto de vista da motivação este tipo de teoria também é problemática, uma vez que, ao considerar um único fim dominante, admitirá somente a descrição sobre o fim que pode ser desejado. Neste sentido, a teoria da justiça como equidade terá mais força motivacional por permitir uma pluralidade de descrições.

Nas teorias utilitaristas de fim predominante, o que confere a unidade do eu – a unidade da razão prática – é a busca do prazer, não há a necessidade de que o prazer buscado seja o do indivíduo, mas simplesmente que seja prazer. Pode-se então definir que o prazer a ser alcançado é o da sociedade como um todo, e nesse caso, a justiça poderia exigir que “inúmeras pessoas, em seus esforços conjuntos, devem esforçar-se por organizar seus atos coletivos de forma a maximizar as experiências agradáveis do grupo” (TJ, §85, p 693). Os próprios indivíduos são instrumentalmente colocados – e em nome da justiça utilitarista devem aceitar se colocar desse modo – a serviço de um fim predominante.<sup>41</sup>

Por tudo isso, Rawls afirma enfaticamente que “a estrutura das doutrinas teleológicas é radicalmente equivocada” (TJ, §84, p. 691). O filósofo defende que o desenvolvimento da teoria moral deve ocorrer na direção oposta daquela dada pelas doutrinas teleológicas: a relação entre bem e correto deve ser invertida, e isso implica que a definição do que seja o correto deve ser feita em instância lexicalmente anterior ao bem, ou, dito de modo mais específico, a definição do correto deve ser feita em um âmbito logicamente independente e não condicionado ao âmbito em que é especificado o bem. Nesse sentido a teoria moral da justiça como equidade se coloca como antagonista das teorias teleológicas. O bem

---

<sup>41</sup> Uma crítica dirigida ao utilitarismo que quando colocado em prática, a medida da utilidade (entendida como felicidade, experiências agradáveis ou satisfação de desejos racionais) pode ser calculada em valores monetários. De acordo com Williams, “Eles [os utilitaristas] poderiam sempre alegar que a medida comum da felicidade proposta por eles não é o dinheiro. Mas, na prática, os seus métodos terão essas implicações – pois eles estão comprometidos com a ideia de que, em última análise, não existem valores incomensuráveis. (...) o utilitarismo é o sistema de valor – o que não surpreende – adotado nas sociedades em que os critérios econômicos são preponderantes; além disso, no nível teórico a quantificação do dinheiro é a única tradução óbvia do critério utilitarista, que é a comensurabilidade do valor (WILLIAMS, 2005, 149-150).

passa a ser definido não mais como um único fim dominante, mas como múltiplos fins que são limitados pela teoria da justiça, e a justiça, por sua vez é definida de modo a permitir a realização – e não a maximização – desses bens. A teoria da justiça como equidade, ao inverter o padrão teleológico de estruturar os conceitos, é entendida como uma teoria do tipo deontológica.

Antes de explicar como as teorias deontológicas representam a terceira alternativa à questão de como relacionar os conceitos na estrutura da teoria moral, é oportuno fazer a distinção entre deontologia e prioridade do correto, uma vez que não é incomum que estes conceitos correlatos sejam tomados como significando a mesma coisa.<sup>42</sup> Em comum, ambas ideias descrevem a “estrutura e conteúdo substantivo de uma concepção moral, não a sua justificação procedimental (...) não é uma reivindicação sobre a ordem de justificação da concepção moral” (FREEMAN, 2007a, p. 62). Ambas ideias dizem respeito ao modo apropriado de relacionar os conceitos de correto de bem, no entanto, o que as difere é o âmbito de aplicação de cada uma delas:

A prioridade do correto descreve o lugar dos princípios de justiça do correto no raciocínio prático de agentes morais motivados por um senso de correto e justiça, como eles devem singular e coletivamente aplicar princípios ‘que recebem um teor definido desde o início’ [T], §6, p.38] ao decidir seu bem e o que eles devem fazer. A “prioridade” em prioridade do correto se refere à ordem lexical dos princípios do correto e da justiça na deliberação social e individual (FREEMAN, 2007a, p. 63).

É importante notar que a prioridade do correto é apenas uma dentre “as diversas regras de prioridade” (TJ, §76, p. 618) atribuídas aos princípios na teoria da justiça como equidade, além da “prioridade da justiça sobre a eficiência, a prioridade das liberdades básicas sobre o princípio da diferença, e a prioridade do valor moral sobre valores não morais” (FREEMAN, 2007a, p. 63).

---

<sup>42</sup> De acordo com Samuel Freeman, Will Kymlicka (em *Liberalism, Community, and Culture*. Oxford: Oxford University Press, 1989) e Michael Sandel (em *Liberalism and the Limits of Justice*. Cambridge: Cambridge University Press, 1982), dentre outros autores, confundem deontologia com prioridade do correto (Cf. FREEMAN, 2007a, pp. 45-74).

Enquanto a prioridade do correto diz respeito à estrutura da razão prática de agentes morais movidos por um senso de justiça, a deontologia é uma reivindicação sobre como princípios morais são definidos e se relacionam com os conceitos de correto e de bem no interior da estrutura da teoria moral. Estes dois conceitos, tomados em conjunto, permitem que o bem racional de pessoas morais seja definido por referência aos princípios deontológicos do correto.

Em momento algum de TJ Rawls define ou se refere às teorias deontológicas como aquelas que dão prioridade ao correto sobre o bem. Sem alongar-se na explicação, Rawls define-as simplesmente como teorias não teleológicas (TJ, §6, p. 36). Significa dizer que as doutrinas deontológicas estabelecem como que os conceitos de bem e de correto não devem ser interpretados e como não devem ser definidos. Rawls apresenta dois sentidos em que a deontologia pode ser compreendida; pode ser uma teoria que

[i] não especifica o bem independentemente (*independently*) do correto, ou

[ii] não interpreta o correto como aquilo que maximiza o bem (TJ, §6, p. 36).

A explicação sobre quais seriam as diferenças entre esses dois sentidos não é aprofundada no livro, o filósofo afirma apenas que a justiça como equidade é uma teoria do segundo tipo, ou seja, não interpreta o correto como aquilo que maximiza o bem. Desse modo, a principal orientação que a deontologia fornece para a teoria moral é vedar que o conceito de correto seja instrumentalizado em favor da maximização de um bem; como a teoria deontológica foi definida por oposição à teleológica, assumimos que o bem em questão também não pode ser definido como um fim predominante (ou supremo) ao qual todos os outros fins estão subordinados. Rawls defende que o projeto mais amplo a ser realizado nas instituições públicas da sociedade, isto é, a realização constitucional dos princípios de justiça, não define um fim dominante ao qual os fins de todos os indivíduos e associações estejam subordinados. A intenção é de que a ordem constitucional realize os princípios de justiça (Cf. TJ, §79, 652).

A deliberação sobre o conteúdo do correto não pode, então, estar de antemão comprometida em promover a maximização do bem. Rawls defende que quando as partes estão escolhendo os princípios de justiça, o que as move é o desejo de garantir uma quantidade maior de bens primários – bens fundamentais para

realizar um projeto racional de vida – e não o de maximizar esses bens (ou a satisfação obtida através deles). Quando o utilitarismo define o correto como o meio mais eficiente de maximização do bem não está preocupado com o impacto e consequências que a aplicação institucional do princípio da utilidade terá na vida das pessoas, pois aceita que o custo da maximização seja exigir (como uma exigência da justiça) dos indivíduos uma disposição para o auto sacrifício e atos heroicos.

É importante manter distinção entre as ideias de deontologia e prioridade do correto, sem perder de vista, no entanto, que ambas funcionam em conjunto. Enquanto a deontologia restringe um modo específico de estruturar os conceitos de correto e bem na teoria moral, a prioridade do correto oferece orientação sobre qual a relação mais adequada entre ambos os conceitos. Temos ainda que considerar que afirmar a primazia do correto em uma determinada teoria moral não pode implicar que o bem seja desconsiderado<sup>43</sup>.

Rawls não aceita que a escolha do correto e dos princípios de justiça seja feita sem levar em conta os efeitos de sua aplicação institucional na vida das pessoas. De acordo com ele,

Todas as doutrinas éticas dignas de atenção levam em conta as consequências ao julgar o que é certo. Aquela que não o fizesse seria simplesmente irracional, insana (*crazy*) (T), §6, p. 36).

A proposta deontologista rawlsiana incorpora no raciocínio e na justificação dos princípios uma preocupação com o estado de coisas resultante da aplicação institucional dos princípios de justiça política. Cabe mencionar que Rawls muda o foco da preocupação deontológica, pois a perspectiva da deontologia voltada apenas a correção das ações individuais é insuficiente para lidar com questões estruturais. Uma interpretação rigorosa da deontologia voltada apenas para ações individuais requer apenas que “você não explore diretamente o trabalho de crianças, mas vocês (e os demais) não têm um dever de evitar que o trabalho infantil continue existindo no mundo” (VITA, 2008, p. 33n17). Muitas situações de opressão e injustiça não são causadas por ações individuais injustas,

---

<sup>43</sup> “Nenhuma doutrina moral pode existir sem uma ou mais concepções de bem, e uma concepção aceitável do correto deve guardar um espaço adequado para tais concepções” (HFM, p. 265).

mas pelo modo como as instituições sociais e arranjos socioeconômicos e políticos são modelados, por isso, o objeto da justiça é a estrutura básica da sociedade.

Nesse sentido, é imperativo que a teoria da justiça se questione e leve em consideração que mundo social seria gerado se tais princípios fossem aplicados incondicionalmente à estrutura básica da sociedade e todos agissem de acordo com as exigências da justiça. Para que a concepção seja, ao final, considerada justa, é preciso que os agentes sejam capazes de desejar viver nesse mundo social, mais do que em qualquer outro, caso venham a pertencer a ele.

Nesse ponto, o leitor familiarizado com uma interpretação mais rigorosa da deontologia – a qual desconsidera por completo as consequências da aplicação dos princípios – pode ficar em dúvida sobre como e em que momento do raciocínio deontológico rawlsiano se insere a discussão sobre qual estado de coisas deve ser considerado justo em termos de conteúdo. Como definir e justificar um conjunto de normas, ou melhor, de princípios de justiça que sejam imparciais e neutros em relação a concepções de bem e ao mesmo tempo levar em conta os efeitos da aplicação dos mesmos? Um dos desafios da teoria da justiça de Rawls é justamente propor um modo de considerar o resultado da aplicação dos princípios no procedimento de escolha sem cair num consequencialismo, isto é, sem que se atribua valor moral para uma ação correta considerando somente suas consequências<sup>44</sup>. O filósofo precisa elaborar um procedimento que permita que a escolha dos princípios de justiça independa da contingência, de interesses pessoais e motivações particulares, mas que, ao mesmo tempo, leve em conta o modo como a aplicação desses princípios vai impactar a vida social e individual.

O tipo de teoria desenvolvida por Rawls é melhor compreendido quando recorremos à distinção feita pelo filósofo Antonio Frederico Saturnino Braga entre Deontologismo Rigorista e Não Rigorista (rawlsiano<sup>45</sup>). A interpretação rigorista<sup>46</sup> da deontologia defende que devem ser priorizadas

---

<sup>44</sup> A distinção que Rawls faz entre teleologia e deontologia não diz respeito a concepções que levam em conta ou não as consequências; não é, portanto, uma oposição entre consequencialismo e deontologia.

<sup>45</sup> Braga considera que Rawls é o teórico que “articulou de forma mais abrangente, sistemática e influente” a posição por ele definida como deontologismo não rigorista ou kantismo não rigorista (Cf. BRAGA, 2011, p. 68; 222).

regras rígidas e inflexíveis, cuja validade não depende de qualquer consideração sobre a justiça (ou a bondade), em termos de conteúdo, do estado de coisas resultante (BRAGA, 2011, p. 25).

Nesta perspectiva, não importam quais os efeitos que a aplicação da lei moral terá sobre a vida das pessoas mais diretamente afetadas pelo ato por ela orientado, nem mesmo “as consequências indiretas, remotas e de longo prazo” (BRAGA, 2011, p. 66-67). Se o objeto de estudo da moral e da política tem a ver com a relação necessária entre as pessoas e entre instituições sociais e cidadãos, uma reflexão que despreze totalmente o resultado de como as ações morais e políticas afetam a vida e o cotidiano dos indivíduos não é plausível.

A aversão pelas consequências que resultam da aplicação de princípios de justiça torna indefensável não só o utilitarismo como também o deontologismo rigorista. Braga sugere que se deve interpretar a deontologia proposta por Rawls levando em conta a preocupação com a natureza e o impacto das consequências da ação moral e das instituições, ou seja, uma deontologia não rigorista, cuja principal característica é introduzir no procedimento para escolha de princípios considerações acerca das circunstâncias, consequências e conteúdos da vontade<sup>47</sup>. Cito:

“Esta é uma forma de deontologismo que procura se caracterizar não pela priorização do dever negativo em relação ao dever de

---

<sup>46</sup> De acordo com Braga, a interpretação rigorista tem origem em um “modo equivocado de entender e aplicar a fórmula da lei universal que Kant apresenta na segunda seção de seu livro [*Fundamentação da metafísica dos costumes*]. Chamarei essa interpretação equivocada de interpretação “formalística” da fórmula universal” (BRAGA, 2011, p. 65). Uma das características dessa interpretação é que ela “evita introduzir [no procedimento de universalização] conteúdos da vontade, de forma a garantir a uniformidade dos resultados do procedimento” (BRAGA, 2011, p. 69).

<sup>47</sup> A formulação do deontologismo não rigorista feita por Braga emerge da interpretação não-formalística do procedimento de universalização do imperativo categórico baseada na interpretação não-universalista de Onora O’neill, a qual permite que a consideração das circunstâncias e das consequências seja incluída no teste de universalização. A interpretação não-formalística do procedimento de universalização precisa cumprir duas condições: (i) admitir a consideração das circunstâncias e consequências do ato ou máxima em questão. (ii) produzir resultados de coloração deontológica. Para essa condição ser cumprida, Braga sugere que o indivíduo-legislador seja colocado na base do procedimento tenha natureza deontológica, isto é, que sua imparcialidade e neutralidade tenham sido configuradas segundo critérios deontológicos (Cf. BRAGA, 2011, p. 193-199).

promover um estado de coisas justo, em termos de conteúdo, mas por um modo próprio de determinar qual é esse estado de coisas – forma distinta daquela adotada pelo consequencialismo em sentido estrito. Desse modo, o debate sobre o (maior ou menor) valor que devemos atribuir às liberdades e aos direitos individuais – assim como a uma distribuição mais igualitária (ou, ao contrário, a uma distribuição mais centrada na noção de maximização) – depende da discussão sobre qual estado de coisas deve ser considerado justo, ou bom, em termos de conteúdo” (BRAGA, 2011, p. 25-26).

A deontologia desenvolvida por Rawls está comprometida com a realização de um determinado estado de coisas que possa ser desejado ou aceito pelos sujeitos: uma ordem política e social razoavelmente justa e capaz de sustentar-se a si mesma. O resultado da aplicação do princípio de justiça deve ser tal que o agente moral, depois de avaliar as consequências da aplicação, e “considerando os conteúdos da sua vontade, ele desejaria, ou aceitaria, ser tratado segundo esse mesmo princípio” (BRAGA, 2011, p. 194).

Como, então, incluir a consideração sobre as consequências e sobre os conteúdos da vontade no procedimento que define os princípios de justiça moralmente corretos?<sup>48</sup> Como essa concepção de deontologismo não-rigorista especifica e relaciona o correto e o bem, já que não concebe o bem como um tipo de satisfação racional que deva ser maximizada pelas instituições sociais?

Veremos a seguir que o deontologismo rawlsiano está estruturado a partir de dois âmbitos logicamente independentes: o âmbito da especificação do procedimento de escolha e seleção dos princípios e o âmbito dos motivos morais do agente para agir a

---

<sup>48</sup> De acordo com Sharon Krause, o conceito de racionalidade usado por Rawls incorpora preocupações afetivas. A racionalidade das partes na PO, mesmo sendo concebida como uma racionalidade instrumental, isto é, a busca de meios para satisfazer determinados fins, “faz referência a uma ou mais concepções de bem, se interpretada de modo estrito como interesses individuais ou, de forma mais ampla, como as convicções ou valores pessoais que orientam o plano de vida de alguém. (...) o papel da racionalidade dentro do procedimento indica que quando refletimos sobre justiça não podemos fazer isso sem desejar o bem”. Nesse sentido, ter uma concepção de bem significa “ter uma ligação afetiva com tal concepção ou um desejo de realizá-la; então, quando somos racionais estamos também desejando” (Cf. KRAUSE, 2008, p. 29-30).

partir dos princípios escolhidos no âmbito precedente. O primeiro âmbito delibera sobre a concepção de justiça, supondo a publicidade e a obediência estrita – ou seja, que o(s) princípio(s) selecionados e as exigências que eles impõem aos indivíduos seriam conhecidos por todos e as pessoas e instituições de fato agiriam de acordo com princípios escolhidos. No segundo âmbito, verifica-se primeiro qual estado de coisas é engendrado pela aplicação dos princípios (que tipo de sociedade resulta e que tipo de senso de justiça é gerado) para em seguida questionar se o agente moral aceitaria ou desejaria, levando em conta os conteúdos de sua vontade, viver sob o estado de coisas resultante da aplicação da concepção de justiça escolhida no primeiro âmbito.

### 3.4 A estrutura do Deontologismo rawlsiano

Com base na terminologia do livro de Charles Larmore, *The Morals of Modernity* (1996), Braga apresenta a oposição entre concepção “atrativa” da ética e concepção “imperativa”<sup>49</sup>. A primeira relacionada às concepções que prevaleceram na época antiga e medieval, períodos em que a questão da realização pessoal era vista como um objeto legítimo de discussão racional e pública, como algo que dizia respeito a todas as pessoas. A realização do ideal de vida boa ocorria na esfera pública, e não no âmbito da vida privada. As éticas atrativas se caracterizam por considerar que nossa obrigação com os outros estão ancoradas, antes de tudo, num ideal comum de realização pessoal que tem pretensão de validade objetiva, compartilhado por todos os indivíduos. Esse ideal de realização pessoal pode ser entendido como uma “concepção unitária e abrangente do bem”: unitária por compreender apenas um único

---

<sup>49</sup> Na introdução a sua *História sobre a Filosofia Moral*, Rawls apresenta uma diferença entre a filosofia moral clássica e a moderna, e o faz com base nos *Métodos da Ética* de Sidgwick. Em *Morals of Modernity*, Larmore parte desta distinção feita por Sidgwick para explicar as éticas atrativas e imperativas. Rawls reconhece que existe uma diferença no modo como a filosofia moral antiga e moderna compreendem os conceitos (e a hierarquia entre eles) de correto e de bem, no entanto supõe que essa diferença talvez não seja tão profunda, “mas simplesmente uma questão do vocabulário utilizado para articular e ordenar o domínio moral. Esse vocabulário pode ter sido determinado por acidentes históricos, e um exame mais atento revela que as duas famílias de conceitos expressas por esses vocabulários são equivalentes, no sentido de que quaisquer ideias morais que pudermos expressar em uma família, poderemos também expressar na outra, ainda que sem a mesma naturalidade” (HFM, p. 5).

bem que é publicamente compartilhado e abrangente pois é esse único bem que determina o sentido da vida de cada cidadão e da própria comunidade como um todo (Cf. BRAGA, 2011, p. 40).

Do ponto de vista da ética atrativa, agir de modo correto com os outros significa um passo dado com vistas a uma vida bem sucedida; antes de ser uma obrigação para com os demais, faz parte da realização do agente. Ou seja, a ideia abrangente do bem (realização pessoal) tem primazia sobre as noções de justo e de justiça. De um modo resumido, a ética atrativa tem duas características principais:

[i] Antes de ser algo obrigatório e correto, o comportamento é pessoalmente bom, favorável à realização pessoal do próprio agente a quem se impõe;

[ii] Para determinar o conteúdo do dever da justiça em diferentes situações e diferentes contextos, o agente toma como ponto de partida o conceito do que é bom para ele; mais precisamente, para saber o que deve fazer, ele toma como base a ideia daquilo que o homem realizado (virtuoso) faria – em vez de uma noção para com os outros, pura e simplesmente (BRAGA, 2011, p. 31).

Do ponto de vista da estrutura da ética atrativa, uma característica relevante é que não conseguimos separar com clareza os âmbitos da especificação de qual é a ação eticamente recomendável do âmbito dos motivos para praticar a ação moral. Isso ocorre porque a noção de bem – compreendido como uma concepção abrangente e unitária – desempenha simultaneamente duas funções: a de determinação da ação eticamente recomendável e de motivo para praticar tal ação. Isso significa que ao refletir sobre qual é curso de ação seguir, o ponto de vista utilizado é aquele do próprio agente, a partir do que ele considera ser o seu verdadeiro bem. E como se trata de buscar o bem supremo do próprio agente, é isto que o motiva a agir corretamente. Basta que o agente tenha um conhecimento do verdadeiro bem para agir eticamente (cf. BRAGA, 2011, p. 42-43).

Durante a modernidade, o problema da realização pessoal e da felicidade é colocado em outra perspectiva: passa a ser visto como uma questão essencialmente privada – e não mais pública – vinculada a tendências e preferências particulares e variadas. As

reflexões éticas que se desenvolvem nesse período estão atentas ao fato de que os indivíduos modernos passaram a atribuir mais valor à satisfação de seus desejos e projetos no âmbito da vida privada e a não aceitar que assuntos pertinentes a sua realização pessoal estejam submetidos ao controle arbitrário e à interferência externa, seja do corpo social, seja do Estado. O ponto de partida da reflexão ética passa a ser o dever incondicionado que as pessoas têm umas para com as outras. Isso implica na necessidade de se formular um imperativo do que é correto a partir de um ponto de vista imparcial e impessoal em relação às várias concepções de bem próprio que os indivíduos venham a adotar. É neste contexto que surge a concepção imperativa da ética. Ao contrário das éticas atrativas, nas quais a função de determinação e a de motivação estão amalgamadas em uma única concepção de bem, nas éticas imperativas essas funções representam dois âmbitos logicamente distintos. A ideia de bem privado como algo socialmente valioso tem papel fundamental nessa mudança de perspectiva, é com vistas a protegê-lo de ingerências que se desvincula a definição do correto da noção de realização pessoal (Cf. BRAGA, 2011).

A concepção imperativa da ética se caracteriza por duas teses interligadas:

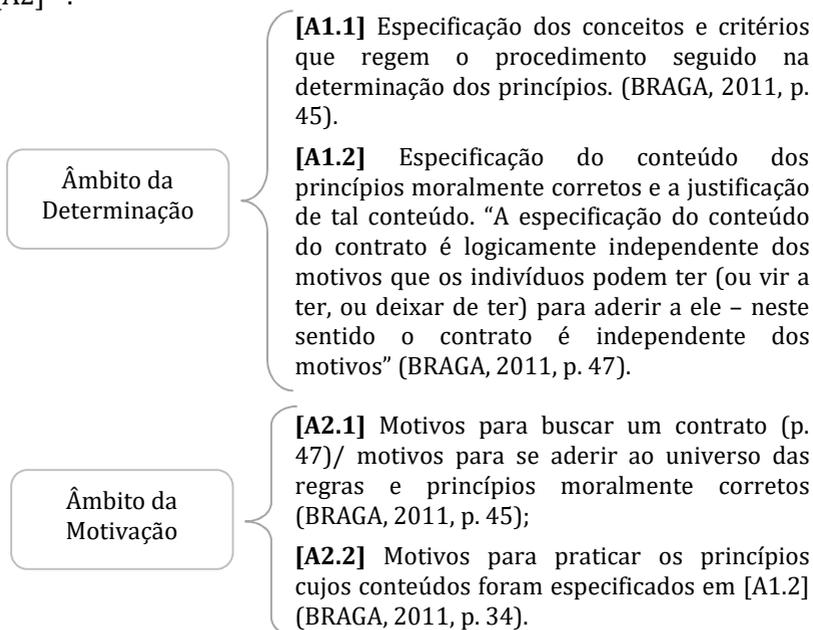
**[1]** a ideia de que a todo indivíduo impõem-se, para com os outros, certas obrigações cuja validade objetiva é logicamente independente da noção de realização pessoal que cada um possa eventualmente assumir;

**[2]** a tese de que a determinação do conteúdo das ações obrigatórias é logicamente independente da questão de se (e em que medida) tais ações promovem (ou não) a realização pessoal do próprio agente de que são exigidas (BRAGA, 2011, p. 33-34).

Na primeira tese constata-se o deslocamento do objeto sobre o qual se busca a validade objetiva. Enquanto nas éticas atrativas (antigas e medievais) o objeto legítimo de deliberação moral era o de ação boa para o próprio agente (com vistas a sua realização pessoal, dado um ideal de realização humana com pretensão de validade objetiva), nas concepções imperativas é o conceito de ação obrigatória que ocupa essa posição. Nesse sentido, as obrigações que os indivíduos impõem-se reciprocamente não podem estar

condicionadas à satisfação de interesses e projetos particulares que cada indivíduo pode ter. Para que algo possa, então, ser considerado justo e correto de se impor a todos, não pode estar condicionado ao que é bom ou vantajoso para um indivíduo ou um grupo de indivíduos específicos. Como implicação da primeira tese, temos que a especificação do conteúdo das obrigações recíprocas também não está condicionada à promoção dos interesses privados do indivíduo. Essas duas teses introduzem uma das características essenciais da concepção imperativa que é a distinção da estrutura argumentativa da teoria moral em dois âmbitos logicamente independentes: o da determinação do conteúdo da ação correta e o da motivação para praticar ação moralmente correta.

Para organizar o raciocínio, vamos elencar os dois âmbitos logicamente distintos que são característica elementar da estrutura das éticas imperativas, âmbito da determinação [A1] e da motivação [A2]<sup>50</sup>:



<sup>50</sup>Para os objetivos de nossa pesquisa, nos referiremos ao âmbito da especificação do procedimento e da determinação do conteúdo do correto como A1, e ao âmbito da motivação moral (ou, motivação para agir a partir de A1) como A2.

A partir da divisão em dois âmbitos logicamente distintos, opera-se um deslocamento da questão da realização pessoal nas éticas imperativas em comparação com as atrativas. Considerando a ideia de que cada indivíduo realiza-se a seu próprio modo, de acordo com seus projetos de vida particulares sem aceitar imposições arbitrárias, busca-se evitar que desejos e interesses particulares e projetos de realização pessoal de uns indivíduos se sobreponham arbitrariamente aos de outros. Para tanto, a noção de bem privado (ou realização pessoal) é retirada do âmbito da determinação do conteúdo das obrigações privilegiando-se nesta esfera outra noção, a de justo e correto, isto é, aquilo que é justo exigir igualmente de todos, aceito como universalmente obrigatório. Por isso,

Na esfera da determinação do conteúdo das regras, a noção-chave continua a ser de dever incondicionado, um dever cuja exigibilidade objetiva não depende dos interesses e desejos privados (variados, variáveis) dos indivíduos (BRAGA, 2011, p. 46).

Como consequência do reposicionamento da realização pessoal no âmbito da vida privada, a noção de bem enquanto autodeterminação dos interesses privados “não pode mais servir como princípio de determinação das regras e atos moralmente recomendáveis” (BRAGA, 2011, p. 44).<sup>51</sup> Esta noção de bem é deslocada para o âmbito da motivação, quando se entende que a busca pela realização de interesses ou bens privados podem ser motivos para aderir ao contrato.

A definição do que é moralmente recomendável não depende mais de uma deliberação pública prévia sobre qual seja o interesse individual que cada pessoa deve seguir, nem quais projetos de vida consideram valioso realizar, tal como nas éticas atrativas. No caso da teoria da justiça como equidade, a ideia de bem utilizada neste âmbito será a de bens primários, isto é, bens que se presume que todos os seres humanos<sup>52</sup> desejariam para poder realizar um plano

---

<sup>51</sup> Um das consequências do reposicionamento do objeto da deliberação moral é que nesse âmbito da teoria imperativa – da determinação do conteúdo do correto – a noção de bem utilizada não será mais a do bem compreendido como a realização pessoal vinculada a contextos particulares.

<sup>52</sup> No Liberalismo Político Rawls passa a considerar que os bens primários são aqueles que cidadãos racionais e razoáveis que desejam, e não os seres humanos em geral.

racional de vida. Durante o procedimento de escolha dos princípios de justiça deve prevalecer uma postura neutra em relação aos diversos interesses, inclinações e circunstâncias particulares dos indivíduos, o único a ser considerado é o interesse por bens primários. As partes devem manter-se neutras e imparciais em relação a qualquer tipo de motivação que poderia levá-las a buscar vantagem com relação a interesses específicos. Por isso, no primeiro âmbito (A1) das éticas imperativas o objetivo é especificar o procedimento para determinar e justificar qual é o conteúdo apropriado da ação moralmente correta, “entendida como aquela que pode ser igualmente exigível de todos, deve ser racionalmente aceitável por todos” (BRAGA, 2011, p. 44) independente de qual seja concepção de bem pessoal que o indivíduo venha endossar.

Se, por um lado, a separação em dois âmbitos logicamente distintos tem como vantagem resguardar o bem privado de violações que o Estado e o corpo social podem cometer contra o indivíduo, por outro, torna mais delicada a questão da motivação moral. Na esfera da motivação, passa-se a lidar com o problema de como um dever incondicionado (à promoção de interesses particulares) vai condicionar os indivíduos a seguirem-no. Como o indivíduo pode se sentir motivado a aceitar ou agir a partir de princípios que, na esfera da determinação, não estão condicionados à promover o bem particular dele? O que motiva o indivíduo a concordar com o procedimento do contrato e a aceitar agir a partir dos princípios escolhidos na situação contratual?

Esta separação de âmbitos engendra aquela que é tensão moral fundamental da filosofia política moderna: como vincular o que é politicamente bom (ou justo) - ou seja, aceitável e desejável por todos - com aquilo que o indivíduo considera como bom para si. O desafio do deontologismo não-rigorista é introduzir no procedimento conteúdo volitivo, avaliação de consequências e circunstâncias. Por que as partes que estão na situação contratual desejariam o estado de coisas gerado pela aplicação dos princípios? O estado de coisas gerado tem que ser congruente, ou pelo menos, não incompatível com o que as partes consideram um bem, tanto do ponto de vista individual quanto coletivo.

A questão que nos trouxe até este ponto era saber como, do ponto de vista da estrutura da teoria moral, o deontologismo rawlsiano insere no raciocínio moral considerações sobre

conteúdos da vontade dos indivíduos e sobre as consequências da aplicação dos princípios de justiça no procedimento de escolha da concepção de justiça sem com isso cair num consequencialismo. A resposta é que o faz distinguindo dois momentos logicamente independentes do contrato, no primeiro âmbito trata da escolha (provisória) de princípios de justiça, no segundo, pretende confirmar se os princípios previamente selecionados, ao serem aplicados pelas instituições, não produzem padrões de inclinações psicológicas prejudiciais ao sistema de cooperação social.

Esse arranjo contratual feito em duas partes logicamente independentes coloca o problema da escolha de um modo específico. E o esforço da TJ “é resolver e defender uma solução para o problema da escolha apresentado por esse arranjo contratual [feito em duas partes]” (MIKHAIL, 2010, p. 7). Veremos agora esses dois âmbitos são vislumbrados na arquitetura de *Uma Teoria da Justiça*.

### **3.5 Localizando o problema da estabilidade na estrutura argumentativa da teoria moral**

Vimos na seção anterior que uma das características das éticas imperativas, no interior da qual a teoria rawlsiana se insere como o mais refinado exemplo de deontologismo não-rigorista, diz respeito à estrutura moral da teoria separada em dois âmbitos logicamente independentes: o âmbito da especificação do procedimento e do conteúdo da concepção de justiça e o âmbito da motivação.<sup>53</sup> O primeiro âmbito diz respeito à noção moral de correto e está vinculado à justificação normativa e a outra diz respeito ao ponto de vista da noção moral de bem, vinculada à formulação descritiva da justiça como equidade.

Acreditamos que é importante esclarecer como a teoria da justiça como equidade realiza essa distinção, não apenas porque nos auxiliará a compreender a localização do argumento da estabilidade, como também pode nos oferecer uma visão um pouco

---

<sup>53</sup> De acordo com Mikhail, Rawls, em sua tese de doutorado, *A Study of the Grounds of Ethical Knowledge: Considered with Reference to Judgements on the Moral Worth of Character* (1950), concebe que a estrutura de uma teoria ética está dividida em duas partes principais: explicação e justificação (*explication and justification*) (MIKHAIL, 2011, p. 3).

mais apurada da arquitetura da teoria. Esperamos também que nos auxilie a compreender como os dois diferentes pontos de vista – um relacionado à noção de correto e o outro a de bem – estão estruturados e como se articulam na estrutura da teoria e nas estratégias de justificação da mesma.

Rawls informa a seu leitor que a teoria da justiça como equidade,

(...) assim como outras visões contratualistas, consiste em duas partes: (1) uma interpretação da situação inicial e do problema da escolha que nela se apresenta e (2) um conjunto de princípios que, segundo se procura demonstrar, seriam acordados (TJ, §3, p. 18-19).

E deixa claro que estas duas partes são independentes entre si:

Pode-se aceitar a primeira parte da teoria (ou alguma variação dela), mas não a segunda, e vice-versa. O conceito da situação contratual inicial pode parecer razoável, embora os princípios propostos possam ser rejeitados (TJ, §3, p. 18-19).

A primeira parte consiste em especificar qual seria a situação inicial<sup>54</sup> filosoficamente mais apropriada a partir da qual os princípios de justiça seriam escolhidos, isso inclui especificar elementos como a natureza das partes e suas capacidades morais, qual o objeto e as circunstâncias do acordo, o grau de conhecimento das circunstâncias particulares, quais são as concepções de justiça que dispõem para escolher, as restrições ao conceito de correto, etc<sup>55</sup>.

Nesta primeira parte, o objetivo é mostrar que as condições e restrições da PO são preferíveis em relação a outras interpretações disponíveis, pois são (i) razoáveis e (ii) moralmente defensáveis, no

---

<sup>54</sup> “Toda doutrina do contrato social precisa explicar a situação em que o contrato, seja ele histórico ou não histórico, deve ser feito. Chamemos essa situação de situação inicial. Para desenvolver uma doutrina de modo minimamente claro, devem ser preenchidos expressamente numerosos aspectos dessa situação” (CHFP, p. 17).

<sup>55</sup> Rawls apresenta em lista os elementos da posição original cuja descrição o filósofo supõe que estarão mais em acordo com pontos fixos e juízos ponderados (Cf. §25, p. 177-178).

sentido que os princípios que seriam escolhidos nesta situação, sejam quais forem, são normativamente válidos, aceitáveis do ponto de vista moral. O próprio procedimento para a escolha dos princípios deve ser justo, isso constitui a ideia que Rawls denomina justiça procedimental pura. Uma vez escolhida a situação inicial, passa-se à segunda etapa da teoria, que consiste em, analisar se os princípios de justiça seriam acordados a partir da PO.

Assim é que, em *Uma Teoria da Justiça* são desenvolvidas duas estratégias de justificação para defender que os princípios da justiça como equidade são uma concepção justiça mais viável que o utilitarismo, as quais consistem em: (i) apresentar argumentos decisivos do ponto de vista da posição original e (ii) testar os princípios de justiça e suas consequências por meio de um confronto com nossos juízos ponderados (procedimento do equilíbrio reflexivo)<sup>56</sup>. Quando o leitor se depara com o estudo da Parte III de TJ, percebe que não há uma delimitação mais precisa entre essas duas estratégias de justificação com relação ao argumento da estabilidade.

Sugerimos que para resolver essa dificuldade é preciso esclarecer duas questões que, ou ainda não estão evidentes e bem delimitadas na TJ - sendo melhor explicitadas nos trabalhos posteriores - ou cuja argumentação foi apresentada de forma esparsa ao longo do livro. No primeiro caso, trata-se da questão de esclarecer quais os pontos de vista que podem ser adotados na teoria, e, no segundo, é preciso evidenciar a estrutura argumentativa da posição original na TJ e localizar dentro dela o argumento (ou parte dele) da estabilidade, bem como evidenciar a estrutura do Equilíbrio Reflexivo.

A teoria da justiça como equidade concebe três diferentes pontos de vista<sup>57</sup> a partir dos quais um mesmo sujeito (no caso, a

---

<sup>56</sup> “ (...) os dois princípios são pelo menos uma concepção plausível da justiça. A questão, porém, é como defendê-los de maneira mais sistemática. Há algumas coisas a fazer. Podemos calcular suas consequências para as instituições e observar suas implicações para a política social fundamental. Desse modo, são testados por um confronto com nossos juízos ponderados acerca da justiça. A parte II deste livro se dedica a isso. Mas podemos também tentar encontrar argumentos em favor dos dois princípios que sejam decisivos do ponto de vista da posição original” (TJ, §26, p. 186).

<sup>57</sup>Estes três pontos de vista já estão presentes na teoria da justiça como equidade desde TJ, no entanto, a distinção clara entre eles começa a ser delineada em *Fairness to Goodness* (1975), em que Rawls chama a atenção para a diferença entre o ponto de vista das partes na PO e dos membros da SBO. A partir de *Kantian Constructivism in Moral Theory* (1980) a

pessoa moral) pode considerar a si mesmo: (i) o ponto de vista das partes na PO; (ii) o dos cidadãos em uma SBO; e (iii) o nosso – o seu e o meu, nós que estamos formulando a ideia de justiça como equidade e examinando-a como a base de uma concepção de justiça que possa fornecer uma compreensão apropriada da liberdade e da igualdade.

A argumentação em favor da estabilidade ocorre, como veremos, na segunda etapa do raciocínio desenvolvido na PO. Nesta segunda etapa, as partes devem, com o auxílio da psicologia moral e teoria social, confrontar os princípios escolhidos na primeira etapa com o ponto de vista dos cidadãos da SBO. Consideramos importante dar destaque a essa dupla etapa de justificação dos princípios na PO, pois não raro é ignorada e acaba-se por tratar a primeira parte como se fosse o argumento inteiro a partir da posição original. Aprofundaremos a análise da PO na próxima seção.

Somente nos escritos posteriores à TJ fica mais evidente que a escolha feita pelas partes do ponto de vista da PO deve ainda passar pelo crivo de pessoas concretas, como você, leitor, e eu, após uma reflexão cuidadosa (teste do equilíbrio reflexivo). Este é o ponto de vista final<sup>58</sup> – o seu e o meu - a partir do qual qualquer concepção política da justiça como um todo deve ser avaliada (Cf. PL, I, §4.6, p. 33). Essa ideia já está presente em TJ, mas não chega a ser explorada com clareza com que está exposta no PL.

Apesar de já estarem presentes em TJ as ideias de uma argumentação que seja decisiva do ponto de vista da PO por um lado, e de um confronto de tais argumentos com nossos juízos ponderados sobre a justiça de outro, a articulação entre elas na teoria aparece um pouco confusa. Em diversas passagens o argumento da segunda etapa argumentativa da PO, ou melhor, o ponto de vista das partes ao analisarem os argumentos da

---

distinção entre os três pontos de vista é apresentada de modo sistemático, o qual é reproduzido no PL (Cf. FG, p. 274; KCMT, p. 320-321; LP, I, §4.6, p. 32).

<sup>58</sup> Em que pese seja o ponto de vista final a partir do qual a teoria é avaliada, o procedimento do equilíbrio reflexivo nunca se encerra por completo. Aliás, sequer há certeza “de que é possível chegar a esse estado” (TJ, §9, p. 59). “O equilíbrio reflexivo (...) é um ponto no infinito que jamais podemos alcançar, embora possamos nos aproximar mais dele no sentido de que, por meio da discussão, nossos ideais, princípios e julgamentos parecem-nos mais razoáveis e mais bem fundamentados do que antes” (LP, p. 455).

estabilidade se confunde com o ponto de vista final – o meu e o seu. Vejamos isso no trecho abaixo:

[i] A posição original é usada em primeiro lugar para definir o conteúdo da justiça, os princípios que a definem. [ii] Somente depois disso a justiça é vista como parte do nosso bem e vinculada a nossa sociabilidade natural (T), §87, p. 721). (ênfase adicionada)

Em (i) Rawls parece tratar a definição do conteúdo da justiça (o que seria tarefa da primeira etapa de argumentação, sendo que a segunda etapa implica em confirmação de tal concepção – mas uma confirmação feita do ponto de vista das partes na PO) como se fosse o raciocínio completo da posição original. O que implica que fiquemos em dúvida se (ii) está se referindo ao segundo estágio do argumento da PO (no qual partes também precisam se questionar se os princípios da justiça podem fazer parte do bem e ser vinculados à sensibilidade moral dos cidadãos da SBO) ou ao segundo tipo de argumento oferecido para justificar a teoria da justiça como equidade: o equilíbrio reflexivo entre nossos juízos ponderados sobre a justiça e os argumentos apresentados a partir da PO. A argumentação em favor dos princípios a partir da PO, embora tenha um papel importante na justificação dos princípios, não pode ser considerada isoladamente. Deve passar ainda pelo procedimento do Equilíbrio Reflexivo. Como afirma Rawls, “a justificação é uma questão do apoio mútuo das muitas ponderações, de tudo se encaixar em uma única visão coerente” (T), § 87, p. 715).

A estratégia profunda da teoria da justiça de Rawls não é sua confiança na posição original tomada isoladamente, mas a estratégia de buscar o equilíbrio reflexivo entre princípios os quais nós provisoriamente (*tentatively*) acreditamos que seriam escolhidos na posição original e nossos juízos ponderados (O’NEIL, 1998, p. 40).

O terceiro ponto de vista que pode ser adotado na teoria da justiça como equidade – o meu e o seu - é uma tentativa de responder à uma frequente objeção ao contratualismo: porque deveríamos aceitar o procedimento do contrato e porque nós deveríamos aceitar como fundamentais princípios de justiça construídos em um acordo hipotético? Qual a força vinculativa que princípios construídos em um acordo hipotético e altamente

abstrato teriam para nós, pessoas concretas, interesses e situações tão distintos entre si? Para Maffettone, essa distinção entre o consenso ideal das partes na Posição Original – “pressuposto sobre o qual a teoria é fundada - e o consenso empírico dos cidadãos, que pressupõe uma situação fática, é frequentemente intangível ou confusa no texto de Rawls” (MAFFETTONE, 2004, p. 569).

A justificativa fundamental dos princípios não é que eles foram escolhidos em uma situação contratual equitativa hipotética, mas sim que esses princípios (e também o procedimento) estão em equilíbrio reflexivo com os nossos juízos mais refletidos sobre a justiça.

Investigaremos agora como a estrutura da teoria moral dividida em dois âmbitos repercute na estrutura do Equilíbrio Reflexivo. Apresentaremos apenas um esboço das duas etapas de sua estrutura. Julgamos importante este esclarecimento, ainda que sem a análise mais profunda da qual deveria estar acompanhado.

Muito embora o método<sup>59</sup> do Equilíbrio Reflexivo tenha um papel importante na teoria moral, Rawls não oferece uma exposição mais detalhada do mesmo. As explicações do Equilíbrio Reflexivo são dadas nos § 4 (*A posição original e sua justificação*) e no § 9 (*Algumas observações acerca da teoria moral*).

John Mikhail chama a atenção para o fato de que há uma diferença crucial entre as explicações, embora Rawls nos leve a pensar, inicialmente, que no §9 está reforçando a explicação dada no §4, pois a definição de equilíbrio reflexivo apresentada é precedida por “conforme já vimos” (*as we have seen*) (TJ, §9, p. 58). (Cf. MIKHAIL, 2011, p17). Vejamos agora como o equilíbrio reflexivo é explicado em cada um destes momentos.

No §4, o Equilíbrio Reflexivo é introduzido para justificar qual seria a interpretação da situação inicial mais apropriada a partir da na qual seriam escolhidos os princípios de justiça. A interpretação mais adequada é aquela em que nossos juízos ponderados estão em equilíbrio reflexivo, sendo este concebido como

---

<sup>59</sup> Mikhail nota que em TJ, o equilíbrio reflexivo ainda não é, literalmente, caracterizado como um método ou estratégia de justificação. Nas poucas menções feitas a ele, é apresentado um “estado de coisas” (*state of affairs*) (Cf. Mikhail, 2011, p. 14). Em IMT, é literalmente descrito como “método” e “procedimento”, que podemos entender como método ou procedimento utilizado para alcançar um determinado estado de coisas. No LP, é tratado como “critério”, “reflexão cuidadosa”.

(...) um estado de coisas (*state of affairs*) que é alcançado uma vez que a descrição da situação inicial tenha alcançado duas condições:  
 [i] que expresse condições razoáveis, e  
 [ii] que gere princípios que com combinem com nossos juízos ponderados devidamente ajustados e apurados (TJ, §4, p. 24-5). (ênfase adicionada)

Uma vez que - após um processo de ajuste mútuo entre princípios (ou seja, construções teóricas) e juízos (isto é, a afirmação de certas coisas sobre a justiça) – se chega a esse estado de coisas, a Posição Original está provisoriamente justificada como sendo a interpretação mais adequada do *status quo* inicial da teoria. E, em sendo um caso de justiça procedimental pura, os princípios escolhidos nesta posição precisam ser levados a cabo, pois “um processo equitativo só traduz sua equidade no resultado quando é realmente levado a cabo” (TJ, §14, p. 105).

O que se deseja nessa explicação do equilíbrio reflexivo é que os princípios de justiça sejam justificáveis a partir da noção de justiça procedimental pura<sup>60</sup>, a qual é definida do seguinte modo:

A justiça procedimental pura, em contraste, verifica-se quando não há um critério independente para o resultado correto: em vez disso [i] existe um procedimento correto ou justo, seja qual for, [ii] contanto que se tenha aplicado corretamente o procedimento (TJ, §14, p. 104).

Podemos dizer, então, que a explicação do equilíbrio reflexivo apresentada no §4 trata de justificar as condições iniciais de possibilidade de um procedimento justo. A condição para que seja considerado justo é que o resultado de sua aplicação também tenha gerado consequências justas. Os princípios precisam ser levados a cabo, e isso é feito aplicando-os na estrutura básica de uma sociedade bem-ordenada.

Ao tratar novamente do Equilíbrio Reflexivo no §9, Rawls enfatiza o aspecto descritivo da teoria moral.

É possível que se considere a teoria moral, a princípio (e saliente a natureza provisória desse enfoque), como a tentativa de descrever

---

<sup>60</sup> Cf. essa sugestão em MIKHAIL, 2011, p. 17.

a nossa capacidade moral; ou, no caso em questão, pode-se considerar a teoria da justiça uma descrição de nosso senso de justiça (TJ, §9, p. 56).

Além desta passagem ora citada, há várias outras no mesmo parágrafo que mencionam que um dos papéis da teoria moral é oferecer uma descrição das nossas capacidades morais. Quando os princípios escolhidos na PO são compatíveis com nossos juízos ponderados, eles descrevem o nosso senso de justiça (cf. TJ, §9, p. 58).

Por isso, podemos sugerir que explicação oferecida no §9, atribui ao equilíbrio reflexivo a tarefa de averiguar o resultado do procedimento para conferir sua correção. Isso implica examinar se princípios que foram escolhidos (numa situação inicial justificada em acordo com o §4) descrevem o nosso senso de justiça.

Cada uma destas explicações do equilíbrio reflexivo se refere a um dos âmbitos da teoria moral e que Rawls está tentando combinar,

Enquanto que no §4 Rawls parece estar discutindo o problema *normativo* de justificar princípios de justiça, o §9 parece ser o problema *descritivo* de descrever e explicar o senso de justiça (MIKHAIL, 2011, p. 17).

Podemos dizer que boa parte do empenho de Rawls em TJ é para tentar tornar congruentes, ou, pelo menos, não incompatíveis, dois campos na estrutura da teoria moral:

uma psicologia moral descritiva que explica o que acontece quando nos engajamos em reflexões morais e políticas, e uma teoria da normatividade, que explica porque os padrões que alcançamos através de tal reflexão devem ser tratados como autoritativos (imperativos) (FRAZER, 2007, p. 759).

Existe uma tensão entre os âmbitos normativo, vinculado correto, e descritivo, vinculado à noção de bem. Com o argumento da estabilidade, Rawls pretende tornar compatíveis estes âmbitos. Mais do que isso, para que sua teoria seja viável, que uma sociedade política justa e duradoura possa existir, ainda que apenas do ponto de vista ideal, Rawls necessita demonstrar que a teoria normativa da justiça não é incompatível com a com as motivações das pessoas.

### 3.6 As duas etapas argumentativas da Posição Original

Uma das dificuldades enfrentadas pelo leitor ao se propor a tarefa de investigar os argumentos em favor da estabilidade em TJ é lidar com argumentos que foram desenvolvidos de modo intermitente e pouco claro na Parte III. Acreditamos que uma das maneiras de organizar as ideias ali expostas é esclarecer o lugar em que os argumentos em favor da estabilidade se localizam na estrutura argumentativa da posição original.

Não basta informar apenas que tal questão é tratada de modo mais aprofundado na Parte III de referido livro, até porque essa informação diz respeito à estrutura do livro, a qual pode não corresponder necessariamente à estrutura da teoria. Perguntar pela localização e enquadramento da estabilidade na estrutura argumentativa da posição original, uma vez que numa primeira leitura de TJ temos uma impressão de que as reflexões acerca da estabilidade ocorrem em algum momento depois de já escolhidos os dois princípios da justiça e de encerrado o raciocínio a partir da PO<sup>61</sup>, como se em algum momento depois de finalizado o experimento, ou em um ponto de vista em que as partes não estejam submetidas às condições e restrições impostas pela da PO.

Esta impressão é fortalecida quando nos guiamos pelos textos dos principais comentários tanto à PO quanto à estabilidade e a motivação moral na TJ: neles é comum que a construção da PO e o raciocínio que conduz aos dois princípios de justiça sejam tomados como todo o procedimento e única função do experimento mental, e que o argumento da estabilidade seja tratado como uma teoria à parte, como algo desvinculado da Posição Original. Edward McClennen, por exemplo, defende a hipótese de que as considerações sobre a estabilidade são fortes e decisivas o suficiente, do ponto de vista da escolha racional, para sustentar uma única concepção de justiça, a ponto de tornar dispensável o procedimento da posição original.<sup>62</sup> Sebastiano Maffettone sugere

---

<sup>61</sup> No Capítulo III, dedicado inteiramente à Posição Original, temos relativamente poucas menções diretas à questão da estabilidade. No entanto, numa das poucas menções à estabilidade, Rawls já informa ao leitor que a mesma será “examinada mais minuciosamente no que denominei segunda parte da argumentação (§§ 79-80)” (TJ), §29, p. 224).

<sup>62</sup> “A mais ambiciosa hipótese seria a de que a estabilidade traz consigo benefícios grandes o suficiente para serem decisivos do ponto de vista da escolha racional, e que as

que a teoria da justiça trabalha com três estratégias de justificação para os princípios<sup>63</sup>: posição original, equilíbrio reflexivo e o que ele denomina Teoria da Estabilidade (Cf. MAFFETTONE, 2010, p. 140). Parece, então, que a estabilidade não é um tipo de argumento ou estágio de justificação à parte da PO.

Tal percepção incorreta – de que as considerações sobre a estabilidade estão localizadas em um âmbito em separado da posição original - ocorre quando não se releva uma importante e nem sempre notada característica da PO: sua estrutura argumentativa é dividida em dois estágios. O primeiro trata da escolha provisória dos princípios e o segundo da confirmação da escolha, e somente depois de analisar ambos é que a justificação a partir da PO está concluída. O argumento da estabilidade está localizado no interior da posição original, mais especificamente, na segunda etapa argumentativa. No entanto, quando o leitor ignora tal estágio, é levado a supor que o primeiro estágio “é o argumento completo a partir da posição original, e uma vez que as partes tenham feito suas escolhas o trabalho delas está concluído e elas podem, por assim dizer, encerrar o negócio” (GAUS, 2014, p. 236). Essa perspectiva empobrece a compreensão da justiça como equidade por não contemplar a análise da descrição dos sentimentos morais e da psicologia moral e nos leva a ter uma visão muito simplificada da PO.

A dúvida quanto à localização do argumento da estabilidade surge pelo próprio modo como o problema é apresentado, pois busca resolver é saber se os princípios escolhidos na PO seriam capaz de gerar nos membros da SBO um senso de justiça forte e se do ponto de vista da racionalidade deliberativa, tais membros da SBO afirmariam os princípios de justiça como reguladores de seu plano de vida. A solução para o problema da estabilidade depende, então, que seja levado em conta o ponto de vista dos cidadãos da

---

considerações da estabilidade selecionam uma única concepção de justiça. Se isso puder ser mostrado, então não seria necessário apelar para o argumento da posição original com sua problemática noção sobre o que seria racional escolher em uma situação de incerteza radical” (McClennen, 1989, p. 9).

<sup>63</sup> Pierce Lehnig e T. M. Scanlon defendem que na teoria da justiça como equidade há três etapas ou ideias de justificação: Método do Equilíbrio Reflexivo, Derivação dos princípios a partir da PO e o papel da Razão Pública na SBO (Cf. Lehnig, 2009, 144; SCANLON, 2003, p. 139). Como nosso objetivo inicial é investigar a questão da estabilidade no primeiro livro de Rawls, não nos deteremos aqui na ideia de Razão Pública, desenvolvida a partir do *Liberalismo Político*.

SBO. O que nem sempre fica claro para o leitor é que este ponto de vista dos membros da SBO deve ser levado em conta pelas partes na PO, e isto será feito com levando-se em conta alguns dos conteúdos não vedados pelo véu da ignorância, como a psicologia moral e a teoria social.

A verdade é que o texto de Rawls também contribui um tanto para essa relativa obscuridade quanto à localização do argumento da estabilidade. Somente quando nos detemos em uma leitura mais atenta da intrincada Parte III – ou melhor, das 100 páginas finais do extenso livro – é que fica textualmente mais evidente que o raciocínio na PO está dividido em duas etapas argumentativas. Supomos que ideia da divisão em duas etapas, embora presente desde o início do livro (cf. TJ, §25, p. 175; §29, p. 224), ficou mais clara para o próprio Rawls no momento do desenvolvimento da parte final de TJ, que é onde aparecem mais menções à divisão da argumentação em dois momentos (Cf. TJ, §76, p. 622; §80, p. 654-5; §82, p. 668-671; p. 674; §87, p. 715). Nestas passagens, o autor usa expressões como “estamos na segunda parte da argumentação”, “o processo de duas etapas”, sem, contudo, afirmar expressamente que se trata da segunda etapa da posição original. O trecho a seguir, entretanto, indica que o *locus* do argumento da estabilidade é no interior da PO: “Mas a decisão na posição original depende de uma comparação: permanecendo constantes os demais fatores, a concepção preferida é a mais estável” (TJ, §76, p. 615).

Nos livros posteriores, as duas etapas argumentativas do artifício de representação passam a ser apresentadas de maneira incontestada. No *Liberalismo Político*, a seção específica que trata a questão da estabilidade inicia com a afirmação: “é melhor formular a justiça como equidade em dois estágios”, momento em que é introduzida a seguinte nota de rodapé:

Esses dois estágios correspondem às duas partes do argumento a partir da posição original para sustentar os dois princípios em *Teoria*. Na primeira parte do argumento, as partes escolhem princípios sem levar em conta os efeitos das psicologias especiais, ao passo que, na segunda parte, elas indagam se uma sociedade bem-ordenada seria estável, isto é, se geraria em seus integrantes um senso de justiça suficientemente forte para

contrabalançar as tendências à injustiça. (...) O argumento a favor dos dois princípios de justiça não está completo até que seja demonstrado, na segunda parte, que são suficientemente estáveis, e o esforço de fazer isso se estende até a penúltima seção de *Teoria*, §86 (PL, §2.1, p. 165n7).

Em JF, Rawls reafirma - agora com mais ênfase e já a partir do parágrafo específico sobre a Posição Original - que um aspecto básico do argumento a partir da posição original é que o mesmo é dividido em duas partes (JF, §25.5, p. 124-5; §26, p.129; §54, p. 257-8; §55, p. 262-3), e que ambas “são desenvolvidas dentro da mesma estrutura e estão sujeitas às mesmas condições incluídas na posição original como mecanismo de representação” (JF, §55, p. 265n8).

Temos aqui três informações muito importantes: (i) a posição original possui duas etapas de argumentação; (ii) apesar de a preocupação com a estabilidade estar presente desde o momento da construção da PO, o argumento da estabilidade faz parte da segunda etapa de argumentação do artifício de representação; (iii) o raciocínio a favor dos princípios na PO só se encerra depois de analisadas as duas etapas, ou seja, depois de analisar a estabilidade. E é justamente nesta segunda etapa tão ignorada da posição original que Rawls desenvolve o fundamento da igualdade (Cf. TJ, §77) e os fundamentos da prioridade da liberdade (Cf. TJ, §82).

Vejamos agora como foram desenvolvidas essas duas etapas argumentativas da posição original. Rawls concebe a posição original como um experimento mental que podemos realizar a qualquer momento por meio do raciocínio - desde que respeitadas certas restrições formais<sup>64</sup> - para chegar a um acordo sobre princípios de justiça que devem regular a estrutura básica da sociedade<sup>65</sup>.

A escolha dos princípios de justiça é feita por partes que representam os interesses de pessoas racionais livres e iguais (no

---

<sup>64</sup> Generalidade, universalidade, ordenamento/hierarquização, finalidade e publicidade (Cf. TJ, §23), além do véu de ignorância (§24), que é um conjunto de restrições às informações que as partes podem considerar ou não durante o procedimento de escolha e avaliação dos princípios de justiça.

<sup>65</sup> O foco é na estrutura básica da sociedade, pois é ela que modela as necessidades e aspirações que os cidadãos vêm a ter. É o sistema social ao qual estamos submetidos que define, em parte: (i) o que as pessoas querem ser e (ii) o tipo de pessoa que são (Cf. JF, §23, p. 118; 122).

caso, os interesses dos cidadãos de uma sociedade bem-ordenada) situadas sob o véu da ignorância - expediente utilizado para garantir não apenas que as partes tenham acesso apenas às informações gerais sobre a sociedade e os cidadãos que representam a fim de que a escolha não beneficie ou favoreça objetivos, planos de vida, pessoas ou grupos específicos, como também de evitar que determinadas aspirações e inclinações psicológicas particulares tenham influência sobre a escolha dos princípios.

Rawls toma o cuidado de elaborar um véu espesso o suficiente para evitar que as partes queiram buscar vantagens para si próprias com base em sua colocação social ou vantagens naturais. Por isso, conhecem apenas fatos genéricos da sociedade humana<sup>66</sup>, quaisquer que sejam desde que relevantes para a escolha da concepção, como assuntos de política e princípios da teoria econômica, leis da psicologia humana e a base da organização social. Em relação aos cidadãos representados, as partes devem desconsiderar (i) a que geração, gênero, raça e grupo étnico<sup>67</sup> pertencem, (ii) seu lugar na sociedade, classe ou status social, (iii) ignoram ainda características especiais de sua psicologia, ou se possuem ou não algum talento natural (isto é, se possuem ou não força física ou inteligência); (iv) não sabem qual concepção de bem possuem, nem as particularidades do projeto racional. Quanto à sociedade<sup>68</sup> em que vivem, devem desconsiderar em que posição econômica ou política a mesma se encontra, e o nível de civilização ou cultura que conseguiu alcançar (Cf. TJ, §24, p. 165-167).

Rawls supõe que as partes na posição original têm alguma noção de que desejam se engajar em algo na vida (através de uma

---

<sup>66</sup> Como tamanho, nível de desenvolvimento econômico, estrutura institucional, ambiente natural, etc. (TJ, §31, p. 245).

<sup>67</sup> Apenas no *Liberalismo Político*, Rawls acrescenta que partes devem também ignorar “informações sobre raça e grupo étnico, gênero” (LP, p. 28).

<sup>68</sup> No *Liberalismo Político*, Rawls acrescenta 5 fatos gerais sobre a sociedade bem-ordenada, os quais supomos que as partes na posição original devem levar em consideração: fato do pluralismo enquanto tal, fato do pluralismo razoável, fato da opressão, fato do apoio livre e voluntário (LP, I, §6.2, p. 43-45) e fato do desacordo razoável (LP, II, §2.4, p. 69). Na TJ: “os únicos fatos específicos conhecidos são os que se podem inferir das circunstâncias da justiça (§TJ, §24, 165). Embora conheçam os princípios básicos da teoria social, as partes não têm acesso ao curso da história; não têm informações sobre frequência com que a sociedade assumiu esta ou aquela forma, ou sobre quais tipos de sociedade existem atualmente”, tampouco conhecem o tamanho, nível de desenvolvimento econômico, estrutura institucional e ambiente natural (TJ, §31, p. 245).

carreira ou vocação), bem como da importância das relações pessoais e de que querem preservar suas identidades enquanto membros de vários grupos (políticos, religiosos, étnicos, etc). A motivação primária das partes na posição original é garantir condições favoráveis e seguras para que possam realizar os vários elementos do plano racional de vida que define o que é uma boa vida para elas. (Cf. FREEMAN, 2007b, p. 148). Para tanto, Rawls especifica como bens primários “coisas que todo indivíduo racional presumivelmente quer” (TJ, §11, p. 75), pois possuem utilidade para realizar projetos de vida: direitos e liberdades, oportunidades, renda, riqueza e as bases sociais do auto-respeito. As partes são, então, concebidas como mutuamente desinteressadas, racionais e como não afetadas pela inveja; também “não têm motivações de afeto nem de rancor” (TJ, §25, p. 176), nem de benevolência (TJ, §30, p. 235). Ou seja, “as partes não estão ligadas por vínculos morais prévios” (TJ, §22, p. 155). É esta motivação puramente racional na busca por bens primários que vai orientar as partes no primeiro momento da escolha dos princípios.

Ao supor que as partes na PO não estão motivadas pela inveja Rawls quer garantir que elas não estarão dispostas ao auto-sacrifício, não abririam mão de garantir bens primários a mais para si apenas para evitar que os demais também consigam mais bens para eles. Com isso, acredita que as partes escolherão a concepção de justiça capaz de lhes garantir uma quantidade maior de tais bens, movidas apenas pela racionalidade. As partes possuem um sistema amplo de objetivos<sup>69</sup>, e a suposição é de estão motivadas a promover apenas esses seus objetivos, elas não estão interessadas no que os outros possuem, não querem conceder benefícios e nem impor prejuízos umas às outras (Cf. TJ, §25, p. 176) e “nem ficam deprimidas ao perceber que os outros têm uma quantidade maior de bens primários sociais” (TJ, §25, p. 174). Rawls ressalta que essa

---

<sup>69</sup> “Os princípios de justiça devem ter nexos racionais com a promoção de interesses humanos definidos de maneira ampla” (TJ, §25, 181).

“Agentes racionais tampouco são sempre autointeressados, isto é, seus interesses nem sempre se voltam para eles próprios. (...) Na verdade, agentes racionais podem ter todos os tipos de afeto por pessoas e de vínculos com comunidades e lugares, incluindo o amor à terra natal e à natureza, e podem selecionar e ordenar seus fins de várias maneiras” (LP, II, §1.2, p. 60). Embora os projetos racionais de longo prazo “definam os objetivos e interesses de um eu, não se presume que esses objetivos e interesses sejam egocêntricos e egoístas” (TJ, §22, p. 156).

suposição motivacional que está fazendo diz respeito apenas à motivação das partes na posição original, e não à descrição do modo como as pessoas agem na vida em sociedade. Essa etapa da derivação dos princípios de justiça pretende mostrar que as partes, movidas pelo interesse racional de garantir mais bens primários do que menos, vão acabar escolhendo os dois princípios de justiça como equidade, e não o princípio da utilidade.

O que descrevemos agora foi a primeira etapa do argumento a partir da posição original, a qual, entretanto, é geralmente tomada como o experimento completo. De acordo com Rawls, quando se examina apenas um dos elementos da posição original pode-se ficar com a impressão de que essa concepção de justiça é egoísta, o que é uma ilusão (Cf. TJ, § 25, p. 181). Qual é então, o lugar apropriado para acomodar as reflexões sobre psicologias especiais?

Como dito anteriormente, as pessoas concretas em suas vidas cotidianas não são mutuamente desinteressadas<sup>70</sup>, mas sim são motivadas pelas mais variadas inclinações e sentimentos, tais como vergonha e humilhação, inveja, desejo de dominar e subjugar, e de privar outrem da justiça, dentre outros sentimentos destrutivos. (Cf. TJ, p. 175; JF, p. 288), e que uma teoria da justiça satisfatória e realista deve lidar com essas questões. Para lidar com as críticas de que sua concepção de justiça se afasta muito da realidade, o lugar que Rawls reserva para a análise dos sentimentos e da psicologia moral é segunda etapa<sup>71</sup> do argumento a partir da posição original. Depois que os princípios são provisoriamente escolhidos na primeira etapa, é preciso verificar se o senso de justiça e a psicologia moral engendradas pela aplicação institucional dos princípios escolhidos provisoriamente são capazes de garantir o apoio às instituições justas ou se têm probabilidade maior de despertar e incentivar propensões e inclinações tais que tornem o sistema social incompatível com o bem dos cidadãos (Cf. TJ, §80, p. 655).

É no segundo estágio que se analisa a estabilidade da concepção de justiça. Rawls defende em TJ que a estabilidade em uma

---

<sup>70</sup> “Não há incoerência em supor que removido o véu da ignorância as partes descobrem que têm laços de sentimento e afeição, e que querem promover interesses dos outros e vê-los atingir seus objetivos” (TJ, §22, p. 157).

<sup>71</sup> É importante frisar que as duas partes do argumento “são desenvolvidas dentro da mesma estrutura e estão sujeitas às mesmas condições incluídas na posição original como mecanismo de representação” (JF, p. 265n8).

sociedade democrática depende de que os princípios de justiça política tenham um suporte emocional apropriado, e é por isso que no segundo estágio da Posição Original o raciocínio das partes passa a levar em conta a psicologia moral e teoria social (T), § 75, p. 607), pois precisam considerar como será a psicologia moral dos cidadãos engendrada pelos princípios de justiça escolhidos no primeiro estágio da teoria. Nesse segundo estágio é preciso verificar se as instituições reguladas pela concepção de justiça selecionada na etapa anterior

têm probabilidade de despertar e incentivar propensões psicológicas tais que tornem o sistema social inviável e incompatível com o bem humano” (...) Se as inclinações engendradas derem sustentação a arranjos justos, ou se puderem ser facilmente acomodadas por eles, a primeira parte da argumentação está confirmada (T), §80, p. 655).

Caso se verifique que a psicologia moral engendrada pelo princípio de justiça estimule tendências destrutivas entre os indivíduos, deve-se reavaliar a escolha da concepção de justiça, retornar à primeira etapa da argumentação e escolher outros princípios para aplicar na estrutura básica da sociedade.



#### 4. O PAPEL DA PSICOLOGIA MORAL

A psicologia moral tem um papel central na teoria da justiça como equidade, pois é elemento indispensável para concluir a argumentação em favor dos dois princípios feita a partir da posição original. Tal estudo cumpre a função de investigar se a motivação moral engendrada – nos cidadãos da SBO – pelos princípios de justiça escolhidos na primeira etapa da posição original é capaz de dar sustentação à concepção de justiça correspondente a estes princípios. É relevante para o argumento da estabilidade uma vez que aponta em que medida os princípios de justiça escolhidos na primeira etapa da posição original conseguem gerar em seus membros um desejo efetivo não só de agir a partir destes princípios como também de apoiar as instituições justas (conforme os critérios definidos pela concepção de justiça escolhida) e se mobilizar para reformar as instituições injustas.

Rawls defende que a filosofia moral é independente de outros ramos da filosofia, o que significa que ela se orienta por problemas próprios, cujo desenvolvimento não está subordinado a outras áreas da filosofia, mas sim à uma disciplina específica, a Teoria Moral. Entretanto, percebemos que em alguns momentos o deslinde da teoria da justiça parece depender mais de uma interação com campos mais empíricos, como a teoria social e, em especial, a psicologia moral do que apenas da teoria moral, a tal ponto que se pode ter a impressão de que a investigação pertence mais ao campo da psicologia moral. Existe uma tensão entre a filosofia moral e a psicologia moral que nos leva a questionar quão autônoma é a filosofia moral se ela está numa forte relação de dependência com a psicologia moral<sup>72</sup>. De fato, Rawls acaba desenvolvendo sua teoria moral com um forte vínculo com a psicologia moral, sobretudo quando investiga o vínculo da estrutura das concepções morais com a sensibilidade humana, a tal ponto que podemos ficar em dúvida se trata estamos diante de uma investigação filosófica ou psicológica. Nestes casos, a orientação do filósofo é que tenhamos em mente que o foco da investigação é filosófico (cf. IMT, p. 302), isto é, há uma

---

<sup>72</sup> Baldwin aponta para o paradoxo que há na abordagem rawlsiana da teoria moral. Por um lado, Rawls alega que a teoria de Hume é muito dependente da psicologia, ao mesmo tempo em que – contra os intuicionistas racionais – defende a importância a teoria moral kantiana (Cf. BALDWIN, 2008, p. 257).

preocupação moral (e política) com a justiça orientando a pesquisa. Quando se fala sobre sentimentos morais, é preciso introduzir “alguma perspectiva de justiça na explicação do desenvolvimento do sentimento correspondente” (TJ, §75, p. 606), por isso, não se trata de uma investigação puramente psicológica. Uma das razões para defender o estudo da filosofia moral e teoria da justiça atrelado à psicologia moral é que

“Mesmo que se faça tudo de acordo com procedimentos constitucionais, precisamos explicar porque esses procedimentos são aceitos. () uma teoria correta da política num regime constitucional justo pressupõe uma teoria da justiça que explique como sentimentos morais influem na condução de assuntos públicos (TJ, §75, p. 608).

Podemos afirmar que há uma relação de dependência mútua entre a filosofia moral e psicologia moral. Se, por um lado, a psicologia moral está no coração da teoria moral, por outro, a teoria do aprendizado moral é fundamentada “explicitamente em uma teoria ética específica” (TJ, § 75, p. 612). A efetividade da concepção moral depende de como seus princípios conseguem mobilizar as atitudes naturais e sentimentos morais das pessoas; e a psicologia moral, por sua vez, não pode ser pensada nem como desvinculada de princípios de justiça e nem de um contexto institucional que seja justo. Cada concepção de justiça resulta num tipo distinto de psicologia moral (Cf. TJ, §75, p. 606).<sup>73</sup>

O que é fundamental para a filosofia moral é a concepção de pessoa que está na base da teoria. O estudo da psicologia moral – que consiste em grande parte na explicação de como os indivíduos adquirem e desenvolvem o senso de justiça – nos auxilia a compreender melhor esta concepção, bem como a noção de relações sociais, da estrutura geral e fins da cooperação social expressas por cada estrutura ou doutrina moral.

Iniciaremos expondo algumas das razões pelas quais o princípio da utilidade não passa no teste da estabilidade inerente, pois supomos que assim conseguiremos ver compreender melhor

---

<sup>73</sup> Essa tensão entre a filosofia moral e a psicologia moral pode nos levar a questionar quão autônoma e independente é a filosofia moral já que ela está numa forte relação de dependência com a psicologia moral.

para qual a problema a teoria da justiça como equidade pretende oferecer uma solução e como a análise da psicologia moral contribui na tarefa da teoria moral de evidenciar as concepções de pessoa e relações sociais subjacentes à teoria.

A estabilidade inerente do utilitarismo clássico depende que o princípio da utilidade, ao ser aplicado institucionalmente na SBO gere e reforce nos cidadãos a capacidade de identificar-se empaticamente com seus concidadãos. A menos que compaixão (*sympathy*) e benevolência sejam intensamente cultivados, a sociedade não será estável. A capacidade para identificação empática (*sympathetic identification*) é que será responsável por operar o mecanismo psicológico nos indivíduos. O problema da compaixão (ou, identificação empática) é, que,

(...) via de regra, não é forte o suficiente nem está seguramente situada dentre o repertório motivacional humano como sendo uma fonte confiável de apoio para princípios e instituições utilitaristas (SCHEFFLER, 2003, p. 435).

A concepção de pessoa que o utilitarismo coloca na base de sua teoria seria a de “altruístas perfeitos” (TJ, §30, p. 232), isto é, pessoas com a forte disposição de se dedicarem aos outros, a praticar ações voluntárias que beneficiam outras pessoas. A estabilidade de uma SBO pelo princípio da utilidade depende de que todos os cidadãos tenham compaixão e identificação empática uns pelos outros. Os princípios dessa concepção parecem exigir, enquanto um dever de justiça, que pratiquemos (ou que estejamos dispostos a praticar) “atos que promovem um bem maior para outrem, seja qual for o custo para nós mesmos” (TJ, §19, p. 140). O problema deste tipo de exigência é que ela é excessiva para os cidadãos de uma SBO, isto é, para pessoas que se concebem reciprocamente como morais, racionais, livres e iguais.

A concepção de justiça utilitarista exige que o indivíduo ou um grupo de indivíduos tenham de estar dispostos abrir mão de bens primários para que um grupo maior desfrute de um maior bem-estar. Este tipo de exigência se torna cruel quando recai sobre as pessoas que estarão entre os menos favorecidos. É difícil imaginar como o princípio da utilidade conseguiria inspirar nos mais desfavorecidos sentimentos amistosos com relação aos mais afortunados que se beneficiam do saldo máximo da satisfação (cf. TJ,

§76, p. 616), pois os esforços exigidos dos menos favorecidos são muito grandes, estão para além do que nossa capacidade humana aceitaria. Somente um altruísta perfeito – ou então, um indivíduo que passou por doutrinação coercitiva – se sentiria motivado a se engajar na cooperação social nos termos do princípio da utilidade.

Fica evidente por que o utilitarista salienta a capacidade de compaixão (*sympathy*). Aqueles que não se beneficiam da situação melhor de outros devem identificar-se com a soma (ou média) mais elevada de satisfação, caso contrário não vão querer adotar o critério da utilidade (TJ, §76, p. 617).

Rawls não está defendendo que atos supererrogatórios<sup>74</sup> não tenham valor moral. Defende que “é bom praticá-los, entretanto não constituem obrigação, nem dever para ninguém” (TJ, §19, p. 140). Nada impede que um indivíduo tenha um projeto de vida orientado por ideais supererrogatórios, entretanto, não pode compor um dever de justiça ao qual todos estão submetidos a obrigação de promover o bem maior de outrem, mesmo que isso implique em auto sacrifício em termos de bens primários. Princípios de justiça que impliquem neste tipo de exigência não seriam capazes de gerar nos cidadãos o firme desejo de segui-los e de apoiar as instituições orientadas por eles. Em outras palavras, o princípio da utilidade não tem força suficiente para garantir, de modo inerente, a sua própria estabilidade.

Dada a natureza humana é muito improvável que as pessoas pudessem formar os extensos laços de simpatia e mesmo de benevolência imparcial que seriam necessários para inculcar em todos um sentimento regulador da justiça informado pelo princípio da utilidade (FREEMAN, 2007a, p. 99).

O senso de justiça formado pelos princípios da justiça como equidade, diferente do moldado pelo princípio da utilidade, procura manter relacionamentos entre as pessoas e apoio às instituições

---

<sup>74</sup> Ação supererrogatória ocorre quando a uma boa ação é realizada em benefício de outra pessoa, mas cuja realização implica em “considerável perda ou risco para o agente, estimando-se para isso os seus interesses interpretados de maneira mais restrita.” São exemplos de atos supererrogatórios os atos de bondade e misericórdia, de heroísmo e auto sacrifício. (Cf. TJ, §66, p. 542; §19, p. 140).

com base na ideia de reciprocidade, e não de auto sacrifício<sup>75</sup>. A suposição de Rawls é que os laços naturais de amor, amizade, companheirismo, confiança mútua e laços com instituições serão reforçados pela aplicação institucional da concepção da justiça como equidade, pois o senso de justiça orientado por estes princípios se desenvolve como continuidade das leis psicológicas, criando assim um vínculo entre as atitudes naturais e os sentimentos morais.

A psicologia moral esboçada na teoria da justiça como equidade toma como fato psicológico profundo, característico da nossa natureza humana – e a partir do qual serão esboçados os estágios do desenvolvimento moral – a reciprocidade, entendida como a tendência para retribuir na mesma medida.<sup>76</sup> É também a partir dessa disposição humana fundamental que Rawls formula as leis psicológicas que orientam o desenvolvimento moral dos membros da SBO. Veremos agora qual o papel da psicologia moral na Posição Original.

#### **4.1. O papel da psicologia moral na escolha dos princípios na Posição Original**

Rawls está elaborando uma teoria ideal, ou seja, não está desenvolvendo uma teoria para sociedades concretas. A concepção de sociedade que tem na base de sua teoria é um “sistema social teoreticamente definido” (IMT, p. 294), caracterizado como um sistema equitativo de cooperação. As leis da psicologia moral são elaboradas com referência a um contexto justo, portanto, a psicologia moral apresentada na TJ é a descrição resumida de como os seres humanos em uma sociedade bem-ordenada (que concretize os princípios da justiça como equidade) podem vir a adquirir um

---

<sup>75</sup> “Ao recorrer diretamente à capacidade de compaixão (*sympathy*) como fundamento para a conduta justa na ausência de reciprocidade, o princípio da utilidade não só requer mais do que a justiça como equidade, como também depende de inclinações mais fracas e menos comuns” (TJ, §76, p. 618).

<sup>76</sup> Na edição revista da TJ, a reciprocidade é assim caracterizada: “*The basic idea is one of reciprocity, a tendency to answer in kind*” (TJ, §75, p. 433). Na tradução para o português brasileiro que estamos usando como referência neste trabalho, foi suprimida a caracterização da reciprocidade como *uma tendência a responder na mesma medida* (Cf. TJ, §75, p. 610).

senso de justiça e os outros sentimentos morais (*moral sentiments*)” (TJ, §69, p. 563).

Retomemos aqui a ideia já presente na TJ, mas tornada explícita apenas no PL: existem três pontos de vista que um mesmo sujeito pode considerar a si mesmo na teoria da justiça da justiça: ponto de vista da das partes na posição original, ponto de vista dos cidadãos da sociedade bem-ordenada e ponto de vista nosso – meu e seu, leitor, nós que estamos examinando a formulação da teoria da justiça como equidade e pensando na sua viabilidade enquanto uma concepção política de justiça (Cf. PL, I, §4.6, p. 32). Entretanto, como é que se articulam estes pontos de vista e as motivações inerentes a cada um deles? Qual o papel da psicologia moral em cada um destes pontos de vista?

O ponto de vista na PO diz respeito às motivações das partes representativas, isto é, ‘criaturas artificiais’ que representam os interesses ‘abstratos’, ora dos cidadãos das sociedades bem-ordenadas, ora de pessoas que se autocompreendem como livres e iguais. A motivação das partes na PO é o autointeresse em garantir bens primários (introduzidos na primeira etapa da PO por meio da teoria restrita do bem como racionalidade).

As partes não escolhem os bens primários movidas por um senso de justiça, por benevolência ou por uma preocupação altruísta com o bem-estar umas das outras. Também não estão motivadas pela inveja, nem pelo ciúme, nem por qualquer outra tendência psicológica em especial. Em razão destas suposições motivacionais das partes na PO, Rawls foi criticado por, supostamente, ter na base de sua teoria uma concepção de pessoa que toma o indivíduo como isolado, deslocado de seu ambiente e desvinculado de relações sociais. A motivação das partes na PO não significa, no entanto, que essa seja a principal motivação dos cidadãos de uma SBO ou de pessoas reais.

Não há incoerência, então, em supor que, removido o véu da ignorância, as partes descobrem que têm laços de sentimentos e afeição, e que querem promover os interesses de outros e vê-los atingir seus objetivos. Porém, o postulado do desinteresse mútuo na posição original visa garantir que os princípios de justiça não dependem de suposições fortes (TJ, §22, p. 157).

(...) embora as partes presentes na situação original não tenham um interesse pelos interesses umas das outras, elas sabem que em uma sociedade precisam contar com a estima de seus concidadãos. Seu auto respeito e sua confiança no valor de seu próprio sistema de fins não tolera a indiferença, muito menos o desprezo de outrem” (TJ, §51, p. 421).

A escolha dos princípios com vistas aos bens primários, no entanto, constitui apenas um dos momentos do raciocínio a favor dos princípios de justiça na PO. Ainda que motivadas pelo autointeresse, as partes sabem que são representativas, isto é, que estão representando os interesses de pessoas morais racionais livres e iguais, dotadas da capacidade para senso de justiça e concepção de bem, situadas no contexto de uma sociedade bem-ordenada, e que são estas pessoas, nesta sociedade, as destinatárias finais dos princípios. Por esta razão, a concepção de justiça escolhida pelas partes na posição original deve ser tal que as pessoas cujas partes estão representando – os cidadãos de uma SBO – possam concordar e desejar a agir a partir deles. Os princípios não podem, por exemplo, gerar expectativas legítimas que depois serão frustradas, e nem fazer uma suposição motivacional forte tal como exigir do indivíduo – como dever de justiça – uma moralidade supererrogatória.

Nem sempre fica claro na teoria da justiça que as partes representativas devem encontrar um modo de incluir no interior do procedimento da PO considerações sobre o consequências e conteúdos da vontade dos cidadãos da SBO para que seja possível verificar – ainda do ponto de vista da PO – se os cidadãos representados poderiam desejar agir a partir dos princípios de justiça. Nesse caso, o leitor pode vir a questionar

(...) como alguém pode averiguar a motivação das pessoas em uma SBO? Resposta: a partir das leis da psicologia; assumindo que as pessoas crescem e vivem sob instituições justas, tal como definidas pelos princípios, alguém deve desenvolver (*to work out*) quais concepções de bem e quais interesses morais as pessoas adquiririam. Quando essas deliberações ocorrem do ponto de vista da posição original, elas são parte do raciocínio

que confirma ou desconfirma a escolha provisória da concepção de justiça (FG, p. 274).

A motivação das partes da SBO não é determinada diretamente pelos motivos das partes na PO. Do pressuposto de que a motivação das partes na PO não estão interessadas em promover os interesses dos demais, apenas o seu próprio (de garantir bens primários) não se segue que na SBO as pessoas irão adquirir e buscar realizar planos modos de vida individualistas. O argumento de Rawls é que tais motivações não irão prevalecer na SBO (Cf. FG, p. 274-5).

As partes utilizarão conhecimentos gerais sobre a psicologia moral (lembrando que estão no rol de informações não vedadas às partes pelo véu de ignorância), leis psicológicas e aprendizado moral, para analisar “se a concepção já adotada [escolhida provisoriamente na primeira etapa da PO] é viável e se não é tão instável que qualquer outra poderia ser melhor” (TJ, §76, p. 622).

#### **4.2. Senso de Justiça e Desenvolvimento Moral**

Na justiça como equidade, senso de justiça é utilizado em dois sentidos: (i) como “capacidade mental que envolve o exercício do raciocínio” (TJ, §9, p. 58), e (ii) como sentimento moral - engendrado por essa capacidade - que resulta de um certo desenvolvimento e do vínculo com atitudes naturais. A descrição do senso de justiça enquanto sentimento é caracterizada por uma “disposição consolidada de se adotar e querer agir com base no ponto de vista moral, pelo menos da forma como isso é definido pelos princípios de justiça” (TJ, §75, p. 606). Rawls supõem que “a capacidade para ter um senso de justiça e os sentimentos morais é uma adaptação da humanidade ao seu lugar na natureza” (TJ, §76, p. 620), e que, portanto, é comum a todos seres humanos, ainda que algumas pessoas possam ter um senso de justiça mais apurado do que outras. Para compartilhar da posição de igual cidadania - conferida pelo senso de justiça - é suficiente que se tenha uma capacidade mínima para ter sentimentos morais (cf. TJ, §77, p. 628)

Rawls privilegia uma dimensão cognitiva dos sentimentos morais, rejeita que sejam identificados com “sensações e manifestações comportamentais características” (TJ, §73, p. 599). A principal característica de um sentimento moral (como senso de justiça, culpa e indignação), e que o distingue de outras emoções

(como a raiva e medo) e das atitudes naturais (como amor, confiança mútua e amizade) é que a explicação do sentimento é feita com relação a um conceito moral e seus princípios associados, e não com referência a uma reação corporal. “A explicação que a pessoa dá de seus sentimentos [morais] faz menção a algo que foi reconhecido como certo ou errado” (TJ, §73, p. 594), e essas explicações diferem de um sentimento para outro.

Uma mesma ação, por exemplo, pode despertar no indivíduo vários sentimentos morais, mas o que irá distinguir uma da outra é o tipo de explicação oferecido. Vejamos o exemplo da culpa e da vergonha. Sentimentos de culpa e de vergonha evocam “princípios que pertencem a partes distintas da moralidade, ou os invocam de pontos de vista distintos” (TJ, §74, p. 599). A explicação do sentimento de culpa, indignação e arrependimento está relacionada com a transgressão do conceito de correto, enquanto que vergonha, escárnio e desdém fazem referência a uma ofensa ao conceito de bem.

Quando uma pessoa moral trapaceia para atingir seus objetivos, ela passa a sentir culpa e vergonha, além de despertar ressentimento e depreciação de sua pessoa quando tal ato se torna público. Culpa por agir contra o seu próprio senso de justiça e violar das reivindicações justas de outros e causar-lhes um mal; ressentimento e indignação são causados nas pessoas que tiveram seus direitos violados pelo trapaceiro (Cf. TJ, §67, p. 551). A vergonha moral é explicada como emoção provocada pela perda de um tipo especial de bem, a autoestima e auto respeito, tem origem num sentimento de depreciação do eu. Ao trapacear, fica evidente a incapacidade de realizar nossos objetivos por meio de nossas excelências e virtudes.

Costumamos nos envergonhar daqueles defeitos de nossa pessoa e das falhas em nossos atos que indiquem perda ou ausência das excelências essenciais para realizar nossos mais importantes objetivos associativos (TJ, §67p. 549).

As pessoas de quem dependemos para confirmar o senso de nosso próprio valor verão em nossa trapaça a incapacidade de realizar os nossos objetivos por meio das nossas excelências, e terão para conosco sentimento de desdém ou escárnio. A ausência das excelências tende “a solapar tanto a nossa autoestima, quanto a

estima que nossos associados sentem por nós” (TJ, §67, p. 550). A culpa e a vergonha são sentimentos morais que expressam “uma preocupação com terceiros e com a própria pessoa que deve estar presente em toda conduta moral” (TJ, p. §73, p. 598). Estes sentimentos, assim como o senso de justiça, são adquiridos, são resultado de um certo desenvolvimento moral.

Durante o processo de aprendizagem moral, as pessoas adquirem e desenvolvem, além do senso de justiça, outros sentimentos morais e atitudes naturais a ele vinculados (como culpa, ressentimento, indignação, amor, amizade, companheirismo, etc.). O desenvolvimento moral é orientado pela reciprocidade, que ora é concebida como uma tendência humana natural, ora como um princípio psicológico, que indica uma disposição para retribuição, para responder na mesma medida. A reciprocidade é fato psicológico profundo que está na base das três leis que orientam os estágios do aprendizado moral:

Primeira lei: [dado que as instituições familiares são justas e que]<sup>77</sup> os pais amam e expressam manifestamente esse amor preocupando-se com o bem da criança, esta, por sua vez, reconhecendo o amor patente que têm por ela, vem a amá-los.

Segunda lei: dado que a capacidade de solidariedade da pessoa se constituiu por meio de vínculos adquiridos de acordo com a primeira lei, e dado um arranjo social justo e publicamente conhecido por todos como justo, então essa pessoa cria laços amistosos e de confiança com outros membros da associação quando estes, com intenção evidente, cumprem com seus deveres e obrigações, e vivem segundo os ideais de sua posição.

Terceira lei: dado que a capacidade de solidariedade da pessoa foi constituída por meio da criação de laços em conformidade com as duas primeiras leis, e já que as instituições da sociedade são justas e publicamente

---

<sup>77</sup> De acordo com Samuel Freeman, por conta de um erro de impressão na edição revista da *Theory*, publicada em 1999, esta primeira oração da primeira lei psicológica, que ora colocamos em colchetes, acabou sendo suprimida. (Cf. FREEMAN, 2007, p. 502, n.12)

conhecidas por todos como justas, então essa pessoa adquire o senso de justiça correspondente ao reconhecer que ela e aquelas com quem se preocupa são beneficiárias desse arranjo (TJ, §75, p. 605).

O aprendizado moral ocorre em três estágios: moralidade da autoridade, moralidade da associação e moralidade de princípios. O desenvolvimento é explicado por meio da relação entre o sentimento de culpa e o princípio psicológico da reciprocidade. No estágio da moralidade da autoridade (TJ, §70), o sentimento de culpa surge quando a criança – que ama e é amada por seus pais – transgredir a ordem. Ao violar o preceito paterno, a criança sente culpa e remorso por quebrar o laço de amor e confiança com seus pais, por não retribuir a eles na mesma medida. Na moralidade de associação (§71), a explicação do sentimento de culpa é feita com relação à amizade e companheirismo (*fellow-feeling*). Quando a pessoa pratica algum ato que fere a moralidade do grupo ao qual pertence (associação, time de futebol, partido político, etc), ela sente culpa por quebrar a confiança que seus amigos e companheiros depositavam nela, se ressentido de não poder mais contar com a confiança deles, e ainda pode despertar nos demais membros da associação indignação pela transgressão da norma do grupo. Esses laços de afeto que desenvolvemos com determinadas pessoas possuem um papel essencial na aquisição da moralidade. Na moralidade de princípios, “quando vamos de encontro a nosso senso de justiça, explicamos nossos sentimentos de culpa recorrendo aos princípios de justiça” (TJ, §72, p. 585). Ao violarmos os princípios da justiça, por exemplo, sonegando tributos, podemos estar prejudicando o fornecimento e acesso a um determinado bem público a pessoas com quem nos preocupamos e que são beneficiadas pelo arranjo justo.

Aos que duvidam da existência da capacidade para um senso de justiça, Rawls sugere o exercício de imaginar não ter um senso de justiça. E responde: “seria como não ter parte da nossa humanidade” (TJ, §75, p. 604). Como os sentimentos morais são considerados uma parte normal da vida humana, se os eliminarmos, também daremos fim a atitudes e vínculos naturais, e que, portanto, constituem a nossa humanidade e nossos laços de sociabilidade.

Quando reconhecemos estes vínculos, temos outros motivos para agir com equidade (para além do desejo de justiça). Neste

processo de desenvolvimento, os cidadãos da SBO irão adquirir quatro desejos, ou motivos, morais, descritos do seguinte modo: (i) desejo de evitar os custos psicológicos da hipocrisia e decepção, (ii) desejo de manter laços afetivos construídos nos três estágios da moralidade, dentre os quais amor, confiança mútua, amizade, companheirismo (*fellow feeling*), (iii) desejo de participar em formas de vida social que evocam o seu talento e o dos outros indivíduos (Princípio Aristotélico), e (iv) desejo de expressar sua natureza de seres naturalmente livres, iguais e racionais (interpretação kantiana da pessoa). Todos estes desejos estão ligados à teoria completa do bem como racionalidade. Rawls sustenta que desejo de agir a partir de princípios de justiça (senso de justiça) não pode ser um desejo dentre outros, mas sim, deve ser o desejo regulador dos demais desejos e motivos morais (vinculados à noção de bem) (TJ, §86, p. 708-709). No entanto, para que a concepção de justiça seja estável (isto é, para que haja um equilíbrio de motivos) não basta que os indivíduos tenham desenvolvido um desejo efetivo de agir com justiça, é preciso que tal desejo seja congruente, ou, pelo menos, não incompatível com o bem. A segunda parte do argumento da estabilidade consiste, então, em tentar demonstrar que manter o senso de justiça constitui um bem para a pessoa, tanto do ponto de vista individual, quanto coletivo.

## CONSIDERAÇÕES FINAIS

Conforme indicamos na introdução deste trabalho, o argumento da estabilidade desenvolvido em *Uma Teoria da Justiça* recebeu pouca atenção de comentadores durante os 30 anos que se seguiram à publicação do livro. Sugerimos que esse lapso ocorreu, em primeiro lugar, pela grande repercussão que a defesa da justiça social em referido livro teve. Ademais, nos anos seguintes, John Rawls se dedicou a responder críticas e interpretar ele próprio sua obra, o que forneceu ainda mais material para discussões. Se colocarmos em perspectiva histórica, perceberemos que 30 anos é um período bem exíguo, portanto, demandar que uma obra filosófica tão extensa, tão densa e inovadora quanto TJ tenha esgotado as discussões em torno de si é uma exigência demasiada. Apontamos que a própria estrutura do livro e a maneira esparsa – e labiríntica – em que foi apresentado o argumento da estabilidade tiveram sua parcela de contribuição para a pouca atenção dispensada à parte final do livro.

Uma das primeiras dificuldades ao estudar o argumento da estabilidade em TJ é compreender em qual parte da estrutura moral e das duas estratégias de justificação apresentadas no livro, o argumento pode ser localizado. Enfrentamos esse impasse investigando a estrutura moral da teoria da justiça como equidade, donde foi possível evidenciar dois âmbitos distintos entre si, o âmbito da determinação (ou, âmbito normativo) e o âmbito da motivação (descritivo). Constatamos que essa estrutura bipartida da teoria moral tem implicações nas estratégias de justificação da justiça como equidade: tanto a posição original, quanto equilíbrio reflexivo, são estruturados de modo que a comportar esses dois âmbitos. Esse resultado foi especialmente importante no que diz respeito à ideia de Posição Original, pois pudemos tornar clara uma característica, até então, raramente notada: possui duas etapas argumentativas, e o argumento da estabilidade é analisado na segunda etapa. A motivação moral, portanto, é um estágio necessário da justificação da justiça social, e não um momento à parte. Com isso, fica evidente a relevância que a psicologia moral passa a ter para a filosofia moral e política.

A despeito das contribuições que, eventualmente possamos ter feito com o resultado de nossa investigação, há ainda ampla

gama de questões intimamente relacionadas com a questão da estabilidade que não pudemos abordar nesta pesquisa, mas que merecem ser aprofundadas, como a relação entre estabilidade e autoestima, o papel da condição de publicidade e finalidade para a estabilidade, os quatro argumentos da congruência (argumento Kantiano, argumento do princípio Aristotélico, argumento da união social de uniões sociais e argumento da analogia dos riscos do amor e da justiça), bem como investigar mais a fundo o papel dos sentimentos morais na teoria da justiça.

Por fim, esperamos que nosso empenho para tornar mais clara a estrutura da teoria moral e da posição original possa, de algum modo, contribuir para uma melhor compreensão de *Uma Teoria da Justiça*.

## REFERÊNCIAS BIBLIOGRÁFICAS

### a) Obras de John Rawls

RAWLS, John. A Kantian Conception of Equality (1975). In: *Collected Papers*. Samuel Freeman (Org.). Cambridge: Harvard University Press, 1999. pp.254-266.

\_\_\_\_\_. *A Theory of Justice*. Revised Edition. Cambridge: Harvard University Press, 1999.

\_\_\_\_\_. Fairness to Goodness (1975). In: *Collected Papers*. Samuel Freeman (Org.). Cambridge: Harvard University Press, 1999. pp.267-285.

\_\_\_\_\_. *Justice as Fairness: a Restatement*. Cambridge: Harvard University Press, 2001.

\_\_\_\_\_. John Rawls: For the Record. *The Harvard Review of Philosophy* (Spring 1991), p. 38-47.

\_\_\_\_\_. Kantian Constructivism in Moral Theory (1980). In: *Collected Papers*. Samuel Freeman (Org.). Cambridge: Harvard University Press, 1999. pp.303-258.

\_\_\_\_\_. *Political Liberalism*. Expanded Edition. New York: Columbia University Press, 1996.

\_\_\_\_\_. Reply to Alexander and Musgrave (1974). In: *Collected Papers*. Samuel Freeman (Org.). Cambridge: Harvard University Press, 1999. pp.232-253.

\_\_\_\_\_. The Independence of Moral Theory (1975). In: *Collected Papers*. Samuel Freeman (Org.). Cambridge: Harvard University Press, 1999. pp.286-302.

\_\_\_\_\_. The Sense of Justice (1963). In: *Collected Papers*. Samuel Freeman (Org.). Cambridge: Harvard University Press, 1999. pp.96-116.

## **b) Traduções utilizadas da obra de John Rawls**

RAWLS, John. *Conferências sobre a história da filosofia política*. Tradução de Fabio M. Said. São Paulo, WMF Martins Fontes, 2012.

\_\_\_\_\_. *Direito dos povos*. Tradução de Luís C. Borges. São Paulo: Martins Fontes, 2004.

\_\_\_\_\_. *História da Filosofia Moral*. Barbara Herman (Org.). Tradução de Ana Aguiar Cotrin. São Paulo, WMF Martins Fontes, 2005.

\_\_\_\_\_. *Justiça como equidade: uma concepção política, não metafísica*. Tradução de Álvaro de Vita. *Lua Nova*. São Paulo. n. 25, 1992.

\_\_\_\_\_. *Justiça como equidade: uma reformulação*. Tradução de Claudia Berliner. Revisão técnica e da tradução Álvaro de Vita. São Paulo: Martins Fontes, 2003.

\_\_\_\_\_. *Prefácio à edição francesa dos textos intermediários (1992)*. In: *Justiça e Democracia*. Catherine Audard (Org.). Tradução de Irene A. Paternot. Seleção, apresentação e glossário Catherine Audard. São Paulo: Martins Fontes, 2000.

\_\_\_\_\_. *Liberalismo Político*. Tradução de Álvaro de Vita. 2 ed. São Paulo: Martins Fontes, 2011.

\_\_\_\_\_. *Uma Teoria da Justiça*. Tradução de Jussara Simões. Revisão técnica e da tradução Álvaro de Vita. 3 ed. São Paulo: Martins Fontes, 2008.

## **c) Comentários e outras referências**

BALDWIN, Thomas. Rawls and Moral Psychology. In: SHAFER-LANDAU, Russ (ed.). *Oxford Studies in Metaethics*. (vol. III), Oxford: Oxford University Press, 2008. pp. 247-270

BARRY, Brian. John Rawls and the Search for Stability. In: *Ethics*, Vol. 105, No. 4 (Jul., 1995), pp. 874-915.

BOBBIO, MATTEUCCI, PAQUINO. *Dicionário de Política*. Vol. I. 16 ed. Brasília: UnB, 1998.

BRAGA, Antonio Frederico Saturnino. *Kant, Rawls e o utilitarismo: justiça e bem na filosofia política contemporânea*. Rio de Janeiro: Contraponto, 2011.

CHAPMAN, John W. Rawls's Theory of Justice. *American Political Science Review* 69, 2 (1975): 588-94.

DE LA MORA, Gonzalo Fernandez. *Egalitarian Envy: The Political Foundations of Social Justice*. New York: Paragon House, 1987.

ESTEVES, Julio. As críticas ao utilitarismo por John Rawls. *Ethic@. Revista Internacional de Filosofia Moral*. Florianópolis v.1, n.1, Jun 2002, p.81-96. Disponível em: <<http://www.cfh.ufsc.br/ethic@/ETHIC1~6.PRN.pdf>>

FRAZER, Michael. John Rawls: Between Two Enlightenments. *Political Theory* 35, 2007, n. 6: 756-780.

FREEMAN, Samuel. Introduction: John Rawls – An Overview. In: FREEMAN, Samuel (Org). *The Cambridge Companion to Rawls*. Cambridge: Cambridge University Press, 2003, pp. 1-61.

\_\_\_\_\_. *Justice and the Social Contract: Essays on Rawlsian political Philosophy*. Oxford: Oxford University Press, 2007. (FREEMAN, 2007a)

\_\_\_\_\_. *Rawls*. New York: Routledge, 2007. (FREEMAN, 2007b)

GARGARELLA, Roberto. *As teorias da justiça depois de Rawls*. São Paulo: Martins Fontes, 2008.

GAUS, Gerald. The Turn to a Political Liberalism. In: MANDLE, Jon; REIDY, David. (Org) *A Companion to Rawls*. Oxford: Wiley Blackwell, 2014, pp. 235-250.

HAYEK, Friedrich August von. *Lei, legislação e Liberdade: uma nova formulação dos princípios liberais de justiça e economia política*. Volume II: A miragem da justiça social. Tradução de Maria Luiza X. de A. Borges. São Paulo: Visão, 1985.

HILL Jr, Thomas. Stability, a Sense of Justice, and Self-Respect. In: MANDLE, Jon ; REIDY, David. (Org) *A Companion to Rawls*. Oxford: Wiley Blackwell, 2014, pp. 200-215.

KLOSKO, George. Rawls's Argument from Political Stability. In: *Columbia Law Review*, Vol. 94, No. 6 (Oct., 1994), pp. 1882-1897.

KRAUSE, Sharon. *Civil Passions: Moral sentiment and Democratic Deliberation*. Princeton: Princeton University Press, 2008.

LADEN, Anthony Simon. Constructivism as Rhetoric. In: MANDLE, Jon ; REIDY, David. (Org) *A Companion to Rawls*. Oxford: Wiley Blackwell, 2014, pp. 59-72.

\_\_\_\_\_. The house that Jack built: Thirty Years of Reading Rawls. *Ethics*, Vol. 113, No. 2 (January 2003), pp. 367-390.

LARMORE, Charles. *The Morals of Modernity*. Cambridge: Cambridge University Press, 1996.

LUCAS, J.R., Against Equality. *Philosophy*, Volume 40, Issue 154. October 1965, pp. 297-307.

MAFFETTONE, Sebastiano. Political liberalism: Reasonableness and democratic practice. In: PHILOSOPHY & SOCIAL CRITICISM, vol 30 n. 5-6, 2004, pp. 541-577.

\_\_\_\_\_. *Rawls: an introduction*. Cambridge: Polity Press, 2010.

MANDLE, Jon. *Rawls's A Theory of Justice: an introduction*. Cambridge: Cambridge University Press, 2009.

McCLENNEN, Edward F. Justice and The Problem of Stability. In: *Philosophy & Public Affairs*, Vol. 18, No. 1 (Winter, 1989), pp. 3-30.

MIKHAIL, John. Rawls' Concept of Reflective Equilibrium and Its Original Function in A Theory of Justice. *In: Washington University Jurisprudence Review*. Volume 3, Issue 1, 2011, pp. 1-30.

NUSSBAUM, Martha Craven. *Political Emotions: Why love matters for Justice?* London: Belknap Press of Harvard University Press, 2013.

NUSSBAUM, Martha C; BROOKS, Thom, *Rawls's Political Liberalism*. New York: Columbia University Press, 2015.

RODILLA, Miguel Ángel. Epílogo : De A Theory of Justice a Political Liberalism - Otra vuelta de tuerca. *In: RAWLS, John. Justicia como equidad : Materiales para Una Teoría de la Justicia*. Madrid : Tecnos, 2013.

ROUSSEAU, Jean-Jacques. *Do Contrato Social*. Coleção Pensadores. 3 ed. São Paulo : Abril Cultural, 1983.

O'NEILL, Onora. Constructivism in Rawls and Kant. *In: FREEMAN, Samuel. (Org) The Cambridge Companion to Rawls*. Cambridge: Cambridge University Press, 2003. pp. 347-367.

\_\_\_\_\_. The Method of A Theory of Justice. *In: HÖFFE, Otfried (Ed.). John Rawls: Eine Theorie der Gerechtigkeit*. Berlin: Akademik Verlag, 1998, pp. 27-45.

SCHEFFLER, Samuel. Rawls and Utilitarianism. *In: FREEMAN, Samuel. (Org) The Cambridge Companion to Rawls*. Cambridge: Cambridge University Press, 2003, pp. 426-459.

SCHOECK, Helmut. *Envy : A theory of social behaviour*. Indianapolis: Liberty Fund, 1969.

STEMPLOWSKA, Zofia; SWIFT, Adam. Rawls on Ideal and Nonideal Theory. *In: MANDLE, Jon ; REIDY, David. (Org) A Companion to Rawls*. Oxford: Wiley Blackwell, 2014, pp. 112-127.

WEITHMAN, Paul. *Why Political Liberalism? On John Rawls's Political Turn*. Oxford: Oxford University Press, 2010.

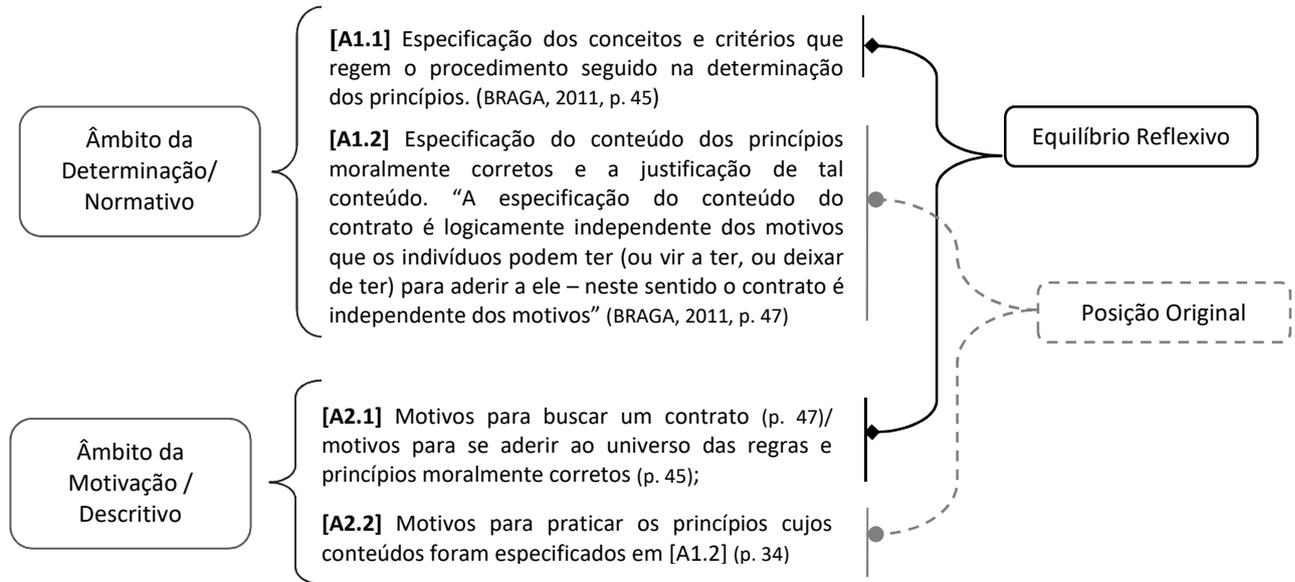
WILLIAMS, Bernard. *Moral: uma introdução à ética*. Tradução de Remo Mannarino Filho. Revisão da Tradução Marcelo Brandão Cipolla. São Paulo: Martins Fontes, 2005.

**ANEXO**  
**DIAGRAMAS SOBRE A ESTRUTURA DA TEORIA**

Os diagramas a seguir são uma tentativa de representar de modo esquemático a estrutura da justiça como equidade e das duas estratégias de justificação apresentadas em *Uma Teoria da Justiça*.

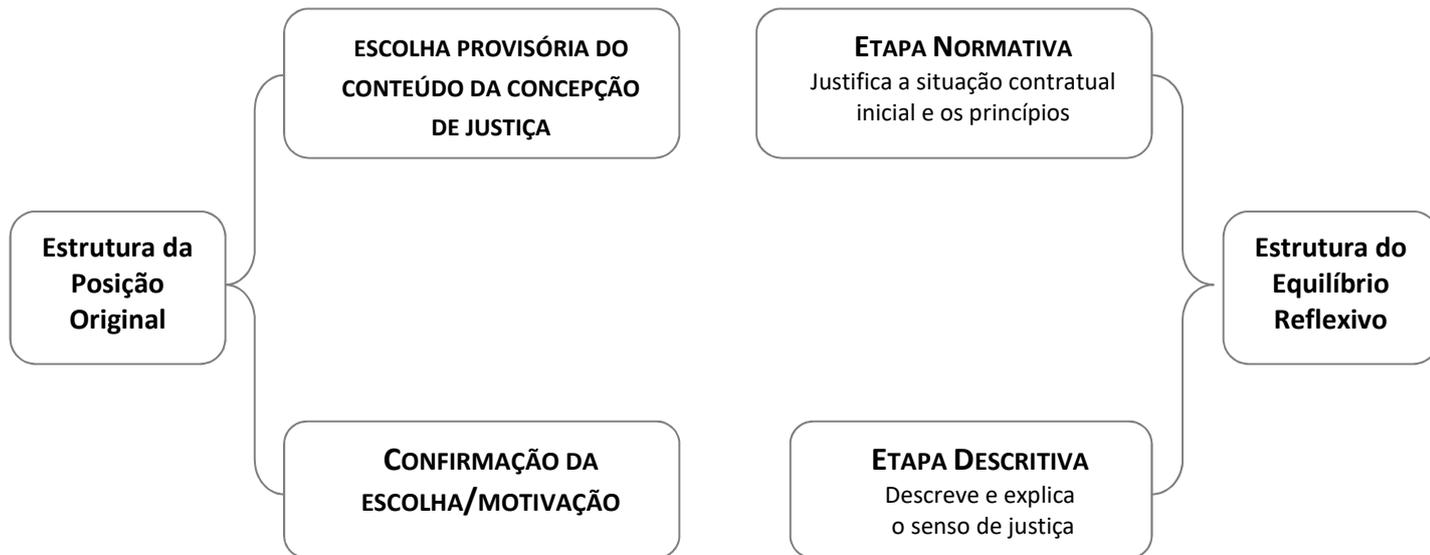


## Estrutura da teoria moral da justiça como equidade (deontologismo não-rigorista)



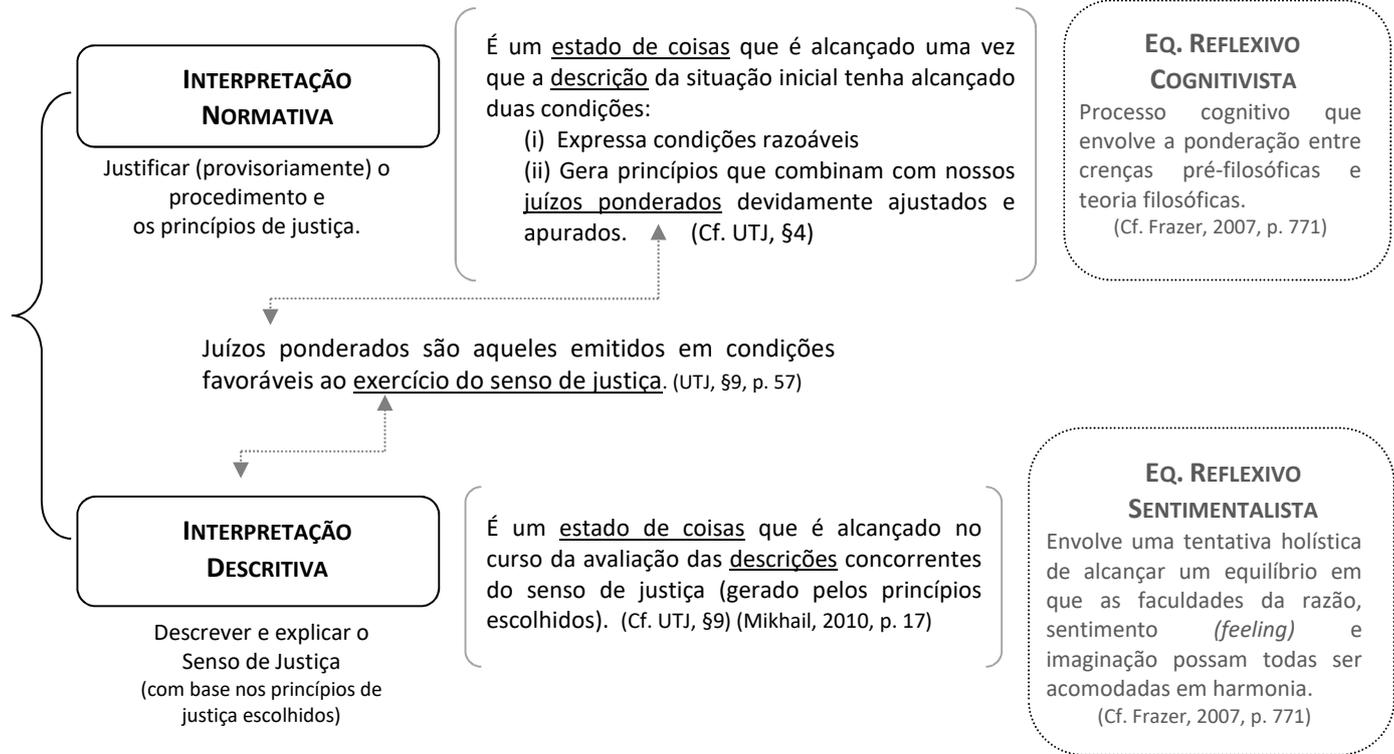


## Posição Original e Equilíbrio Reflexivo estruturados em duas etapas





## Duas explicações do Equilíbrio Reflexivo em *Uma Teoria da Justiça*





## Estrutura argumentativa da PO dividida *duas etapas*

