

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE DIREITO
DEPARTAMENTO DE DIREITO PÚBLICO E FILOSOFIA DO DIREITO

Jerônimo Erig Weiller

**UMA CRÍTICA À TEORIA DE JUSTIÇA DE JOHN RAWLS SOB A PERSPECTIVA
DA TEORIA DA ESCOLHA RACIONAL**

Porto Alegre

2021

JERÔNIMO ERIG WEILLER

**UMA CRÍTICA À TEORIA DE JUSTIÇA DE JOHN RAWLS SOB A PERSPECTIVA
DA TEORIA DA ESCOLHA RACIONAL**

Trabalho de Conclusão de Curso apresentado ao Departamento de Direito Público e Filosofia do Direito da Faculdade de Direito da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do grau de Bacharel em Ciências Jurídicas e Sociais.

Orientador: Prof. Dr. Paulo Baptista Caruso MacDonald

Porto Alegre

2021

JERÔNIMO ERIG WEILLER

**UMA CRÍTICA À TEORIA DE JUSTIÇA DE JOHN RAWLS SOB A PERSPECTIVA
DA TEORIA DA ESCOLHA RACIONAL**

Trabalho de Conclusão de Curso apresentado ao Departamento de Direito Público e Filosofia do Direito da Faculdade de Direito da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do grau de Bacharel em Ciências Jurídicas e Sociais.

Orientador: Prof. Dr. Paulo Baptista Caruso MacDonald

Aprovado em 19/05/2021

BANCA EXAMINADORA:

Professor Doutor Paulo Baptista Caruso MacDonald (Orientador)

Professor Doutor Wladimir Barreto Lisboa

Doutorando Edison Dri Consiglio Filho

Porto Alegre

2021

Dedico este trabalho à minha avó Anelise e aos meus pais, que tanto estimularam a minha curiosidade e o meu gosto pela leitura.

AGRADECIMENTOS

Agradeço primeiramente à minha família. Sem o apoio dela, esta graduação nunca teria acontecido.

Agradeço também a todos os meus amigos por terem feito com que estes anos fossem prazerosos. Obrigado ao pessoal do Americano que está comigo desde antes do fundamental, aos grupinhos do Dohms e do Leonardo que a gente conheceu no ensino médio, a todos os amigos que fiz na faculdade, ao pessoal do SAJU e do NEDEP, ao time de basquete e ao time de League of Legends da Associação Atlética Advogados do Diabo e a todos os demais que conheci em cursos, festas, na praia, pelos *videogames* ou por amigos em comum.

Dentre os meus amigos, gostaria de agradecer em especial a todos aqueles que tiveram saco para conversar comigo sobre filosofia e me ajudaram a desenvolver meus pensamentos: um obrigado especial para o Eduardo, o Doncatto, o Nikolas, o Renato, o Gabri, a Júlia, o Zanuzzi, o Mariani, o Salton, o Estéfano, o Raphael, o Matheus, o Bohm, o Santiago, o Nunes, o Gui, a Mari e a Nia. Dentre estes, dois se destacam por serem os com quem as conversas foram mais frequentes e mais profundas: um muito obrigado para o Becker e para o Thales.

Por fim, gostaria de agradecer a todos os professores de filosofia que tive do fundamental até hoje, por terem despertado o meu interesse nesta área e terem me ensinado muito (mesmo aqueles de quem discordei bastante). Dentre estes, meu agradecimento especial vai para o meu orientador, Paulo Baptista Caruso MacDonald. Muito obrigado por ter sido este orientador incrível, que tanto trabalhou para me ajudar a fazer uma ótima monografia, que sempre tinha sugestões e indicações de leitura relevantes e com quem tive tantos debates frutíferos.

“There is only one thing a philosopher can be relied upon to do, and that is to contradict other philosophers.”

- William James

RESUMO

O presente trabalho busca desenvolver uma crítica à teoria de justiça de John Rawls, na forma em que esta foi caracterizada em sua obra “Uma Teoria da Justiça”. Inicialmente, no primeiro capítulo, faz-se uma análise da chamada “teoria da escolha racional”, que é um modelo para a compreensão de comportamentos econômicos e sociais. Conceitos trabalhados neste capítulo serão relevantes para as críticas posteriores à teoria de Rawls. Em seguida, no segundo capítulo, é realizada uma breve reconstrução do raciocínio de Rawls para chegar aos seus princípios de justiça: ele imagina uma situação hipotética, chamada de “posição original”, em que várias pessoas idealmente racionais que não conhecem diversas características sobre si mesmas deliberam sobre os princípios de justiça que organizarão sua sociedade. A isonomia garantida por este processo deliberativo justificaria os princípios de justiça que Rawls afirma que elas escolheriam. Por fim, no terceiro capítulo, serão elencadas algumas críticas ao pensamento de Rawls, tanto de autoria do economista John Harsanyi quanto de autoria própria. Conclui-se que a teoria de justiça de Rawls, embora muito relevante e com diversos méritos, é passível de muitas críticas, o que faz com que tanto ela quanto o utilitarismo, considerado por Rawls seu principal adversário, sejam defensáveis.

Palavras-chave: Ética. Filosofia política. Filosofia do direito. Teoria da escolha racional.

ABSTRACT

The present work aims to develop a criticism of John Rawls' theory of justice, as it was developed in his "A Theory of Justice". Initially, in the first chapter, an analysis of the so called "rational choice theory", which is a model for the comprehension of economical and social behavior, will be made. Concepts elaborated in this chapter will be relevant for the following criticisms to Rawls' theory. Subsequently, in the second chapter, a brief reconstruction of Rawls' reasoning to reach his principles of justice will be done: he imagines a hypothetical situation, known as the "original position", in which many ideally rational people that don't know various characteristics about themselves deliberate over the principles of justice that will organize their society. The isonomy guaranteed by this deliberative process would justify the principles of justice that Rawls says they would choose. Lastly, in the third chapter, a few criticisms of Rawls' theory, both by John Harsanyi and myself, will be listed. It is concluded that Rawls' theory of justice, although very relevant and with many merits, is liable to many criticisms, which makes it so that both it and utilitarianism, which Rawls considered its main adversary, are defensible.

Keywords: Ethics. Political philosophy. Philosophy of law. Rational choice theory.

LISTA DE ILUSTRAÇÕES

Figura 1 – Gráfico da utilidade que as partes na posição original obteriam por bens primários.....	77
Figura 2 – Estado de coisas que deveria ser preferido segundo o princípio da diferença.....	77
Figura 3 – Estado de coisas que deveria ser preterido segundo o princípio da diferença.....	78

LISTA DE TABELAS

Tabela 1 – Opções de loterias do experimento de pensamento de Allais.....	26
Tabela 2 – Interpretações do segundo princípio de justiça de Rawls.....	42
Tabela 3 – Resultados das escolhas no experimento de pensamento de Harsanyi..	65

SUMÁRIO

INTRODUÇÃO.....	11
1. A TEORIA DA ESCOLHA RACIONAL.....	15
2. A TEORIA DE JUSTIÇA DE RAWLS.....	35
3. CRÍTICAS À TEORIA DE JUSTIÇA DE RAWLS.....	63
3.1. CRÍTICAS DE HARSANYI.....	63
3.2. CRÍTICAS AUTORAIS.....	74
CONCLUSÃO.....	90
REFERÊNCIAS.....	92

INTRODUÇÃO

Em 1971, foi publicada uma das mais influentes obras de filosofia política da tradição ocidental: “Uma Teoria da Justiça”, de John Rawls. Este livro tornou-se um dos tratados políticos mais comentados da história e seu autor passou a ser considerado por muitos o maior filósofo político do século XX. A obra exerceu grande influência em diversos autores posteriores, mantendo-se relevante até os dias de hoje.

Nela, John Rawls tinha um objetivo específico: desenvolver uma teoria ética que fundamentasse as instituições das sociedades democráticas ocidentais distinta do utilitarismo e do que ele chamou de “intuicionismo”.

Isso é deixado bastante explícito pelas palavras utilizadas por ele em seu prefácio:

“Talvez eu possa explicar meu objetivo neste livro da seguinte maneira. Durante muito da filosofia moral moderna a teoria sistemática predominante foi alguma forma de utilitarismo. (...) Aqueles que os criticaram [os utilitaristas] com frequência o fizeram de uma maneira muito mais limitada. Eles apontaram as obscuridades do princípio da utilidade e notaram as aparentes incongruências entre muitas de suas implicações e nossos sentimentos morais. Mas eles falharam, creio eu, em construir uma concepção moral viável e sistemática para se opor a ele. O resultado é que com frequência parece que nós somos forçados a escolher entre o utilitarismo e o intuicionismo”¹ (tradução livre).²

Rawls não considerava o utilitarismo uma teoria ética satisfatória para fundamentar as instituições das sociedades democráticas, pois, como a teoria tem como objetivo único a maximização da felicidade, ela não levaria os direitos a sério. Vide o seguinte trecho: “Particularmente, eu não acredito que o utilitarismo possa fornecer uma avaliação satisfatória dos direitos e liberdades básicos dos cidadãos

¹ RAWLS, John Bordley. **A Theory of Justice**. Edição revisada. Cambridge, Massachusetts: Harvard University Press, 1999, p. xvii.

² Texto original: “Perhaps I can best explain my aim in this book as follows. During much of modern moral philosophy the predominant systematic theory has been some form of utilitarianism. (...) Those who criticized them often did so on a much narrower front. They pointed out the obscurities of the principle of utility and noted the apparent incongruities between many of its implications and our moral sentiments. But they failed, I believe, to construct a workable and systematic moral conception to oppose it. The outcome is that we often seem forced to choose between utilitarianism and intuitionism”.

como pessoas livres e iguais, um requisito de absoluta primeira importância para uma avaliação de instituições democráticas”³ (tradução livre).⁴

Aquilo que ele chamou de intuicionismo, por sua vez, consiste na negação do utilitarismo apenas quando ele levar a consequências que forem consideradas antiéticas por aquele que as estiver examinando. Rawls também não considerava isso uma base segura para as instituições, pois carecia de qualquer sistematicidade e dependia demasiadamente da subjetividade daquele que julgava as diferentes situações.

Rawls então desenvolveu uma teoria própria, que culminou em sua obra “Uma Teoria de Justiça”. Após diversas críticas, ele reformulou sua teoria em obras posteriores, passando esta a possuir ao fim de sua vida um aspecto bastante diferente daquele que possuía originalmente.

O presente trabalho de conclusão de curso busca desenvolver uma crítica à teoria rawlsiana de justiça; contudo, o enfoque será apenas na teoria de Rawls como desenvolvida em “Uma Teoria de Justiça”. Rawls sofreu várias críticas pelo modo como se utilizou nesta obra da teoria da escolha racional, um modelo para a compreensão de comportamentos econômicos e sociais. Isso fez com que em uma obra posterior, “Liberalismo Político”, ele se utilizasse de argumentos diferentes para defender sua teoria de justiça^{5 6}, tendo se afastado um pouco da ideia de apreendê-la apenas da teoria da escolha racional.^{7 8} Estes outros argumentos, contudo, não serão abordados devido ao escopo limitado deste trabalho e ao enfoque dado nele à teoria da escolha racional.

O argumento central de Rawls em “Uma Teoria de Justiça” é o de que, para saber que princípios de justiça devem organizar uma sociedade, devemos imaginar

³ RAWLS, John Bordley. **A Theory of Justice**. Edição revisada. Cambridge, Massachusetts: Harvard University Press, 1999, p. xii.

⁴ Texto original: “In particular, I do not believe that utilitarianism can provide a satisfactory account of the basic rights and liberties of citizens as free and equal persons, a requirement of absolutely first importance for an account of democratic institutions”.

⁵ RAWLS, John Bordley. **Political Liberalism**. Nova York: Columbia University Press, 1996, p. 24.

⁶ URBINATI, Nadia. **On John Rawls’ Representational Judgement**. Contemporary Readings in Law and Social Justice, vol. 1, no. 1. Nova York: Addleton Academic Publishers, 2009, p. 10.

⁷ GAUS, Gerald Francis; THRASHER, John. **Rational Choice and the Original Position: The (Many) Models of Rawls and Harsanyi**. In: HINTON, Timothy (org.). **The Original Position**. Cambridge: Cambridge University Press, 2015, p. 51.

⁸ RAWLS, John Bordley. **Political Liberalism**. Nova York: Columbia University Press, 1996, p. 53.

com que princípios pessoas em uma situação ideal de igualdade concordariam para esse fim.

Ele chamou esta situação ideal de igualdade de *posição original*. Seria uma situação hipotética em que aqueles que nela estivessem não saberiam diversos fatos sobre si mesmos, tais como: se são ricos ou pobres, homens ou mulheres, brancos ou negros, se possuem ou não deficiências etc. Ele ainda afirma que as pessoas nessa posição seriam idealmente racionais e desinteressadas quanto à situação dos demais.

Ele sustenta que, nesta posição, seriam escolhidos os seguintes princípios de justiça:

(1) “Todas as pessoas devem ter um direito igual ao esquema mais extensivo de liberdades básicas iguais compatível com um esquema similar de liberdade para os outros”

(2) “Desigualdades econômicas e sociais devem ser arranjadas de maneira que sejam tanto (a) razoavelmente esperadas como sendo para a vantagem de todos e (b) ligadas a posições e cargos igualmente abertos a todos”

Esta monografia irá criticar o raciocínio de Rawls a favor destes princípios ao longo de três capítulos. Inicialmente, será feita uma análise da teoria da escolha racional, já que esta é pressuposta por Rawls em sua argumentação, ao afirmar que as pessoas na posição original seriam idealmente racionais, e também porque essa teoria será utilizada posteriormente quando a teoria de justiça de Rawls for criticada. Esta análise irá basear-se em boa medida na obra “Philosophy, Politics, and Economics”, de Gerald F. Gaus.

No segundo capítulo, será realizada uma reconstrução dos pontos principais da argumentação de Rawls a favor de seus dois princípios de justiça, baseando-se na primeira parte de “Uma Teoria de Justiça”.

Por fim, o terceiro capítulo será dedicado às críticas à teoria de Rawls e à alegação de que ela seria superior ao utilitarismo. Ele será dividido em duas seções: a primeira irá descrever as críticas do economista John Harsanyi à teoria de Rawls, tanto devido ao valor em si dessas críticas quanto devido ao fato de elas servirem de

base para algumas das críticas da segunda seção; a segunda seção, por sua vez, irá tratar das críticas de autoria própria deste trabalho, não se limitando àquelas baseadas na teoria da escolha racional ou nos argumentos de Harsanyi.

Nesta seção, inicialmente é criticado o raciocínio que Rawls afirma que as partes na posição original realizariam para chegar aos seus dois princípios de justiça; em seguida, são oferecidas respostas à crítica de Rawls ao raciocínio que poderia ser realizado na posição original para chegar-se ao princípio da utilidade média; o argumento de Rawls a favor de seus princípios baseado na ideia de que as pessoas na posição original estariam entrando em um acordo (conhecido como argumento dos “esforços do comprometimento”) é, então, questionado, assim como o argumento de que seus princípios seriam superiores ao utilitarismo por mais facilmente inculcarem na população um desejo de agir conforme eles e o argumento de que eles seriam superiores por mais facilmente cultivarem o autorrespeito dos cidadãos; posteriormente, questiona-se o pressuposto de Rawls de que a concepção de justiça de uma sociedade deve ser pública; por fim, são listados alguns temas que deveriam ser objeto de uma teoria de justiça, mas que a teoria de Rawls não é capaz de tratar adequadamente, enquanto o utilitarismo o é.

A conclusão deste trabalho é a de que, embora a obra de Rawls seja um grande marco para a filosofia moral, política e do direito, suas conclusões e seu raciocínio podem ser questionados, sendo igualmente defensável que o utilitarismo seria escolhido pelas partes na posição original buscando princípios de justiça para sua sociedade.

1. A TEORIA DA ESCOLHA RACIONAL

O conceito de racionalidade é extremamente importante para a compreensão da teoria de justiça de Rawls, já que é o que determina a escolha de seus princípios de justiça na posição original. Como essa escolha ocorre será analisado posteriormente. Inicialmente, o foco desta monografia será a discussão sobre o conceito de racionalidade, utilizando-se a obra de Gerald F. Gaus, “On Philosophy, Politics and Economics”, como referência principal.

“Ser racional” é um conceito até certo ponto vago, não havendo unanimidade entre os filósofos e economistas quanto a se certas ações em certas condições seriam ou não racionais. Para os fins desta monografia, serão evitadas teses muito controversas sobre a ideia de racionalidade, utilizando-se apenas algumas de suas características mais aceitas. Axiomas por vezes questionados serão justificados.

A primeira destas características é a de que *a ação de uma pessoa é racional se esta ação, de acordo com suas crenças, é uma maneira efetiva de se atingir um de seus objetivos.*⁹ Por exemplo, se alguém deseja comer um sanduíche e acredita que em sua geladeira encontrará ingredientes não estragados para preparar um sanduíche, sua ação de abrir a porta da geladeira é racional.

O enfoque nas crenças da pessoa é importante porque, se caracterizássemos a ação racional simplesmente como “uma ação efetiva para se atingir seus objetivos”, chegaríamos à conclusão contraintuitiva de que, se a pessoa houvesse se enganado e não restassem mais ingredientes para um sanduíche em sua geladeira, ela teria agido irracionalmente ao abrir a porta.

A caracterização de racionalidade mencionada acima foca apenas em que ações são racionais para a obtenção de um dado objetivo, sem questionar a possibilidade de um objetivo ser ou não racional.

Há um grande debate entre filósofos sobre a possibilidade de objetivos serem ou não irracionais.¹⁰ David Hume foi um grande defensor da tese de que objetivos

⁹ GAUS, Gerald Francis. **On Philosophy, Politics and Economics**. Wadsworth: Cengage Learning, 2008, p. 11.

¹⁰ Ibid., 2008, p. 12.

não podem ser irracionais em si, apenas os meios para atingi-los podem sê-lo.¹¹ Embora tal debate seja bastante frutífero, será deixado de lado na presente monografia, pois, para os fins de uma crítica às conclusões que Rawls extrai do experimento de pensamento da posição original, é desnecessário um comprometimento com teses controversas sobre a irracionalidade de certos fins.

Outro conceito muito importante para a compreensão da teoria da escolha racional moderna é o de “preferência”. A “preferência” se dá sempre entre pelo menos dois estados de coisas, X e Y, podendo um destes ser valorizado mais que o outro (o “preferido”) ou todos serem igualmente desejáveis.

Alguns autores, como Louis Kaplow e Steven Shavell, afirmam que “possuir uma preferência por X” equivaleria a “gostar de X”.^{12 13} Esta visão, contudo, é incorreta, pois podemos imaginar diversos casos em que alguém prefira X dentre as alternativas possíveis e, ainda assim, não goste de X. Por exemplo, alguém pode ir a um restaurante e não gostar de nenhum prato do cardápio, mas prefere comer o ensopado da casa a comer qualquer outro prato, passar fome ou ter de buscar outro local para se alimentar. “Comer o ensopado” é a preferência da pessoa, mesmo não sendo algo de que ela goste.

Também é afirmado por alguns economistas que “ter uma preferência por X a Y” equivaleria a “escolher X entre X e Y”.¹⁴ Esta visão pode ser atraente por aparentemente tornar questões sobre a determinação de preferências mais claras e científicas, livrando-se de aspectos mentais. Ela peca, contudo, por ser demasiadamente reducionista e não explicar adequadamente todos os casos em que alguém possui uma preferência. Imagine alguém que vai ao cinema assistir a um filme novo indicado ao Oscar. Contudo, ao chegar lá, descobre que o filme já havia saído de cartaz, e decide assistir sua segunda opção, uma comédia nacional. Segundo a concepção de preferência aqui desenvolvida, não seria o caso que a

¹¹ HUME, David. **A Treatise of Human Nature**. Oxford: Clarendon Press, 1960, p. 416.

¹² LUCE, Robert Duncan; RAIFFA, Howard. **Games and Decisions: Introduction and Critical Survey**. Nova York: John Wiley and Sons, 1957, p. 21.

¹³ GAUS, Gerald Francis. **On Philosophy, Politics and Economics**. Wadsworth: Cengage Learning, 2008, p. 31.

¹⁴ SAMUELSON, Paul Anthony. **Consumption Theory in Terms of Revealed Preference**. *Economica*, vol. 15, no. 60. Londres: Wiley-Blackwell, 1948.

pessoa preferia ter assistido ao filme indicado ao Oscar, pois ela não efetivamente o escolheu dentre as opções (já que ele não constava dentre elas).¹⁵

Chegamos, por fim, ao conceito de preferência a ser utilizado neste trabalho: um *ranking* que as pessoas estabelecem mentalmente entre diversos estados de coisas, sendo uns preferíveis a outros ou indiferentes entre si.¹⁶ É claro, ninguém cria efetivamente em sua cabeça um *ranking* de todos os estados de coisas possíveis, mas o fato é que as pessoas possuem tal habilidade e, muitas vezes, estabelecem tais *rankings* inconscientemente, tomando decisões cotidianas sem refletir muito sobre suas preferências. Uma pessoa preferir X a Y, portanto, equivale a X estar acima de Y no ranking de preferências daquela pessoa, em um dado contexto.

É uma crítica comum a certas concepções de racionalidade que elas deixariam de lado questões deontológicas, focando apenas em consequências. Não há, contudo, nada na descrição de racionalidade até agora desenvolvida que impeça que questões deontológicas influenciem racionalmente a tomada de decisões de uma pessoa.

Uma pessoa preocupada com questões deontológicas, como, por exemplo, um kantiano que acredite que não pode mentir em nenhuma circunstância, irá simplesmente, em seu *ranking* de preferências, colocar em último lugar aqueles resultados que surgiriam a partir de ela contar uma mentira.¹⁷

Tal pessoa pode colocar em seu *ranking* os estados de coisas que resultarem de não tomar nenhuma ação acima daqueles que resultarem de uma mentira, e, racionalmente, nunca contará uma mentira.

Embora alguns possam alegar que um *ranking* que dê preferência a deixar um amigo morrer a contar uma mentira seja irracional, uma crítica destas entra na questão controversa de se qualquer *ranking* de estados de coisas é racional, ultrapassando o escopo desta monografia.

¹⁵ GAUS, Gerald Francis. **On Philosophy, Politics and Economics**. Wadsworth: Cengage Learning, 2008, p. 33.

¹⁶ Ibid., p. 34.

¹⁷ Ibid., p. 51.

Esta noção de um *ranking* de preferências exige certos axiomas para ser formalizada e utilizada em estudos econômicos. Os axiomas mais comuns para este fim são os da chamada “teoria da utilidade”, que será exposta a seguir. Embora seu nome advinha dos filósofos utilitaristas, como Jeremy Bentham,¹⁸ atualmente a teoria da utilidade é usada para formalizar quaisquer preferências que uma pessoa tenha, não apenas a obtenção do prazer.

Começaremos com os axiomas da “teoria da utilidade ordinal”. Trata-se dos axiomas para uma formalização do *ranking* de preferências de alguém que levam em consideração apenas a ordem das preferências, e não “o quanto mais” alguém prefere X a Y. A teoria da utilidade ordinal é aceita pela maioria dos economistas, sendo muito menos controversa que a chamada “teoria da utilidade cardinal”, pois esta necessita de mais axiomas, cujas justificativas e críticas serão desenvolvidas adiante.

O primeiro destes axiomas é o da *completude*. O que ele define é que o ranking de preferências de uma pessoa racional é completo, ou seja, sendo apresentado um par de opções, ela sempre pode ordenar sua preferência, definindo que X é melhor, pior ou indiferente a Y.¹⁹

Este axioma às vezes é atacado mediante alegações de que uma pessoa racional não precisaria, necessariamente, ter um ranking completo de preferências. Alguém que nunca enfrentasse a possibilidade de experimentar comida asiática não precisaria ter uma preferência entre *sushi* e *tempura* para ser considerado um agente racional.²⁰

O importante, contudo, como mencionado anteriormente, é a potencialidade de ter um *ranking* completo. Ou seja, sendo apresentada a oportunidade de ordenar suas preferências por comidas asiáticas, a pessoa racional deve ter tal capacidade.

Imaginemos que uma pessoa que nunca experimentou comida asiática é levada a um restaurante japonês por seus novos colegas de trabalho. Lá, ela lê o

¹⁸ BENTHAM, Jeremy. **An Introduction to the Principles of Morals and Legislation**. Kitchener: Batoche Books, 2000.

¹⁹ GAUS, Gerald Francis. **On Philosophy, Politics and Economics**. Wadsworth: Cengage Learning, 2008, p. 36.

²⁰ *Ibid.*, p. 40.

menu e precisa escolher entre uma porção de *sushi* ou *tempura*. É senso comum acreditar que uma pessoa racional é capaz de tomar uma decisão.

Ela pode, por exemplo, perguntar os ingredientes da receita e como é preparada para tentar determinar de qual ela gostaria mais. Ainda que isso não fosse possível, ou que as descrições das comidas não tornassem uma mais atraente que a outra, a pessoa ainda deveria ser capaz de dizer “tanto faz” e deixar seus amigos escolherem, mostrando-se indiferente entre as alternativas.

Uma pessoa sem um ranking completo de preferências não conseguiria decidir que comida pedir e, paradoxalmente, não poderia alegar que não possui preferência. Talvez ela pudesse não comer nada, mas, caso seu ranking de preferências colocasse “comer *sushi*” acima de “passar fome” e “comer *tempura*” acima de “passar fome”, porém não estabelecesse qualquer relação entre “comer *sushi*” e “comer *tempura*”, nem mesmo esta opção estaria aberta a ela. Sua reação em uma situação dessas seria um ótimo exemplo de uma atitude considerada irracional.

Após esta breve defesa do axioma da completude, passamos a outro dos axiomas da teoria da utilidade ordinal, o *axioma da transitividade*. Este axioma define que as relações de preferência do *ranking* de preferências de uma pessoa devem ser transitivas.

O que isso significa, em termos mais simples, é que, se uma pessoa prefere X a Y e prefere Y a Z, deve preferir X a Z. Do mesmo modo, se alguém é indiferente entre X e Y e entre Y e Z, também deve ser indiferente entre X e Z.²¹

O axioma da transitividade também é bastante questionado, com algumas pessoas afirmando que não é um pré-requisito da racionalidade ter relações transitivas de preferência. Deve ficar claro, contudo, que apenas preferências não transitivas *simultâneas* violam o axioma da transitividade. Se alguém em um momento prefere ver um filme a ler um livro e ler um livro a ir ao parque, mas, posteriormente, prefere ir ao parque a ver um filme, não estará violando o axioma da

²¹ Ibid., p. 36.

transitividade. Afinal de contas, a pessoa pode ter visto um filme recentemente e não querer fazer a mesma coisa de novo tão cedo.

Ainda assim, como é possível justificar que pessoas com preferências simultâneas não transitivas seriam irracionais? Uma resposta bastante comum é utilizar o experimento de pensamento da “bomba de dinheiro”.²²

Imaginemos uma pessoa que mantenha relações de preferência não transitivas ao mesmo tempo com relação a certas comidas. Ela prefere uma barra de chocolate a uma coxinha, uma coxinha a um pastel e um pastel a uma barra de chocolate. Um dia, no intervalo da faculdade, ela compra uma coxinha no bar da universidade, enquanto seu colega comprou uma barra de chocolate e um pastel.

Ao ver seu colega com uma barra de chocolate, ela sugere trocar sua coxinha pela barra. Como ela prefere a barra, ela está disposta a dar alguma quantidade de dinheiro para compensar seu colega. Digamos que ela lhe dê um real. Após a troca, ela vê que seu colega tem um pastel, que ela prefere a uma barra de chocolate. Então, novamente, ela sugere trocar os lanches, pagando um real. Por fim, ela estará com um pastel, e, como prefere uma coxinha, estará disposta a pagar um real novamente para trocá-los.

Esse ciclo poderia ser repetido indefinidamente, e a pessoa com preferências simultâneas intransitivas perderia cada vez mais dinheiro, voltando sempre ao ponto de partida. Isso seria uma atitude irracional da parte dela, pois estaria perdendo dinheiro sem obter qualquer compensação por isso.

Essa resposta à crítica do axioma da transitividade, contudo, se baseia na ideia de que uma pessoa racional prefere, restando as demais coisas iguais, mais de algo que valoriza do que menos. Sem esta suposição, é mais difícil definir como irracional a atitude da pessoa que é uma “bomba de dinheiro”. Ainda assim, parece uma posição muito implausível sustentar que alguém que gaste dois reais para acabar com o mesmo lanche que já possuía estaria agindo racionalmente.

Há ainda dois outros axiomas da teoria da utilidade ordinal, não tão questionados quanto o da completude e o da transitividade. Estes são o da *reflexividade* e o da *assimetria da preferência e simetria da indiferença*.

²² Ibid., p. 38.

O axioma da reflexividade é, possivelmente, o mais incontroverso. Segundo ele, a relação geral de preferência é reflexiva. O que isso quer dizer é que um objeto deve ser tão preferível quanto ele mesmo no *ranking* de preferências de uma pessoa racional.²³

Uma pessoa racional, portanto, não pode afirmar que não gosta tanto de uma dada fatia de pizza de queijo quanto dela mesma. Se esse fosse o caso, as preferências da pessoa seriam incompreensíveis. É difícil até mesmo de imaginar como uma pessoa dessas poderia agir no dia a dia.

O último axioma determina que, se uma pessoa racional prefere X a Y, ela não prefere Y a X. Ou seja, suas relações de preferência são *assimétricas*. Similarmente, se ela é indiferente entre X e Y, também é indiferente entre Y e X, sendo suas relações de indiferença *simétricas*.²⁴

Este axioma também é menos controverso. É geralmente aceito como uma característica do pensamento racional que a relação entre duas coisas em um *ranking* (no caso, de preferências) não deva ser influenciada pela ordem em que as coisas são mencionadas, já que a ordem de menção não muda as características das coisas em si.

Após esta breve defesa dos axiomas da teoria da utilidade ordinal, será realizada uma defesa dos axiomas da teoria da utilidade *cardinal*. Será sustentado que uma pessoa racional também deve agir de acordo com tais axiomas. Este ponto é muito mais controverso que o anterior e, conseqüentemente, os argumentos são menos conclusivos. Ainda assim, acredito que as críticas aos axiomas não sejam suficientes para levar à rejeição da teoria da utilidade cardinal.

O que é, exatamente, um ordenamento *cardinal* de preferências? É um ranking de preferências em que cada estado de coisas não é apenas definido como preferível ou indiferente a outro, mas, além disso, são definidos *graus* de preferência.²⁵ Em um ranking ordinal de preferências, por exemplo, poderíamos apenas dizer que uma pessoa prefere receber uma nota “A” a uma nota “C” e que prefere receber “C” a ser reprovada em uma matéria da faculdade. Contudo, com um

²³ Ibid., p. 36.

²⁴ Ibid., p. 36.

²⁵ Ibid., pp. 40-41.

ranking cardinal de preferências, podemos dizer que a preferência dessa mesma pessoa em receber a nota “C” está mais próxima à preferência em receber “A” do que à preferência em reprovar.

Isso pode ser representado, por exemplo, mediante números. Em um ranking ordinal, poderíamos representar as preferências da pessoa pelos números 3, 2 e 1 ou por 1.000, 999 e 1 e o resultado seria indiferente: apenas o que importaria seria qual número fosse maior que qual. Em um ranking cardinal, por outro lado, se atribuíssemos os números 1.000, 999 e 1 aos estados de coisas, estaríamos dizendo que a pessoa prefere muito mais o estado de coisas 999 ao estado de coisas 1, e está perto de ser indiferente entre aquele e o estado de coisas 1.000.

Existem diversos axiomas possíveis para o estabelecimento de uma teoria da utilidade cardinal. Esta monografia, contudo, ficará restrita aos axiomas conhecidos como axiomas von Neumann-Morgenstern (devido aos seus desenvolvedores, o matemático John von Neumann e o economista Oskar Morgestern). Estes axiomas são alguns dos mais populares dentre os possíveis para teorias de utilidade cardinal.²⁶

Eles se baseiam, essencialmente, na ideia de *loterias*. As loterias possuem resultados incertos, mas com probabilidades atribuídas a cada estado de coisas. Uma pessoa possui um ranking de preferências ordinal entre os possíveis resultados de uma loteria. A partir do ranking ordinal de preferências dessa mesma pessoa entre diferentes loterias com diferentes probabilidades de gerar cada um dos resultados, é possível estabelecer um ranking cardinal de preferências entre os resultados.^{27 28}

Um exemplo pode tornar esta ideia mais clara. Suponhamos que uma pessoa prefira um *taco* a um *burrito* e um *burrito* a uma *tortilla*. Se quisermos definir a *intensidade* das preferências da pessoa, podemos nos utilizar do conceito de loterias.

²⁶ Ibid., p. 41.

²⁷ NEUMANN, John von; MORGENSTERN, Oskar. **Theory of Games and Economic Behavior**. 3. ed. Princeton: Princeton University Press, 1953.

²⁸ GAUS, Gerald Francis. **On Philosophy, Politics and Economics**. Wadsworth: Cengage Learning, 2008, p. 41.

Inicialmente, atribuímos um valor arbitrário ao *taco* e à *tortilla*, que são os resultados mais e menos preferidos, respectivamente. Utilizemos 1 e 0, por questões de simplicidade. O número entre 1 e 0 que a pessoa atribuir ao *burrito* definirá a intensidade de suas preferências.

Para encontrar este número, oferecemos à pessoa uma escolha dentre duas opções: receber um *burrito* com certeza ou jogar uma loteria que possui 99% de chance de lhe dar uma *tortilla* e 1% de chance de lhe dar um *taco*. Caso a pessoa escolha a certeza do *burrito*, o valor atribuído a ele é maior do que 0,01.

Oferecemos, em sequência, diversas loterias à pessoa, cada uma com probabilidades diferentes de resultar em uma *tortilla* ou um *taco*. Quando encontrarmos uma loteria em que a pessoa esteja indiferente entre jogar nela ou a certeza do *burrito*, teremos encontrado o valor cardinal da utilidade do *burrito* (nesse conjunto de resultados possíveis). Se a pessoa for indiferente, por exemplo, entre a certeza do *burrito* ou uma loteria que lhe dê 50% de chance de obter uma *tortilla* e 50% de chance de obter um *taco*, o *burrito* terá a utilidade de 0,5.

Restando a ideia por trás da teoria de von Neumann e Morgenstern mais clara, passemos à listagem dos axiomas de sua teoria em si. O primeiro destes é o axioma da *continuidade*. Ele determina que as preferências de um agente devem ser contínuas.

Em outras palavras, se o agente prefere X a Y e Y a Z, deve haver uma loteria entre X e Z à qual o agente seja indiferente entre jogar nela ou simplesmente obter Y. Trata-se da garantia de que o raciocínio que desenvolvemos acima será possível em todos os casos, não havendo conjuntos de preferência em que tal indiferença não ocorra nunca.²⁹

Este axioma é, por vezes, criticado. Um experimento de pensamento comum que demonstra tal crítica é o seguinte. Imagine que uma pessoa deva escolher entre a certeza de obter um centavo ou uma loteria que lhe dê uma certa chance de obter um real e uma certa chance de morrer. Segundo o axioma da continuidade, deve haver uma distribuição das chances entre obter um real e morrer em que seja indiferente para a pessoa jogar na loteria ou garantir um centavo. É alegado,

²⁹ Ibid., p. 41.

contudo, que muitas pessoas racionais podem simplesmente nunca aceitar jogar em tal loteria, não havendo, portanto, continuidade entre as suas preferências.³⁰

Eu defendo, contudo, que este argumento é equivocado. Uma pessoa racional deve estar disposta a jogar a loteria em alguma distribuição de chances específica. Para sustentar esta posição, creio que uma análise atentada de nossas decisões cotidianas seria útil.

Constantemente, em nosso dia a dia, tomamos decisões que possivelmente nos matariam. Dirigimos carros, pegamos aviões, andamos por regiões perigosas de uma cidade, compramos comida de vendedores na rua, nadamos na praia e (pelo menos no caso dos estudantes da Universidade Federal do Rio Grande do Sul) entramos em prédios sob risco de desabamento.

Todas essas decisões são tomadas porque se acredita que a probabilidade de morte é suficientemente baixa a ponto de as recompensas por tomá-las compensarem o risco. Igualmente, é de se imaginar que uma pessoa deva preferir jogar na loteria antes mencionada caso a probabilidade de morte seja de 0,0000(imagine aqui 10^1000000 zeros)0000001%. Afinal de contas, nossas ações cotidianas possuem uma probabilidade muito maior de nos matar. Sustento, portanto, que mesmo no contraexemplo antes mencionado, uma pessoa racional deveria ser indiferente entre garantir um centavo ou jogar uma loteria em que a probabilidade de morte seja inimaginavelmente baixa.³¹

O axioma seguinte é o axioma dos *prêmios melhores*. Ele determina que uma pessoa racional, quando oferecida a escolha entre duas loterias, ambas com a mesma probabilidade de gerar cada resultado, sendo um resultado igual entre as duas e o outro resultado de uma superior ao da outra no ranking de preferências desta pessoa, deve escolher a loteria com o prêmio melhor.³²

³⁰ Ibid., p. 44.

³¹ É possível que algumas pessoas ainda se recusassem a participar da loteria que possui a morte como resultado possível, mesmo que a possibilidade de morte fosse incrivelmente menor que em ações nossas do dia a dia. Isto ocorreria pelo *efeito do enquadramento*, em que as pessoas levariam mais a sério a morte como um resultado interno à loteria do que a morte como uma casualidade externa às escolhas tomadas, como um acidente ao se escolher andar de carro. O efeito do enquadramento será trabalhado mais adiante neste capítulo, e será argumentado que ele não é uma crítica válida à teoria da utilidade.

³² GAUS, Gerald Francis. **On Philosophy, Politics and Economics**. Wadsworth: Cengage Learning, 2008, p. 42.

Um exemplo para ilustrar essa noção melhor: imagine que uma pessoa prefira livros de romance policial a livros de terror e livros de terror a romances românticos. Essa pessoa deve escolher entre jogar uma moeda e receber uma cópia de “Um Estudo em Vermelho” caso o resultado seja cara e receber uma cópia de “A Culpa é das Estrelas” caso seja coroa ou jogar a mesma moeda e receber uma cópia de “O Chamado de Cthulhu” caso o resultado seja cara e receber uma cópia de “A Culpa é das Estrelas” caso seja coroa. Supondo que essa pessoa ainda não tenha lido nenhum desses livros ou os tenha em casa, ela deve escolher a loteria que pode lhe proporcionar uma leitura de Sherlock Holmes para agir racionalmente.

Uma objeção ao axioma dos prêmios melhores é desenvolvida por James Dreier, em “Decision Theory and Morality”.³³ ³⁴ Ele nos pede para imaginar que temos um gato e queremos doá-lo para Talia ou Horace. Sabemos que doá-lo para Talia é pelo menos tão bom quanto doá-lo para Horace. Contudo, por questões de justiça, decidimos jogar uma moeda para definir quem receberá o gato.

Essa atitude aparentemente viola o axioma dos prêmios melhores. Podemos redescrever a situação como a escolha entre uma loteria com 50% de chance de doar o gato para Talia e 50% de chance de doar para Horace sendo preferida diante de uma loteria com 50% de chance de doar para Talia e mais 50% de chance de doar para Talia. Como doar para Talia é pelo menos tão bom quanto doar para Horace, preferir a loteria com a possibilidade de doar para Horace deveria ser irracional, mas isso não parece ser o caso.

Trata-se, aqui, de um caso em que é útil lembrar-se que é possível possuir preferências sobre ações e meios de se obter resultados. Se uma pessoa decide jogar a moeda para definir quem terá o gato, é porque essa pessoa não valoriza apenas quem terá o gato, mas também que o processo de definição do dono seja justo em sua visão. Nesse caso, a escolha de jogar a moeda não estaria violando o axioma dos prêmios melhores, já que a loteria que com certeza dará o gato para Talia de maneira injusta e a loteria que possivelmente dará o gato para Horace de

³³ DREIER, James. **Decision Theory and Morality**. In: MELE, Alfred R.; RAWLING, Piers (orgs.). **The Oxford Handbook of Rationality**. Oxford: Oxford University Press, 2004.

³⁴ GAUS, Gerald Francis. **On Philosophy, Politics and Economics**. Wadsworth: Cengage Learning, 2008, p. 44.

uma maneira justa e possivelmente dará o gato para Talia de maneira justa não possuem um prêmio em comum.³⁵

Outro axioma é o das *chances melhores*. Ele determina que, tendo de escolher entre duas loterias que resultam nos mesmos prêmios, uma pessoa racional deve escolher aquela que possui a maior probabilidade de resultar no prêmio preferido.³⁶

Agora que tanto o axioma dos prêmios melhores quanto o das chances melhores foram explicados, podemos nos debruçar sobre uma famosa crítica aos axiomas da teoria da utilidade cardinal, desenvolvida pelo economista Maurice Allais.³⁷

Allais nos pede para imaginar duas loterias hipotéticas, em que os resultados são determinados pela cor da bola que retirarmos de uma urna. Em cada loteria, podemos escolher entre duas opções de recompensa pela retirada das bolas. As loterias são conforme a tabela abaixo:

Opções:	Bola Vermelha (1):	Bola Branca (89):	Bola Azul (10):
Loteria 1 Opção A	US\$ 1.000.000,00	US\$ 1.000.000,00	US\$ 1.000.000,00
Loteria 1 Opção B	US\$ 0,00	US\$ 1.000.000,00	US\$ 5.000.000,00
Loteria 2 Opção C	US\$ 1.000.000,00	US\$ 0,00	US\$ 1.000.000,00
Loteria 2 Opção D	US\$ 0,00	US\$ 0,00	US\$ 5.000.000,00

Segundo os axiomas dos prêmios melhores e chances melhores, nossas escolhas entre loterias devem ser determinadas pelos prêmios e nossas chances de obtê-los. Ou seja, na loteria 1 acima, as bolas brancas não deveriam influenciar nossa escolha, já que possuem o mesmo resultado tanto na opção A quanto na B.

³⁵ Ibid., pp. 45-46.

³⁶ Ibid., p. 42.

³⁷ Ibid., p. 46.

Portanto, se uma pessoa escolher a opção A na loteria 1, isso significa que para ela os US\$ 4.000.000,00 a mais que ela receberia caso pegasse uma bola azul não compensam o risco de não receber nada caso pegue uma bola vermelha. Essa pessoa prefere uma loteria que lhe dá US\$ 1.000.000,00 em 11 de cada 11 jogadas a uma que lhe dá US\$ 5.000.000,00 em 10 de cada 11 jogadas e nada em 1 de cada 11 jogadas.

Se esse for o caso, essa mesma pessoa deveria, na loteria 2, escolher a opção C. Afinal de contas, o resultado das bolas brancas é o mesmo entre C e D, não devendo influenciar a decisão, e as diferenças entre as bolas vermelhas e brancas são as mesmas que na loteria A. O que acontece é que, no experimento de Allais, muitas pessoas escolheram a opção A na loteria 1 e a opção D na loteria 2.

Isto é muitas vezes tido como uma objeção poderosa contra os axiomas von Neumann-Morgenstern. Demonstraria que as pessoas nem sempre agem de acordo com eles, pondo sua plausibilidade em xeque. Entendo, contudo, que não se trata de uma objeção tão forte quanto pode parecer à primeira vista.

Isto porque é possível que as pessoas que escolhem as opções A e D ainda estejam agindo de acordo com os axiomas. No argumento de Allais, há uma premissa oculta: os participantes apenas se importam com o resultado monetário. Este pode, contudo, não ser o caso. Na opção A, as pessoas possuem US\$ 1.000.000,00 garantidos não importa o que aconteça, enquanto que na opção C elas necessariamente possuem o risco de não ganhar nenhum centavo.

É possível que as pessoas valorizem não apenas o dinheiro que podem receber, mas também a garantia de obter uma boa quantia de dinheiro. Elas podem querer evitar o risco de pensar “eu tinha US\$ 1.000.000,00 garantidos, mas perdi isso por ganância”. Se este for o caso, esta preferência entra na função de utilidade das pessoas e os axiomas von Neumann-Morgenstern não mais as obrigariam a escolher a opção C após ter escolhido A.³⁸

Por fim, o último axioma a ser mencionado é o da *redução de loterias compostas*. Ele apenas afirma que, se o prêmio de uma loteria for outra loteria, isso pode ser reduzido para uma loteria simples. Isso faz com que apenas os resultados

³⁸ Ibid., p. 47.

sejam relevantes, e não a quantidade de loterias, retirando da equação o desejo de apostar.³⁹

Por exemplo, se a loteria A possui 50% de chance de lhe fazer jogar a loteria B (que lhe dá 50% de chance de obter um morango e 50% de chance de obter uma banana) e 50% de chance de lhe fazer jogar a loteria C (que lhe dá 50% de chance de obter uma manga e 50% de chance de obter uma uva), podemos simplesmente dizer que a loteria A lhe dá 25% de chance de receber cada uma das frutas.

Embora os axiomas acima explicados sejam diversas vezes criticados, acredito que uma defesa razoável foi aqui desenvolvida, respondendo a algumas das críticas mais comuns.

Há, também, outros motivos para se levar a ideia de utilidade cardinal a sério. O filósofo americano John Pollock desenvolveu um modelo computacionalmente realista de tomada de decisões que utiliza utilidade cardinal e, no processo, estabeleceu argumentos a favor de tal conceito.

Ele afirma que um modelo que utilizasse somente *rankings* de utilidade ordinal para tomar decisões teria de comparar tantos pares possíveis que não seria capaz de funcionar, necessitando de uma estrutura de dados infinita. Ele argumenta, portanto, que agentes reais guardam suas informações de utilidade em formato cardinal.^{40 41}

É relevante também mencionar duas propriedades que normalmente são tidas como necessárias para a tomada de decisões racional: a propriedade da *contração* e a propriedade da *expansão fraca*. A propriedade da contração determina que se um resultado X é escolhido dentre todas as opções de um conjunto C, X também deve ser escolhido em todos os subconjuntos de C em que seja uma opção. De maneira similar, a propriedade da expansão fraca determina que se X é escolhido

³⁹ Ibid., p. 42.

⁴⁰ POLLOCK, John Leslie. **Thinking about Acting: Logical Foundations for Rational Decision Making**. Oxford: Oxford University Press, 2006.

⁴¹ GAUS, Gerald Francis. **On Philosophy, Politics and Economics**. Wadsworth: Cengage Learning, 2008, p. 48.

dentre todas as opções dentro de dois conjuntos, X também deve ser escolhido quando estes conjuntos são unificados.⁴²

Amartya Sen questiona a plausibilidade destas duas propriedades em “Maximization and the Act of Choice”.⁴³ ⁴⁴ Ele menciona o exemplo de alguém que, convidado a uma festa, vê uma cesta de frutas com mangas e maçãs. Essa pessoa prefere mangas a maçãs, mas, como só restou uma manga na cesta, decide pegar uma maçã.

Esta pessoa teria violado a propriedade da contração, pois teria escolhido uma manga em um conjunto que possuísse algumas maçãs e algumas mangas, mas não manteve tal decisão no subconjunto em que há algumas maçãs e apenas uma manga. Sen afirma que tal pessoa não é irracional e que, para compreendermos sua atitude, precisamos diferenciar preferências de deveres ao modelarmos a ação humana.

Creio, contudo, que tal distinção não seja necessária, podendo criar zonas cinzentas e confusões. A atitude de tal pessoa pode ser perfeitamente modelada utilizando-se apenas o conceito de preferência.

Se fosse o caso que a única coisa que importa para essa pessoa é pegar a fruta mais saborosa, ela deveria escolher a manga para agir racionalmente. Contudo, este claramente não é o caso. A pessoa valoriza outras questões, como, por exemplo, não ser a responsável por fazer com que nenhum outro convidado possa mais provar uma manga.

Este valor entraria na função de utilidade da pessoa, assim como o desejo de garantir um mínimo de lucro no caso do paradoxo de Allais e o desejo de se utilizar de um procedimento imparcial no caso de definir se Talia ou Horace receberia um gato. Vemos, portanto, que a escolha no caso em que há apenas uma manga não é

⁴² Ibid., p. 54.

⁴³ SEN, Amartya. **Maximization and the Act of Choice**. *Econometrica*, vol. 65, no. 4. Nova York: The Econometric Society, 1997.

⁴⁴ GAUS, Gerald Francis. **On Philosophy, Politics and Economics**. Wadsworth: Cengage Learning, 2008, p. 54.

um subconjunto da escolha quando há várias, pois no primeiro caso entra uma nova questão em jogo: ser responsável por acabar com as mangas.⁴⁵

Este novo elemento pode ser mais importante para o agente do que a sua preferência por mangas perante maçãs e, portanto, o agente escolherá pegar uma maçã. Vemos, portanto, que não ocorreu no caso nenhuma violação às propriedades da contração e da expansão fraca, e não houve a necessidade de introduzir uma diferenciação entre preferências e deveres.

Dadas todas estas características de como uma pessoa racional deveria organizar suas preferências, resta a pergunta de como tal pessoa lidaria com situações de risco ao tomar escolhas. Como resposta, farei uma breve exposição da chamada *teoria da utilidade esperada* e, posteriormente, mencionarei algumas críticas contra ela. Este ponto será de grande importância para argumentações posteriores desta monografia, em que a decisão dos agentes na posição original sustentada por Rawls será questionada.

A teoria da utilidade esperada é um sistema formal de tomada de decisão em situações de *risco*. Situações de risco são aquelas em que um agente não sabe qual será o resultado de uma ação, mas é capaz de atribuir probabilidades para cada resultado. Diferenciam-se de situações de *incerteza*, em que a falta de informação é tão grande que impede até mesmo a atribuição de uma probabilidade a cada um dos possíveis resultados.⁴⁶

A teoria da utilidade esperada torna-se possível com a utilização de uma ordem cardinal de preferências, não sendo possível apenas com o uso de um *ranking* ordinal. A utilidade esperada de uma ação é equivalente à soma da utilidade de cada um dos possíveis resultados multiplicada pela probabilidade de cada um desses resultados.

Um exemplo de um destes cálculos é o seguinte: imagine uma pessoa em um arraial, onde há uma barraquinha vendendo chocolates. Para alavancar as vendas, o vendedor da barraca estabeleceu um jogo em que cada pessoa possui direito a uma

⁴⁵ Ibid., p. 55.

⁴⁶ Ibid., p. 56.

amostra grátis de bombom. Cada pessoa pode, apenas uma vez, escolher pegar um bombom aleatório de dentro de um saco.

Há dois sacos, um vermelho com bombons de chocolate preto e bombons de cereja e outro azul com bombons de chocolate branco e bombons de creme de avelã. O dono da barraquinha avisa que no saco vermelho $\frac{3}{4}$ dos bombons são de chocolate preto e no saco azul $\frac{3}{4}$ dos bombons são de chocolate branco, pois os de cereja e creme de avelã são mais caros.

Digamos que, para a pessoa, a utilidade de um bombom de creme de avelã seja 10, a de um de cereja seja 8, a de um de chocolate preto seja 4 e a de um de chocolate branco seja 3. Podemos, então, definir qual saco mais provavelmente maximizará sua utilidade. A utilidade do saco vermelho é $(4 \times \frac{3}{4}) + (8 \times \frac{1}{4}) = 5$ e a do saco azul é $(3 \times \frac{3}{4}) + (10 \times \frac{1}{4}) = 4,75$. Para maximizar sua utilidade, portanto, a pessoa deveria escolher o saco vermelho.

Muitos psicólogos e economistas comportamentais vêm criticando a teoria da utilidade esperada, afirmando que ela muitas vezes não representa adequadamente os processos de decisão dos seres humanos. Dois grandes nomes dentre estes críticos são os dos psicólogos cognitivistas Daniel Kahneman e Amos Tversky. Serão mencionados alguns dos pontos levantados por estes autores contra a teoria da utilidade esperada.

O primeiro é com relação à tendência humana de cometer erros em questões que envolvam probabilidade. Richard Nisbett e Lee Ross, por exemplo, desenvolveram o seguinte problema: eles afirmavam ter um amigo professor, tímido, que gostava de escrever poesia e possuía baixa estatura. Em seguida, perguntavam se era mais provável que esse amigo fosse da área da psicologia ou dos estudos chineses.^{47 48}

A pesquisa de Kahneman e Tversky aponta que em situações como esta, a grande maioria das pessoas afirma que o amigo seria da área dos estudos chineses. Isso, contudo, é um equívoco, pois, considerando-se que há muito mais professores

⁴⁷ NISBETT, Richard Eugene; ROSS, Lee. **Human Inference: Strategies and Shortcomings of Social Judgement**. Englewood Cliffs: Prentice-Hall, INC., 1980, p. 25.

⁴⁸ GAUS, Gerald Francis. **On Philosophy, Politics and Economics**. Wadsworth: Cengage Learning, 2008, p. 57.

na área da psicologia que na dos estudos chineses, é muito mais provável que o amigo seja da área da psicologia. Mesmo que tais dados de probabilidade sejam apresentados, as pessoas ainda costumam cometer esta confusão, pois nossos cérebros evoluíram de tal modo que se utilizam de estereótipos para buscar uma coerência narrativa e dar sentido às nossas experiências.^{49 50}

Outra questão que põe em xeque a teoria da utilidade esperada é a *teoria dos prospectos*. Ela questiona o axioma da teoria da utilidade cardinal que afirma que as pessoas apenas possuem interesse nos resultados das loterias.

Um exemplo desenvolvido por Kahneman e Tversky é o da aceitação de duas loterias diferentes.^{51 52} A primeira possuía participação gratuita e tinha uma chance de 10% de dar US\$ 95,00 ao participante e uma chance de 90% de fazê-lo perder US\$ 5,00. A segunda custava US\$ 5,00 para jogar e possuía uma chance de 10% de premiar US\$ 90,00 ao participante e uma chance de 90% de não resultar em nada.

Embora ambas possuam a mesma utilidade esperada de US\$ 5,00, 55 dos 132 participantes aceitaram jogar em uma e não em outra. Destes, 42 rejeitaram a primeira loteria, mas aceitaram a segunda. Uma possível explicação é a de que a primeira loteria invoca a ideia de perda e a segunda a de ganho. As pessoas parecem valorizar mais “não perder x” do que “ganhar x”, o que violaria a ideia da teoria da utilidade esperada de que apenas o valor dos resultados das loterias e suas probabilidades importam.

Há também o chamado *efeito do enquadramento*. Em síntese, a maneira com que as opções são apresentadas influencia a escolha que as pessoas farão, ainda que as opções sejam exatamente as mesmas, apenas “enquadradas” de uma maneira diferente.

⁴⁹ KAHNEMAN, Daniel; TVERSKY, Amos Nathan. **On the Psychology of Prediction**. Psychological Review, vol. 80, no. 4. Washington: American Psychological Association, 1973.

⁵⁰ GAUS, Gerald Francis. **On Philosophy, Politics and Economics**. Wadsworth: Cengage Learning, 2008, pp. 57-58.

⁵¹ KAHNEMAN, Daniel; TVERSKY, Amos Nathan. **Choices, Values, and Frames**. Nova York: Cambridge University Press, 2000, p. 15.

⁵² GAUS, Gerald Francis. **On Philosophy, Politics and Economics**. Wadsworth: Cengage Learning, 2008, pp. 59-60.

Um dos experimentos de Kahneman e Tversky para demonstrar isso é a seguinte pesquisa: os participantes ouviam uma pequena história de que uma doença asiática chegaria aos Estados Unidos e, se não combatida, mataria 600 pessoas. Um primeiro grupo de participantes tinha de escolher entre a adoção de uma política que salvaria 200 pessoas com certeza e uma que tinha uma chance de 1/3 de salvar 600 pessoas e uma chance de 2/3 de não salvar ninguém. 72% dos participantes escolheram a primeira opção.

Posteriormente, a mesma história foi contada a outros participantes, mas, dessa vez, as políticas foram descritas de outro modo: a primeira política faria 400 pessoas morrerem e a segunda tinha uma chance de 1/3 de fazer com que ninguém morresse e uma chance de 2/3 de fazer com que todos morressem. Pode-se perceber que os resultados são os mesmos. Ainda assim, nesse caso, 78% dos participantes escolheram a segunda opção.^{53 54}

As dificuldades humanas em lidar com probabilidade, a teoria dos prospectos e o efeito do enquadramento são apenas alguns exemplos das diversas maneiras com que o raciocínio humano destoa daquilo que é ditado pela teoria da utilidade esperada.

Essas questões tornam, porém, a teoria da utilidade esperada inútil? Não. Elas demonstram, sem dúvida, que a teoria da utilidade esperada não prevê adequadamente o comportamento humano em certas situações. Isto não significa, contudo, que ela não seja um bom esquema para realizar previsões aproximadas em várias ocasiões (como no exemplo citado anteriormente de uma pessoa tendo de escolher de que saco pegaria um bombom, baseando-se em que sabores são seus preferidos).

Mais importante do que isso é o fato de que esses resultados apenas questionam a validade da teoria da utilidade esperada como uma maneira de prever ações de pessoas concretas. Não possuem qualquer influência sobre a teoria de como uma pessoa tida como racional deveria agir.

⁵³ KAHNEMAN, Daniel; TVERSKY, Amos Nathan. **Choices, Values, and Frames**. Nova York: Cambridge University Press, 2000, p. 5.

⁵⁴ GAUS, Gerald Francis. **On Philosophy, Politics and Economics**. Wadsworth: Cengage Learning, 2008, pp. 60-62.

É um fato inegável que os seres humanos não são completamente racionais e muitas vezes cometem erros de julgamento. Os exemplos citados anteriormente demonstram isso, pois são decisões que a maioria das pessoas consideraria irracionais. Se fosse apontado a uma pessoa que tivesse feito escolhas diferentes no caso do efeito do enquadramento que as opções eram na verdade as mesmas, mas descritas de forma diferente, ela provavelmente ficaria surpresa, se sentiria levemente envergonhada por não ter percebido isso e definiria uma escolha independentemente da maneira em que as opções são apresentadas. Dificilmente ela insistiria em escolher opções diferentes a depender de serem mencionadas pessoas salvas ou mortas.

Como a teoria da justiça de Rawls é normativa e pressupõe pessoas idealmente racionais na posição original, podemos utilizar a teoria da utilidade esperada para tentar determinar que princípios de justiça essas pessoas escolheriam. O fato de as pessoas no dia a dia muitas vezes não agirem assim não possui influência sobre esta argumentação.

2. A TEORIA DE JUSTIÇA DE RAWLS

Neste segundo capítulo, será feita uma análise da teoria de justiça de Rawls. Serão mencionados alguns de seus objetivos ao desenvolvê-la, algumas de suas características e será feito um resumo dos principais argumentos a favor dela, como construídos em *Uma Teoria de Justiça*.

Rawls tinha como objetivo, ao desenvolver sua teoria, criar uma teoria de justiça sistemática que se provasse uma alternativa preferível ao utilitarismo, que ele considerava ser a teoria de justiça que havia prevalecido na filosofia política nas décadas anteriores.⁵⁵

Ele via o utilitarismo com maus olhos porque, na visão utilitarista, todos os direitos são estabelecidos e mantidos como meios para se obter a maximização da utilidade. Eles não são, portanto, valorizados em si, e estão sujeitos a limitações e revogações em certos contextos, contanto que isso se prove um meio efetivo de maximizar a utilidade.⁵⁶

Rawls também não desejava recorrer àquilo que ele chamou de intuicionismo: uma teoria de justiça que fosse utilitarista na maioria das situações, mas que, em momentos em que o princípio da maximização da utilidade parecesse levar a decisões inadequadas, simplesmente o abandonasse e determinasse que fosse feito o que parecesse intuitivamente o mais justo em tais situações.⁵⁷

O intuicionismo não se mostrava suficiente, pois não possuía qualquer sistematicidade e dependia exacerbadamente dos julgamentos subjetivos de a quem coubesse tomar decisões. Daí advém a preocupação de Rawls em tornar sua teoria de justiça o mais sistemática, clara e objetiva possível.

A sua teoria de justiça segue a tradição contratualista, encabeçada por filósofos como Locke, Rousseau e Kant. A tradição contratualista é caracterizada por tentar justificar a existência do Estado ou princípios de justiça ao buscar aquilo com

⁵⁵ RAWLS, John Bordley. **A Theory of Justice**. Edição revisada. Cambridge, Massachusetts: Harvard University Press, 1999, p. xvii.

⁵⁶ Ibid., p. xii.

⁵⁷ Ibid., pp. xvii-xviii.

que pessoas racionais concordariam para regular suas relações e vidas em uma situação inicial de igualdade.

Para Rawls, a situação inicial equivale a um acordo hipotético entre pessoas iguais e racionais que não possuem acesso a diversos fatos específicos sobre si, visando garantir imparcialidade. Tal situação é chamada por ele de *posição original* e será desenvolvida em mais detalhe posteriormente nesta monografia.

Rawls afirma que um de seus objetivos era tomar a tradição contratualista e levá-la a um nível maior de abstração, visando atingir os princípios de justiça que idealmente seriam escolhidos após ponderação em uma situação ideal.⁵⁸

Para este fim, e visando evitar atingir o utilitarismo como conclusão, Rawls se utiliza de uma concepção de ser humano autônomo. Para ele, as pessoas devem ser encaradas como agentes que são livres e iguais para exercerem seus dois *poderes morais*: sua capacidade de possuir um *senso de justiça* e sua capacidade de possuir uma *concepção de bem*.

A capacidade de possuir um senso de justiça é o que permite que um agente não apenas seja capaz de atuar de acordo com o que dita a justiça, mas também de compreender, aplicar e agir conforme os princípios de justiça apenas em respeito aos princípios em si.⁵⁹

A capacidade de possuir uma concepção de bem, por sua vez, é o poder de formar e revisar concepções de bem, além de seguir um plano de vida racional. Os seres humanos como caracterizados por Rawls em seu argumento, portanto, não estão presos naturalmente a uma concepção de bem específica, mas são capazes de alterá-la ao longo de suas vidas, a depender das circunstâncias.⁶⁰

Esses dois poderes, combinados, fazem com que cada agente possa ser considerado *completamente autônomo* em relação aos demais.⁶¹ Este ponto se mantém relevante ao longo da argumentação de Rawls, pois ele desenvolve seus

⁵⁸ Ibid., p. xviii.

⁵⁹ FREEMAN, Samuel (org.). **The Cambridge Companion to Rawls**. Cambridge: Cambridge University Press, 2003, p. 5.

⁶⁰ Ibid.

⁶¹ Ibid.

princípios de justiça de uma maneira tal que eles permitam a fruição máxima dos dois poderes morais de cada cidadão.

O senso de justiça de uma pessoa é aquilo que permite a teorização moral. Entender como Rawls considera que a teorização moral ocorre é importante para compreender como ele desenvolve sua teoria de justiça. Ele entende que a teorização moral ocorre por meio do *equilíbrio reflexivo*, que seria a maneira adequada para analisarmos e alterarmos nossas crenças sobre questões éticas. Rawls afirma que sua noção de equilíbrio reflexivo deriva em grande parte de “Outline of a Decision Procedure for Ethics”.⁶²

Ele começa sua caracterização do equilíbrio reflexivo com a suposição de que toda pessoa, atingindo certa idade e tendo as capacidades mentais necessárias, desenvolve um senso de justiça em situações normais. Isso equivale a dizer que essa pessoa é capaz de julgar certas situações como sendo justas ou injustas e de fornecer razões que fundamentem seu juízo. Além disso, as pessoas costumam ter um desejo de agir de acordo com tais juízos e esperam que os demais façam o mesmo.

A teorização moral (ou, especificamente no caso de “Uma Teoria de Justiça”, a teorização sobre a justiça) poderia ser encarada, então, como uma tentativa de encontrar princípios que, quando unidos a nossas crenças e conhecimento das circunstâncias, nos levem à definição de um juízo moral, fundamentado em razões.⁶³

Esses princípios seriam adequados quando refletissem de fato a forma com que realizamos nossos juízos cotidianos sobre questões morais. Eles poderiam ser utilizados, então, como premissas em um argumento que visasse atingir certo juízo moral.

Mas que juízos são adequados para tal fim? Aqueles que forem suficientemente ponderados. Rawls começa afirmando que certos juízos podem ser descartados de antemão: estes são, por exemplo, aqueles dos quais não temos muita certeza, aqueles que foram imaginados em situações não propícias, como em

⁶² RAWLS, John Bordley. **Outline of a Decision Procedure for Ethics**. The Philosophical Review, vol. 60, no. 2. Durham: Duke University Press, 1951.

⁶³ RAWLS, John Bordley. **A Theory of Justice**. Edição revisada. Cambridge, Massachusetts: Harvard University Press, 1999, p. 41.

momentos de medo ou raiva, ou aqueles que sabemos que podem ser utilizados para nosso benefício próprio, sendo, portanto, enviesados. Os juízos devem ser desenvolvidos em circunstâncias favoráveis para o uso de nosso senso de justiça, ou seja, em situações onde as desculpas típicas para erros sobre tais questões não se aplicam.⁶⁴

Rawls passa então a expor a noção de equilíbrio reflexivo. Ela se faz necessária porque, mesmo nossos juízos tendo sido tomados em situações favoráveis, ainda estão sujeitos a irregularidades e distorções. Uma pessoa, ao encarar certos princípios de justiça que parecem incorporar bem suas considerações morais, pode estar disposta a alterar alguns de seus juízos que não estejam de acordo com tais princípios. Isso é ainda mais provável caso a pessoa consiga encontrar uma explicação para tais discrepâncias e os princípios levem a juízos que ela possa aceitar.

Essa é, então, a ideia de um “equilíbrio reflexivo”: alguém em tal estado possui seus juízos ponderados, obtidos em circunstâncias propícias. Passa, posteriormente, a analisar diversas concepções de justiça, checando se elas levariam a seus juízos.

Em caso de discordância com uma concepção de justiça, duas ações podem ser tomadas: a pessoa pode se manter fiel a seu juízo, passando a buscar uma concepção que melhor o reflita (o que pode ser o caso, por exemplo, quando uma concepção de justiça sustentar o instituto da escravidão); ou a pessoa pode alterar seu juízo, tendo considerado a concepção suficientemente adequada (como pode ser o caso de uma pessoa que abandone juízos de superioridade racial ao ser atraída por uma concepção de justiça baseada na dignidade da pessoa humana).⁶⁵

Uma teoria de justiça adequada, portanto, não é simplesmente aquela que se conforma imediatamente a todos os juízos ponderados de uma pessoa. Ela é, na verdade, a teoria que será adotada em um estado de equilíbrio reflexivo, tendo os juízos sidos devidamente questionados.

⁶⁴ Ibid., p. 42.

⁶⁵ Ibid., pp. 42-43.

Rawls levanta ainda a questão de que o equilíbrio reflexivo pode ser encarado de uma maneira restrita, sendo analisadas apenas concepções de justiça que não destoem muito dos juízos de alguém, ou de uma maneira ampla, sendo analisadas todas as concepções de justiça a que alguém possa plausivelmente conformar seus juízos.

Ele afirma que é difícil de imaginar a segunda opção ocorrendo, sendo necessário que nos limitemos a concepções de justiça herdadas pela tradição filosófica e quaisquer demais concepções que nos surjam durante reflexão. Esse é o método que ele utiliza em “Uma Teoria de Justiça”, comparando seus princípios de justiça em especial a duas concepções de utilitarismo, o princípio da utilidade média e o princípio clássico da maximização da utilidade.⁶⁶

Após expor sua noção de teorização moral, Rawls passa a descrever detalhadamente seus dois princípios de justiça, deixando para mais tarde o desenvolvimento dos argumentos a seu favor.

Inicialmente, ele afirma que esses princípios de justiça visam regular o que ele chama de “a estrutura básica da sociedade”. Trata-se das instituições mais básicas de uma sociedade, cujos efeitos sobre a população são vastos e permitem a cooperação de seus membros. Esses princípios determinam a designação de direitos e deveres dentro dessas instituições, além de definir a distribuição dos benefícios e encargos da vida em sociedade.⁶⁷

Esses princípios, portanto, não devem ser confundidos com aqueles que governam as ações de pessoas privadas, como, por exemplo, a maneira com que uma pessoa deveria reagir a um amigo deixando de lhe devolver uma quantia de dinheiro que havia pegado emprestado. Para tais situações, menos abrangentes, princípios próprios de justiça devem ser definidos, e essa não é a tarefa a que Rawls se propôs.

Instituições que Rawls cita como constituindo a estrutura básica da sociedade são: a Constituição política e os principais arranjos econômicos e sociais, aí incluídos a proteção legal da liberdade de pensamento e de consciência, mercados

⁶⁶ Ibid., p. 43.

⁶⁷ Ibid., p. 47.

competitivos, propriedade privada sobre os meios de produção e a família monogâmica.⁶⁸ Ele cita, então, uma versão provisória dos princípios que acredita serem os que seriam escolhidos na posição original:

(1) “Todas as pessoas devem ter um direito igual ao esquema mais extensivo de liberdades básicas iguais compatível com um esquema similar de liberdade para os outros”

(2) “Desigualdades econômicas e sociais devem ser arrançadas de maneira que sejam tanto (a) razoavelmente esperadas como sendo para a vantagem de todos e (b) ligadas a posições e cargos igualmente abertos a todos”⁶⁹

Rawls afirma que há duas ambiguidades nessa formulação do segundo princípio, quais sejam: os termos “vantagem de todos” e “igualmente abertos a todos”. Essas ambiguidades serão esclarecidas mais à frente.

Ele também sustenta que há uma regra de prioridade entre os princípios: o primeiro princípio possui prioridade sobre o segundo.⁷⁰ Isto significa que, em nenhuma circunstância, é possível realizar uma diminuição no esquema de liberdades básicas dos cidadãos visando vantagens sociais ou melhorias econômicas.

Esses princípios são um caso especial de uma regra mais geral de justiça, que é caracterizada da seguinte maneira: “Todos os valores sociais – liberdade e oportunidade, renda e riqueza e as bases sociais do autorrespeito – devem ser distribuídos igualmente, a menos que uma distribuição desigual de algum desses/todos valores seja para a vantagem de todos”.⁷¹

Mas como seria possível medir valores sociais para garantir uma distribuição igualitária destes? Para este fim, Rawls desenvolve o conceito de *bens primários*. Bens primários são bens que se espera que qualquer pessoa racional desejaria ter mais, pois são necessários para a realização de qualquer concepção de bem.⁷²

⁶⁸ Ibid., p. 6.

⁶⁹ Ibid., p. 53.

⁷⁰ Ibid., pp. 53-54.

⁷¹ Ibid., p. 54.

⁷² Ibid., p. 79.

Eles permitem que se tenha uma medida geral de o quão boa seria a situação em que uma pessoa se encontra. Tentar medir a felicidade ou utilidade de alguém é muito difícil, e realizar comparações entre essa pessoa e outros membros da sociedade é uma tarefa ainda mais complexa. Ao listar certos bens que, se assume, qualquer pessoa desejaria possuir mais, Rawls passa a ter uma medida mais objetiva e comparável do bem-estar entre as pessoas. É um esquema que, embora possa gerar certas simplificações, possui a vantagem de ser muito mais efetivo na prática.

Alguns dos principais bens primários são direitos (como o direito à saúde), liberdades (como a de associação), oportunidades (como as de se candidatar a empregos), renda e riqueza e as bases sociais do autorrespeito. As bases sociais do autorrespeito podem à primeira vista não parecer ser algo tão importante, mas são muito relevantes, pois são o que fazem com que uma pessoa se valorize e busque seguir seu plano de vida e atingir sua concepção de bem. Sem tais bases, as pessoas podem deixar de ver um propósito em seguir seus objetivos, o que deixaria todos os demais bens primários sem qualquer utilidade.⁷³

Os bens primários até aqui citados são bens sociais. Há também saúde e vigor e inteligência e imaginação, que são bens naturais. Embora a sua posse seja influenciada pela estrutura básica da sociedade, sua distribuição não se encontra sob um controle tão fácil quanto os bens sociais.⁷⁴

A situação de uma pessoa é definida, portanto, pelas suas expectativas de obter bens primários ao longo da vida. E, pelos princípios de justiça acima mencionados, ela não pode diminuir liberdades básicas suas ou de outros visando aumentar a expectativa de renda, riqueza e oportunidades sua ou de outros.

Após tratar da prioridade do primeiro princípio de justiça sobre o segundo, Rawls passa a desenvolver o segundo princípio, tratando das ambiguidades dos termos “vantagem de todos” e “igualmente abertos a todos”. Ele afirma que esses termos são independentes e, a depender de como cada um for interpretado, o segundo princípio poderia apoiar quatro organizações sociais distintas.⁷⁵

⁷³ Ibid.

⁷⁴ Ibid., p. 54.

⁷⁵ Ibid., p. 57.

	“Vantagem	de todos”
“Iguamente abertos a todos”	Princípio da eficiência	Princípio da diferença
Igualdade enquanto carreiras abertas a talentos	<i>Sistema de Liberdade Natural</i>	<i>Aristocracia Natural</i>
Igualdade enquanto oportunidades iguais	<i>Igualdade Liberal</i>	<i>Igualdade Democrática</i>

Ele começa explicando a interpretação que chamou de “sistema de liberdade natural”. Segundo essa interpretação, o princípio determinaria que as instituições básicas da sociedade devem ser moldadas para ser mais eficientes e, no sistema social, as carreiras são abertas a talentos.

É feita, então, uma explicação do conceito de eficiência, que para Rawls equivale, nesse caso, ao ótimo de Pareto. Segundo esse princípio, uma distribuição de bens é eficiente se for impossível alterá-la sem deixar pelo menos uma pessoa em situação pior.

Esse princípio é negado por Rawls porque leva a conclusões claramente injustas quando aplicado à estrutura básica de uma sociedade. Segundo tal princípio, por exemplo, seria impossível acabar com um regime escravagista que estivesse estabelecido em um país. Afinal de contas, acabar com a escravidão deixaria os donos e comerciantes de escravos em uma situação pior do que a em que estavam antes. O princípio da eficiência, portanto, não é adequado por si só para a estrutura básica de uma sociedade.

Esse princípio deve ser limitado, portanto, e isso ocorreria, nessa interpretação, pela liberdade formal assegurada pelo primeiro princípio de justiça e pelas oportunidades de emprego abertas a todos, assim como um livre mercado. Essa interpretação não se esforça, contudo, para manter uma igualdade entre os cidadãos. Assim, embora as oportunidades de empregos estejam abertas a todos, as capacidades das pessoas de atingi-las são influenciadas por fatores arbitrários de um ponto de vista moral, como a acumulação de recursos financeiros pela sua

família devido a talentos naturais de seus antepassados. Essa interpretação do segundo princípio de justiça é, portanto, rechaçada.⁷⁶

A chamada interpretação da “igualdade liberal” tenta resolver essa questão. Segundo essa interpretação, a sociedade deveria tomar medidas que minimizassem as influências de fatores como classe social e renda e riqueza na obtenção de empregos pelas pessoas.

Essa interpretação, contudo, não resolve todos os problemas. Ela ainda permite que certas pessoas possuam vantagem ao obter empregos por questões arbitrárias, como talentos que possuem naturalmente. Além disso, questões tão básicas quanto uma pessoa ter ambição e capacidade de se esforçar são influenciadas por questões familiares e naturais, o que também não é tratado por essa interpretação.⁷⁷

Rawls afirma, portanto, que é necessário adotar a interpretação da “igualdade democrática”, mas antes menciona brevemente a interpretação da “aristocracia natural”. Trata-se de um sistema em que não se busca limitar contingências sociais para além do requisitado pelas liberdades mínimas, mas em que as vantagens das pessoas que possuem mais dotes devem ser apenas aquelas que melhorem a situação dos menos afortunados. Há aí, portanto, a ideia de *noblesse oblige*, tendo os mais afortunados o direito de manter sua posição, mas também o dever de ajudar os mais pobres.

Tanto a interpretação liberal quanto a da aristocracia natural seriam inerentemente instáveis. Isso se deve ao fato de que cada uma combate um tipo de desigualdade arbitrária (contingências sociais ou dotes naturais), mas permite a manutenção do outro. Os membros de uma sociedade que seguisse uma dessas interpretações, conseqüentemente, se perguntariam por que um tipo de desigualdade é combatido e outro não, e isso os levaria a rechaçar o sistema.⁷⁸

Chegamos, enfim, à interpretação que Rawls considera representar o que seria escolhido na posição original. Segundo a interpretação democrática, todos devem ter oportunidades iguais de obter empregos e posições de poder e as

⁷⁶ Ibid., pp. 62-63.

⁷⁷ Ibid., p. 64.

⁷⁸ Ibid.

desigualdades econômicas são justas apenas enquanto a vantagem dos mais bem-afortunados fizer parte de um esquema que aumenta os prospectos daqueles em posições sociais mais baixas.⁷⁹

Essa segunda parte do princípio é denominada *princípio da diferença*, e se distingue muito tanto do princípio da maximização da utilidade quanto do princípio da eficiência (entendido acima como equivalente ao ótimo de Pareto).

O princípio da utilidade determina, simplesmente, que a utilidade total da sociedade deve ser maximizada, levando-se em consideração todos os seus membros. É possível, portanto, diminuir um pouco a utilidade da classe social mais desfavorecida se isso resultar em um benefício grande para as classes sociais mais abastadas a ponto de compensar tal perda. Isso não é permitido pelo princípio da diferença.

O princípio da eficiência, por sua vez, não permite prejudicar uma classe social em benefício de outra. Ele permite, contudo, que as preferências de uma classe social mais alta sejam satisfeitas, desde que isso não piore a situação dos mais pobres. O princípio da diferença adiciona uma cláusula nesse caso, e determina que, para que uma classe seja mais rica que outra, não basta que os mais pobres não sejam prejudicados por isso, mas também se faz necessário que essa desigualdade seja vantajosa para eles.

O princípio da diferença é, em muitos aspectos, similar a um princípio da ciência econômica, conhecido como *regra maximin* (de “maximizar o mínimo”). Essa regra é um critério de decisão em situações de grande aversão a risco e determina que as escolhas devem ser tomadas de maneira tal que maximizem o mínimo que uma pessoa pode receber em dada situação. Rawls nota, contudo, que se referir ao princípio da diferença simplesmente como “regra *maximin*” pode levar a certas confusões, pois o princípio da diferença é um princípio de justiça decidido na posição original, onde grande aversão ao risco não é postulada.⁸⁰

Ainda desenvolvendo a interpretação do segundo princípio, Rawls afirma que há, dentro dele, um *ordenamento léxico* entre suas partes. Assim como o primeiro

⁷⁹ Ibid., p. 65.

⁸⁰ Ibid., pp. 72-73.

princípio de justiça é lexicamente anterior ao segundo (ou seja, não é possível sacrificar o cumprimento do primeiro princípio visando um melhor cumprimento do segundo), a parte do segundo princípio referente a manter todas as carreiras e posições de poder igualmente abertas a todos é lexicamente anterior ao princípio da diferença. Isso significa, em outras palavras, que não é permitido pelos princípios de justiça impedir que certas pessoas possam ter certos empregos, mesmo que isso implicasse um aumento na expectativa de bens primários de todos, inclusive a parcela mais pobre da população.⁸¹

Rawls posteriormente aponta outro fato que torna sua teoria de justiça mais fácil de aplicar na prática que o utilitarismo. Não apenas as comparações interpessoais de utilidade são evitadas com o uso de expectativas sobre bens primários, mas também há o fato de o princípio da diferença permitir que, para avaliar a justiça de certa distribuição de bens primários, baste que se identifique o grupo social em maior desvantagem.

Embora essa seja uma tarefa que traga seus próprios desafios (como definir que grupos sociais devem ser considerados?), ela é mais simples que medir a utilidade de cada grupo de pessoas, e permite que se utilize apenas um ranking ordinal, não cardinal, pois basta identificar o grupo social em maior desvantagem, sendo desnecessário determinar o quão mais desvantajosa seria a posição deste grupo quando comparado aos demais. Assim que o grupo mais desafortunado for definido, para determinar a justiça da distribuição de bens basta checar se ele está na melhor posição que poderia estar e se as desigualdades da sociedade contribuem para a melhoria de suas expectativas sobre bens primários.⁸²

Analisar a justiça da distribuição de bens primários, contudo, não é algo necessário em qualquer circunstância. Isto só é necessário no que Rawls chama de “as circunstâncias da justiça”. Trata-se de uma ideia exposta originalmente por Hume, e equivale às condições normais em que a cooperação humana é tanto possível quanto necessária.⁸³

⁸¹ Ibid., pp. 73, 77 e 266.

⁸² Ibid., pp. 78-79.

⁸³ Ibid., pp. 109-110.

Trata-se de um estado de coisas em que há tanto uma identidade quanto um conflito de interesses entre as pessoas. Há uma identidade de interesses porque, mediante cooperação e ajuda mútua, todas as pessoas podem avançar seus objetivos mais do que conseguiriam isoladamente. Há um conflito de interesses, por outro lado, pois as pessoas não são indiferentes quanto à distribuição dos resultados de seus trabalhos conjuntos e cada uma prefere uma parcela maior desses para atingir seus objetivos. Princípios de justiça são necessários, conseqüentemente, para definir de que maneira serão distribuídos os recursos. Esse é o papel da justiça, e as circunstâncias de justiça são aquelas que tornam os princípios necessários.

As circunstâncias de justiça podem ser subdivididas em circunstâncias objetivas e subjetivas. As circunstâncias objetivas se referem a fatos sobre o mundo distintos da subjetividade de cada uma das pessoas de uma sociedade. Aí são incluídos: o fato de muitas pessoas viverem ao mesmo tempo em uma região geográfica, serem aproximadamente similares quanto a capacidades físicas e mentais (a ponto de não ser possível uma única pessoa dominar os demais) e serem vulneráveis a ataques e impedimentos de seus planos por uma união das demais pessoas.

Há ainda a circunstância da escassez moderada. Isso significa que os recursos disponíveis para a sociedade não são abundantes a ponto de tornar desnecessária qualquer cooperação, mas, ao mesmo tempo, não são escassos a ponto de tornar cooperações infrutíferas. Empreitadas mutuamente vantajosas são possíveis, mas os benefícios advindos delas são insuficientes para satisfazer todos os desejos dos homens.

As circunstâncias subjetivas, por sua vez, se referem a aspectos dos sujeitos que cooperam entre si na sociedade. Esses sujeitos possuem necessidades e interesses até certo ponto similares ou, pelo menos, complementares, o que permite a cooperação. Ainda assim, cada um possui um plano de vida próprio, ou concepção de bem.

Essas concepções de bem distintas geram certos conflitos, pois as pessoas farão reivindicações conflitantes sobre os recursos visando atingir seus objetivos.

Embora tais concepções não precisem ser egoístas, são consideradas por cada um como dignas de reconhecimento e devendo ser satisfeitas.

As pessoas nas circunstâncias de justiça também possuem várias deficiências quanto a conhecimento, pensamento e julgamento. Elas podem constantemente errar, seu raciocínio é muitas vezes defeituoso e seus julgamentos são influenciados por questões como ansiedade e preconceito. Isso faz com que existam, nas circunstâncias de justiça, diversas visões de mundo diferentes, aí contidas doutrinas filosóficas, religiões e ideologias políticas.⁸⁴

Rawls esclarece que ele assume que as pessoas na posição original sabem que as circunstâncias de justiça se aplicam à sua sociedade. Afinal, se não fosse esse o caso, não haveria qualquer motivo para buscar um consenso sobre princípios de justiça.

Ele também assume que os sujeitos na posição original são mutuamente desinteressados. Isso significa que eles não estão dispostos a sacrificar seus interesses para satisfazer os interesses dos outros, mas isso não implica que seus interesses tenham de ser egoístas. Isso é assumido porque, caso não fosse esse o caso, todas as pessoas trabalhariam em conjunto para satisfazer ao máximo os desejos de todos, não havendo conflitos e, conseqüentemente, necessidade para princípios de justiça.⁸⁵

O último tópico relevante antes de se reconstruir os argumentos de Rawls a favor de seus princípios de justiça são os limites formais a que as concepções de justiça a serem analisadas na posição original estão sujeitas. Esses limites provêm da noção de moralmente correto adotada por Rawls.⁸⁶

A primeira condição desses princípios é a de que eles devem ser *gerais*. Isso implica que deve ser possível formulá-los sem apelar a nomes próprios ou “descrições definidas fraudadas” (maneiras de se referir a um objeto específico sem utilizar o seu nome próprio). Os predicados utilizados nos princípios devem, portanto, expressar relações e propriedades gerais. Rawls reconhece que parece haver grandes dificuldades filosóficas em definir exatamente quando uma

⁸⁴ Ibid., p. 110.

⁸⁵ Ibid., p. 112.

⁸⁶ Ibid.

propriedade ou relação seria geral, mas afirma que não acha imperativo resolver tais questões para os fins de uma teoria de justiça como a que ele busca desenvolver.⁸⁷

Essa condição se deve em parte ao fato de que os princípios de justiça devem organizar uma dada sociedade ao longo do tempo, podendo diferentes gerações compreendê-los. Para esse fim, é importante que os princípios não façam menção a contingências como pessoas ou associações específicas.

A segunda condição é a de que os princípios sejam *universais*, ou seja, se apliquem a todos devido a seu *status* de pessoa moral. Isso faz com que eles necessariamente tenham de ser de uma complexidade tal que todos os membros da sociedade possam compreendê-los. Além disso, todos devem agir de acordo com os princípios, então princípios que derrotem a si mesmos caso todos os sigam ou que exijam que alguns sigam outros princípios para que possam ser cumpridos não estão dentre as alternativas.⁸⁸

A terceira condição é a da *publicidade*, típica de visões contratualistas de justiça. As partes na posição original escolhem princípios de justiça para compor uma concepção pública de justiça. Isto significa que todos sabem os fatos mais básicos sobre esses princípios, fatos estes que alguém que tivesse concordado com os princípios saberia. Esse conhecimento geral possui consequências benéficas, como uma contribuição para a estabilidade da cooperação social e a efetividade da aplicação dos princípios no cotidiano.⁸⁹

A quarta condição é a de que uma concepção de moralmente correto imponha um *ordenamento em reivindicações conflitantes*. Isso decorre da função da justiça de resolver demandas antagônicas entre as pessoas. É desejável que esse ordenamento seja completo e transitivo (como tais noções foram desenvolvidas no primeiro capítulo desta monografia), mas certas maneiras de obter tais ordenamentos, como julgamento por combate, são rechaçadas por Rawls por se utilizarem de “aspectos não relevantes” de pessoas e suas situações.⁹⁰

⁸⁷ Ibid., p. 113.

⁸⁸ Ibid., p. 114.

⁸⁹ Ibid., p. 115.

⁹⁰ Ibid., pp. 115-116.

A última condição é a da *finalidade*. Os princípios de justiça devem ser os últimos critérios para se julgar o raciocínio prático. Não há, portanto, considerações superiores para se utilizar em argumentos sobre ética, e o raciocínio a partir desses princípios é conclusivo.⁹¹

Essas cinco condições, tomadas em conjunto, não excluem a grande maioria das teorias éticas legadas pela tradição filosófica. Elas tornam, contudo, inviáveis diferentes formas de egoísmo. A ditadura de uma pessoa e o egoísmo “*free-rider*” violam o requisito da generalidade, enquanto o egoísmo geral, em que cada um possui a liberdade de buscar seus objetivos sem levar os demais em consideração, é excluído por não ser capaz de produzir um ordenamento adequado das diferentes reivindicações.⁹²

Iniciamos agora uma reconstrução da argumentação de Rawls a favor de seus princípios de justiça, cujo primeiro passo consiste em uma análise mais detalhada de como é caracterizada a posição original. Ele busca com tal caracterização utilizar a justiça puramente procedimental como base de sua teoria, anulando os efeitos de contingências que tentam as pessoas a agir umas contra as outras ou usar o ambiente para sua vantagem.⁹³

Rawls então cria a ideia do “véu da ignorância”. Trata-se de uma metáfora, pois as pessoas na posição original desconheceriam diversos fatos sobre suas vidas e até mesmo sobre sua sociedade, como se estivessem por trás de um véu que lhes impede ter acesso a todas informações que normalmente conheceriam.

Uma das coisas que as pessoas não saberiam na posição original é seu lugar na sociedade. Ninguém saberia se é uma pessoa rica ou pobre, de uma família abastada ou humilde, com um bom emprego estável ou desempregada. Tampouco saberia que dons e habilidades naturais possui. Questões como inteligência, saúde e talentos para certas funções, como tocar instrumentos, são, portanto, uma incógnita.

Uma pessoa na posição original também não sabe sua concepção de bem. Consequentemente, não tem ideia de qual é seu plano de vida. Rawls chega a

⁹¹ Ibid., pp. 116-117.

⁹² Ibid., p. 117.

⁹³ Ibid., p. 118.

afirmar que a pessoa até mesmo não saberia sobre as características específicas de sua psicologia, como sua aversão a risco ou tendência a encarar situações de maneira otimista ou pessimista.

Tampouco possuem as partes informações sobre o estado de sua sociedade, tais como: nível econômico, situação política, nível de civilização ou cultural. Elas também não sabem a que geração da sociedade pertencem, o que é justificado pelo fato de questões de justiça não ocorrerem apenas dentro de uma dada geração, mas também entre diferentes gerações, como no caso de se determinar uma poupança adequada de capital ou nível de preservação do meio ambiente para a geração seguinte.

O único fato específico sobre sua sociedade a que as pessoas possuem acesso é o fato de esta se encontrar nas circunstâncias de justiça. Contudo, elas também sabem fatos gerais sobre a sociedade humana. Questões políticas e princípios de teoria econômica são compreendidos, assim como as bases da organização social e as leis da psicologia humana.

Esses conhecimentos gerais são justificados porque uma teoria de justiça deve ser aplicável e, caso entre em conflito com leis básicas de economia ou psicologia, não é viável. Além disso, uma concepção de justiça adequada deveria gerar nos membros da sociedade o desejo de mantê-la, o que não seria possível caso ela contradissesse tais leis.⁹⁴

Rawls esclarece, buscando responder a possíveis críticas, que a posição original não se refere a qualquer acordo realizado de fato por pessoas em algum momento histórico, mas sim a um acordo hipotético que pessoas racionais escolheriam estando sujeitas às restrições do véu da ignorância. Ele afirma que qualquer um pode imaginar esse acordo, simplesmente abstraindo todas as informações irrelevantes que possui de si e se imaginando na posição original.⁹⁵

Ele aponta também que, como todas as pessoas na posição original possuem as mesmas informações, nenhuma pode barganhar buscando obter uma vantagem

⁹⁴ Ibid., pp. 118-119.

⁹⁵ Ibid., p. 119.

específica para si, já que não sabe quem será na sociedade e, conseqüentemente, que princípios serão vantajosos para si.⁹⁶

Como último apontamento, Rawls relembra o fato de que os princípios de justiça devem ser públicos e guiar a cooperação entre os membros da sociedade. Isso implica que, mesmo todos os tipos de fatos gerais sobre sociedades estando disponíveis para a análise das partes na posição original, concepções de justiça que se utilizem de fatos menos complexos devem ser preferidas. Ele encara isso como um ponto favorável a seus dois princípios de justiça quando comparados com visões utilitaristas, pois estas exigiriam muita informação e cálculos complexos.⁹⁷

Rawls passa então a analisar a racionalidade das pessoas na posição original. Afirma que, embora elas não saibam sua concepção de bem, sabem que, muito provavelmente, preferirão mais bens primários a menos, já que esses são efetivos em promover qualquer concepção razoável de bem. Também possuem o desejo de proteger suas liberdades básicas e aumentar suas oportunidades.

Ele sustenta que sua visão de racionalidade é a visão tradicional da teoria social ("*standard one familiar in social theory*"), de pessoas com um conjunto coerente de preferências, ordenadas com base em quão bem elas promovem seus propósitos.⁹⁸ Temos aqui, portanto, uma ideia de ranking de utilidade ordinal como desenvolvida no primeiro capítulo desta monografia.

Há mais uma característica da racionalidade dos indivíduos, decorrente de seu mútuo desinteresse: eles não sofrem de inveja, não estando dispostos a sofrer uma perda apenas se os demais também a sofrerem e não se sentindo pior ao ver outros em situação melhor que a sua. É reconhecido que tal ideia é controversa, mas Rawls afirma que primeiramente estabelecerá um argumento a favor de seus princípios de justiça assumindo que a inveja não existe e apontará que o arranjo social conseqüente, por sua própria natureza, dificulta o surgimento da inveja, sendo inerentemente estável.⁹⁹

⁹⁶ Ibid., pp. 120-121.

⁹⁷ Ibid., pp. 122-123.

⁹⁸ Ibid., pp. 123-124.

⁹⁹ Ibid., pp. 124-125.

Por fim, é assumido que as pessoas na posição original possuem um senso de justiça e que esse fato é conhecimento comum entre elas. Essa condição faz com que todos tenham a garantia de que, após princípios de justiça serem definidos, todos serão capazes de compreendê-los e agir de acordo com eles. Rawls chama isso de “esforços do comprometimento” (tradução livre de “*strains of commitment*”). Isso não implica, contudo, que não se deva levar em consideração o fato de alguns princípios serem mais estáveis que outros.¹⁰⁰

Passamos, então, a analisar a justificação dos dois princípios de justiça a partir da posição original. Rawls formula seu argumento por meio de uma comparação hipotética pelas pessoas na posição original entre seus princípios e outros princípios possíveis.

Em “Uma Teoria de Justiça”, é feita a comparação dos princípios de Rawls com o princípio da utilidade média e o princípio clássico da maximização da utilidade. Nesta monografia, contudo, me restringirei a analisar a comparação com o princípio da utilidade média, pois minha crítica à teoria de Rawls será em grande parte uma defesa de que esse princípio seria escolhido por pessoas racionais na posição original.

Embora o princípio da maximização da utilidade também possa ser defendido, sua defesa possuiria uma natureza muito distinta daquela dos princípios de Rawls ou do princípio da utilidade média. Enquanto o princípio da utilidade média visa maximizar a “utilidade *per capita*” da sociedade, o princípio da maximização da utilidade implica que, restando as demais coisas iguais, se a população de uma sociedade dobra, sua utilidade dobra. Este princípio exige, portanto, que as instituições da sociedade sejam arranjadas de maneira tal que estimulem o crescimento indefinido da população, desde que isso cause uma diminuição na utilidade média devagar a ponto de o aumento da população ainda estar aumentando a utilidade total. Do ponto de vista das partes na posição original, esse princípio não parece avançar seus interesses e, conseqüentemente, seria difícil desenvolver uma defesa dele utilizando-se da ideia de consenso na posição original, como Rawls mesmo aponta.¹⁰¹

¹⁰⁰ Ibid., pp. 125-126.

¹⁰¹ Ibid., pp. 160-161.

Rawls inicia a justificação de seus princípios de justiça mostrando que raciocínio seria utilizado pelas partes para chegar a eles. Ele afirma que, devido ao fato de elas não possuírem informações sobre si e não poderem buscar princípios que lhes favorecessem, o primeiro princípio que lhes surgiria na mente seria uma garantia igual de liberdades mínimas, igualdade de oportunidades e distribuição igualitária de renda e riqueza.¹⁰²

Deve-se considerar, contudo, que diferenças na distribuição de recursos financeiros podem aumentar a eficiência econômica da sociedade, o que beneficiaria a todos. Não haveria motivos, portanto, para alguém se opor a isso, especialmente considerando-se que Rawls imaginou as pessoas na posição original sem inveja. Estando garantidas liberdades mínimas e igualdade de oportunidades, chega-se, assim, ao princípio da diferença.¹⁰³

Em seguida Rawls tenta justificar seus princípios de justiça utilizando-se da semelhança entre eles e a regra *maximin* de escolha sob incerteza. Embora a regra *maximin* não seja um critério adequado para escolha em diversas situações, há momentos em que ela é a escolha racional a ser tomada, como no caso em que a posição na sociedade de uma pessoa que deve definir a distribuição de recursos será definida por um inimigo desta.

Esse não é, claramente, o caso da posição original. Isso demonstra, porém, que há certas condições que, se cumpridas, tornam a regra *maximin* um critério racional de decisão. Rawls tenta demonstrar que essas condições são cumpridas na posição original, o que daria uma justificativa maior aos seus princípios de justiça.¹⁰⁴

A primeira dessas aparentes condições é a de que se trate de uma situação em que conhecimento sobre a probabilidade de diferentes resultados é impossível ou extremamente inseguro. Isso levaria a um ceticismo sobre cálculos probabilísticos, especialmente quando a decisão tivesse de ser justificada a terceiros.

A segunda é a de que a pessoa tomando a decisão possua uma concepção de bem tal que ela se importe muito pouco com o que pode ganhar acima do mínimo

¹⁰² Ibid., p. 130.

¹⁰³ Ibid., pp. 130-131.

¹⁰⁴ Ibid., p. 133.

que obteria seguindo a regra *maximin*, não valendo à pena tomar riscos visando ganhos maiores.

A terceira e última é a de que as alternativas recusadas sejam tais que dificilmente poderiam ser aceitas, ou seja, a situação envolve riscos de grandes perdas.¹⁰⁵

Analisando como a posição original poderia incorporar tais condições, Rawls lembra inicialmente que o véu da ignorância remove todas as informações sobre a situação da sociedade das partes, impossibilitando cálculos probabilísticos. Além disso, as partes devem levar em consideração que sua escolha de princípios deve ser justificável a terceiros, como seus descendentes que viverão sob eles. A primeira condição, portanto, é cumprida.

A favor da segunda condição, Rawls sustenta que os dois princípios de justiça são suficientes para garantir um mínimo satisfatório aos membros da sociedade e, sendo a liberdade prioritária aos ganhos socioeconômicos, as partes na posição original não teriam interesse em arriscá-la visando obter ganhos maiores.

A terceira condição é cumprida pelo fato de concepções alternativas de justiça, como o utilitarismo, poderem, teoricamente, sustentar regimes inaceitáveis, como a escravidão. Rawls admite que essa é uma questão controversa, não sendo claro se alguma forma de regime escravagista poderia, em alguma situação, maximizar a utilidade de uma sociedade, mas afirma que não é necessário para sua argumentação analisar a verdade dessa questão, bastando apontar que o utilitarismo teoricamente poderia levar a um resultado inaceitável, ficando a cargo das circunstâncias esse ser o caso ou não.¹⁰⁶

Com isso, Rawls pretende ter demonstrado que o contexto da posição original incorpora em grande parte as condições que tornam racional a utilização da regra *maximin* de escolha sob incerteza. As partes em tal posição, portanto, escolheriam racionalmente seus dois princípios de justiça.

Em seguida ele responde a uma objeção que já imaginava que seria comum: a de que, segundo tais princípios de justiça, seria correto sacrificar enormes ganhos

¹⁰⁵ Ibid., p. 134.

¹⁰⁶ Ibid., p. 135.

para uma grande parte da sociedade em troca de ganhos ínfimos aos menos favorecidos, o que não pareceria ser uma escolha racional ao se levar em conta os interesses de todos os membros da sociedade.¹⁰⁷ Uma política pública hipotética que levaria a um aumento geral de R\$ 4.000,00 no salário de todos os membros da classe média, por exemplo, seria injustificada se tivesse como efeito colateral uma perda de R\$ 0,10 nos ganhos da parcela mais pobre da população, mesmo que essa parcela equivalesse a 0,1% da população e a classe média a 50% dela.

Como resposta a tal objeção, Rawls aponta que seus princípios de justiça não têm como objetivo determinar distribuições de renda, propriedade ou serviços dentre dados indivíduos. Trata-se de princípios a serem aplicados à estrutura básica da sociedade, ou seja, suas regras e instituições mais elementares, afetando questões como serviços públicos e direitos de propriedade. Esses princípios seriam aplicados conjuntamente à estrutura e os princípios da igual liberdade mínima e de iguais oportunidades justas impossibilitariam o surgimento de desigualdades econômicas tão extremas. Já que os ganhos dos mais favorecidos seriam tais que melhorariam as expectativas dos menos favorecidos, a tendência seria a de que desigualdades fossem minimizadas e fossem menores do que as de sistemas que foram tolerados ao longo da história da humanidade. Casos como o da objeção acima, portanto, dificilmente ocorreriam.¹⁰⁸

Rawls passa então a apresentar o raciocínio que levaria as partes ao princípio da utilidade média e afirma que, caso ele fosse sólido, tornaria desnecessária qualquer comparação entre alternativas na posição original, já que seria a única alternativa razoável.¹⁰⁹

Imaginemos uma pessoa a quem são apresentadas várias sociedades possíveis, com diferentes políticas públicas, e que tem o dever de definir em qual destas sociedades preferiria viver. Se a pessoa soubesse seus talentos e que posição ocuparia na sociedade, poderia fazer uma escolha em benefício próprio. Mas, caso sigamos o raciocínio da posição original e definamos que ela não possui tais informações, ela poderia realizar o seguinte raciocínio: imaginar que ela possui a mesma chance de ser qualquer membro da sociedade, com sua concepção de bem

¹⁰⁷ Ibid., pp. 135-136.

¹⁰⁸ Ibid., pp. 136-137.

¹⁰⁹ Ibid., p. 141.

e talentos. Ela naturalmente somaria a utilidade de cada membro e dividiria o total pelo número de pessoas, chegando à sociedade em que o membro médio mais satisfaz suas preferências e, conseqüentemente, ao princípio da utilidade média.¹¹⁰

Para o raciocínio acima desenvolvido, não é necessário encarar “utilidade” como uma propriedade mental, como o prazer, bastando utilizar uma medida cardinal de satisfação de preferências, como os axiomas von Neumann-Morgenstern desenvolvidos no capítulo anterior desta monografia.¹¹¹

Esse raciocínio pressupõe que as partes na posição original não possuem qualquer aversão especial a riscos e utilizam o princípio da razão insuficiente ao calcular probabilidades.¹¹² O princípio da razão insuficiente, defendido por Thomas Bayes e Pierre Laplace, determina que, não havendo razões para atribuir quaisquer probabilidades específicas para diferentes possibilidades, deve ser adotada a equiprobabilidade.

A seção seguinte de *Uma Teoria de Justiça* foca em apontar alguns problemas com o princípio da utilidade média. Rawls começa questionando a utilização do princípio da razão insuficiente e a ideia de que as pessoas na posição original têm a mesma chance de se tornar qualquer membro da sociedade. Ele afirma que essa utilização não é sustentada por qualquer característica da sociedade, pois as partes não possuem acesso a tais características sob o véu da ignorância. É defendido que as partes na posição original desconsiderariam probabilidades atingidas com base apenas neste princípio. Julgamentos probabilísticos em tal posição deveriam possuir uma base objetiva em fatos particulares ou, pelo menos, em crenças razoáveis.¹¹³

Isso se deve ao fato de o acordo na posição original ser extremamente importante e as partes desejarem tomar uma decisão que pareça responsável a seus descendentes, que serão afetados por ela. Esses fatos diminuem a disposição das pessoas de tomar riscos visando ganhos maiores e, sendo os dois princípios de justiça outra opção, parece que a posição original os favoreceria.

¹¹⁰ Ibid., pp. 141-143.

¹¹¹ Ibid., pp. 143-144.

¹¹² Ibid., p. 143.

¹¹³ Ibid., pp. 145-146.

A segunda dificuldade apontada é uma que Rawls afirma não poder lidar com profundidade em “Uma Teoria de Justiça”, mas que ainda assim merece menção. Na última etapa do raciocínio a favor do princípio da utilidade média, as pessoas devem escolher entre sociedades em que elas possuem a mesma chance de ser qualquer um dos cidadãos, com suas utilidades, interesses e *rankings* de preferências.

Isso claramente pressupõe a possibilidade de comparações interpessoais de utilidade, mas, ainda que se deixe de lado as dificuldades desta ideia, resta outro problema. As partes na posição original são conceitualizadas como não possuindo objetivos próprios, mas simplesmente adotando aqueles da pessoa cujo lugar elas tomariam na sociedade. Já que não há nenhum esquema de objetivos que determina suas estimativas, ele não possuiria “a unidade necessária”.¹¹⁴

É feita uma distinção entre a avaliação de situações objetivas e a avaliação de aspectos de uma pessoa, como suas habilidades, traços de caráter e sistema de objetivos.

Rawls afirma que é simples julgar, do nosso ponto de vista, a situação de outro indivíduo devido a sua posição social, riqueza, expectativa de bens primários ou outro critério qualquer. Simplesmente nos imaginamos em seu lugar, com nosso caráter e preferências, e vemos como nossos objetivos seriam afetados por tais circunstâncias.

Rawls afirma que ainda é possível ir além, nos imaginando também com algumas características e objetivos da pessoa cuja situação estamos avaliando. Sabendo nosso plano de vida, checamos se tais características e objetivos seriam úteis para cumpri-lo. Em caso positivo, poderíamos até tentar cultivá-los.

Contudo, ao julgar a forma de vida e sistema de fins últimos de alguém, o argumento do contrato exige que utilizemos nossos fins como critério. O valor para nós do modo de vida de outra pessoa não é o valor que tal modo de vida possui para ela.¹¹⁵

Além disso, as circunstâncias de justiça implicam que esses valores variam muito, ainda que certos bens básicos sejam do interesse de todos. Os fins de

¹¹⁴ Ibid., p. 150.

¹¹⁵ Ibid., p. 151.

diferentes pessoas muitas vezes são opostos e não sujeitos a uma medida comum. O valor das circunstâncias de uma pessoa para nós, portanto, não é o valor dessas circunstâncias para ela e, portanto, o último passo do argumento a favor do princípio da utilidade média não pode ser realizado.¹¹⁶

Posteriormente, a mesma dificuldade é formulada de uma maneira um pouco distinta. O raciocínio a favor do princípio da utilidade média precisa definir alguma forma de expectativa unificada. Isso pode ser feito se as partes concordarem com certas regras para fundamentar a comparação interpessoal de fins últimos. Essas regras podem advir, por exemplo, de leis da psicologia que definam a satisfação de pessoas dadas certas questões como intensidade de desejos, habilidades naturais e atributos físicos, bens públicos e privados possuídos etc.

Indivíduos iguais em tais parâmetros teriam a mesma satisfação e seria possível, conseqüentemente, determinar uma satisfação média. Todos teriam a mesma “função de utilidade profunda” (tradução livre de “*deep utility function*”), como Rawls se refere, e, tendo todos as mesmas expectativas unidas, se atingiria o princípio da utilidade média.

Esse raciocínio, contudo, pressupõe certa noção de pessoa que Rawls questiona. Tais pessoas não teriam interesses de última ordem ou fins fundamentais por meio dos quais decidem que tipo de pessoa ser. Elas não teriam um “caráter de vontade” próprio. Seriam, segundo Rawls, “pessoas-vazias” (tradução livre de “*bare-persons*”), pois estariam dispostas a aceitar como sua concepção de bem o que quer que tais regras de comparação definissem como sendo sua concepção de bem devido a suas circunstâncias.¹¹⁷

Foi assumido no começo da argumentação a favor dos dois princípios de justiça, contudo, que as pessoas têm um caráter e vontade determinados, mesmo que se esqueçam deles na posição original. São “pessoas-determinadas”, com interesses de última ordem e fins fundamentais pela referência aos quais decidem que tipo de vida e objetivos subordinados lhes são aceitáveis. Elas buscam proteger

¹¹⁶ Ibid.

¹¹⁷ Ibid., pp. 151-152.

esses interesses e fins e, como as liberdades básicas dos dois princípios de justiça os asseguram, elas devem escolhê-los ao invés do princípio da utilidade média.¹¹⁸

Como último ponto de sua argumentação a favor de seus princípios de justiça, Rawls menciona algumas razões favoráveis a esses princípios quando comparados com o princípio da utilidade média.

A primeira destas se refere aos esforços do comprometimento mencionados anteriormente. As partes na posição original assumem que seus esforços não são em vão e que a concepção de justiça a ser definida será efetivamente seguida pela grande maioria das pessoas, considerando-se os fatos da psicologia humana.

Elas não podem, portanto, entrar em contratos que podem ter consequências inaceitáveis e evitarão aqueles a que somente podem aderir com grandes dificuldades. Uma vez que o acordo original é final e perpétuo, não há segundas chances para se escolher outros princípios. Ao entrar em um acordo, é preciso estar disposto a cumpri-lo mesmo que o pior resultado possível dele se revele. Caso contrário, o acordo não foi feito de boa-fé.

Isso seria um ponto a favor dos princípios de justiça de Rawls. Sendo eles escolhidos, cada pessoa tem um conjunto de liberdades mínimas assegurado, enquanto que no princípio da utilidade média é possível que uma parcela da população tenha de arcar com sacrifícios grandes a sua liberdade em nome de um aumento na utilidade média, o que pode ser mais do que as leis da psicologia humana permitem que se cumpra de bom grado.¹¹⁹

Outra consideração invoca também a condição de publicidade. Rawls considera um grande ponto a favor de uma concepção de justiça o fato de ela poder criar seu próprio apoio ao longo do tempo. Em outras palavras, é preferível que o conhecimento público de que a estrutura básica da sociedade satisfaz uma concepção de justiça desenvolva nos cidadãos um desejo de agir de acordo com esta.

Que concepção de justiça melhor cumprirá tal requisito depende das regras da psicologia humana. Todavia, à primeira vista pode ser percebido que o princípio

¹¹⁸ Ibid., p. 152.

¹¹⁹ Ibid., pp. 153-154.

da utilidade média exige uma grande capacidade de se identificar com a posição dos demais membros da sociedade para que seja seguido de boa vontade, já que algumas pessoas teriam de aceitar sacrifícios a suas liberdades e interesses em nome de uma maior utilidade para terceiros, enquanto os dois princípios de justiça não exigem isso de maneira tão forte, tendo-se em vista que as desigualdades econômicas seriam sempre tais que melhorariam a posição dos menos favorecidos. Como é difícil esperar uma capacidade de identificação com o próximo tão grande da maioria das pessoas, os dois princípios de justiça parecem ser mais capazes de desenvolver nos cidadãos um desejo de cumpri-los.¹²⁰

Além disso, o reconhecimento público dos dois princípios oferece mais suporte para o autorrespeito dos membros da sociedade, o que aumenta a cooperação social. É racional para as pessoas na posição original buscar garantir seu autorrespeito, pois uma percepção de valor próprio é necessária para que as pessoas busquem sua concepção de bem e se sintam realizadas ao atingi-la. O autorrespeito não é parte de um plano de vida, mas sim a ideia de que vale a pena seguir um plano destes.

O autorrespeito de cada pessoa muitas vezes depende do respeito que os outros possuem por ela, e é possível assumir que aqueles que possuem autorrespeito são mais inclinados a respeitar os demais e vice-versa. Desrespeito aos outros leva ao desrespeito a si mesmo, e o autorrespeito leva ao respeito dos demais, o que fortalece o autorrespeito de cada um.

É, portanto, desejável que uma concepção de justiça expresse publicamente o respeito dos homens uns pelos outros, e os dois princípios de Rawls o fazem. Como cada membro da sociedade teria liberdades mínimas iguais garantidas e todas as desigualdades existiriam para o bem dos menos favorecidos, todos se considerariam uma parte valiosa da sociedade.¹²¹

Rawls reconhece que não faz parte da doutrina utilitarista a ideia de que algumas pessoas teriam menos vantagens por seu bem-estar valer menos que o dos outros. Contudo, ao se levar em conta as características da psicologia humana,

¹²⁰ Ibid., pp. 154-155.

¹²¹ Ibid., pp. 155-156.

ele acredita que aqueles com menos vantagens teriam seu autorrespeito abalado, pois as vantagens dos outros não contribuiriam para o bem de todos.¹²²

Os dois princípios são equivalentes, na verdade, a uma tentativa de distribuir as habilidades naturais de cada cidadão, como se fossem um bem coletivo.¹²³ A abstenção dos mais afortunados em se aproveitar de seus dons de maneiras que não beneficiem os demais incorpora o respeito mútuo entre os cidadãos na estrutura básica da sociedade.

Rawls afirma que essa argumentação demonstra uma conexão entre seus princípios de justiça e uma das formulações do imperativo categórico de Kant: “trate os outros seres humanos não como meios, mas como fins em si mesmos”.¹²⁴ De um ponto de vista contratualista, ele acredita que isso poderia ser representado por um acordo na posição original, em que todas as partes são iguais e se encaram como fins em si mesmas.

Agora, tentando encontrar princípios substanciais que realizem tal ideia, ele chega à conclusão de que seus princípios de justiça o fazem. Isso porque todos têm liberdades mínimas asseguradas e o princípio da diferença faz com que todos sejam tratados como fins, pois os ganhos que não sejam para o benefício de todos são abandonados. *A contrario sensu*, fazer com que aqueles que já possuem menos sofram um ônus para o ganho de outros seria o mesmo que tratá-los apenas como meios.

Se ainda for suposto que a cooperação social entre aqueles que se respeitam mutuamente é mais efetiva e harmoniosa, o nível de expectativas geral pode acabar sendo maior do que se o princípio da utilidade média fosse adotado.¹²⁵

É apontado que um utilitarista poderia responder a tal crítica afirmando que tais questões já são levadas em conta pelo princípio da utilidade. Ou seja, caso as liberdades básicas fossem necessárias para o autorrespeito e o autorrespeito maximizasse a utilidade, o princípio da utilidade média advogaria a favor de tais liberdades.

¹²² Ibid., p. 157.

¹²³ Ibid., p. 156.

¹²⁴ KANT, Immanuel. **Metafísica dos Costumes**. São Paulo: Editora Vozes, 2013.

¹²⁵ RAWLS, John Bordley. **A Theory of Justice**. Edição revisada. Cambridge, Massachusetts: Harvard University Press, 1999, pp. 156-157.

Deve-se levar em conta, porém, a necessária publicidade da concepção de justiça adotada. Não é possível escolher um princípio utilitarista na posição original e, posteriormente, aplicar princípios não utilitaristas na estrutura básica da sociedade, pois o princípio utilitarista não estaria sendo publicamente reconhecido. Dada tal estipulação, e assumindo que um princípio utilitarista aplicado publicamente diminuiria o bem-estar das pessoas, os dois princípios de justiça se revelam uma opção mais atraente para as partes na posição original.¹²⁶

Estes são, em resumo, os principais pontos da argumentação de Rawls a favor de seus dois princípios de justiça. Ele reconhece que seu raciocínio se baseia, em muitos pontos, em intuições e não é uma prova definitiva. Ainda assim, levando-se em consideração suas críticas ao princípio da utilidade média (o fato de ele não encarar as pessoas como tendo fins últimos próprios e se utilizar de distribuições probabilísticas não fundamentadas em fatos) e suas colocações a favor de seus dois princípios (que eles seriam mais facilmente cumpridos pelos menos favorecidos, contribuiriam para a estabilidade da sociedade e poderiam aumentar o bem-estar de todos ao garantir o autorrespeito dos cidadãos), Rawls acredita que a balança de considerações pode pender claramente para o lado de seus princípios.¹²⁷

¹²⁶ Ibid., p. 158.

¹²⁷ Ibid., p. 159.

3. CRÍTICAS À TEORIA DE JUSTIÇA DE RAWLS

No terceiro capítulo desta monografia, será realizada uma crítica à teoria de justiça de Rawls e dos argumentos em seu favor. Esta crítica será dividida em duas seções, para fins de clareza. Na primeira seção, será feita uma reconstrução dos argumentos de John Harsanyi, importante economista húngaro, contra a teoria de justiça de Rawls em um famoso artigo seu: “Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls’s Theory”.¹²⁸ Esses argumentos servem como uma crítica inicial à teoria de Rawls e fundamentam críticas posteriores desta monografia.

A segunda seção consistirá em argumentos de autoria própria contra a teoria de justiça de Rawls. Estes argumentos serão em parte baseados no artigo de Harsanyi, mas não se restringirão a ele. As partes mais intuitivas do raciocínio de Rawls serão criticadas, assim como sua tendência de retirar algumas questões espinhosas do escopo de sua teoria de justiça. Além disso, alguns dos pressupostos mais básicos de sua teoria de justiça serão questionados: a sua concepção de pessoa e sua exigência de que a concepção de justiça seja uma concepção pública.

Para fins de abreviação, todas as vezes que o “princípio da utilidade” for mencionado neste capítulo, a referência será ao princípio da maximização da utilidade média. Do mesmo modo, quando houver menção à ideia de “maximização da utilidade”, refere-se à utilidade média da população. Isto se justifica pelo fato de não ser do escopo desta monografia abordar o princípio da maximização da utilidade absoluta, conforme justificado no capítulo anterior.

3.1. CRÍTICAS DE HARSANYI

Harsanyi inicia seu artigo com uma breve introdução em que explica os objetivos de Rawls em “Uma Teoria de Justiça”, define o conceito de posição original e aponta sua discordância principal com o raciocínio de Rawls: o fato de Rawls

¹²⁸ HARSANYI, John Charles. **Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls’s Theory**. The American Political Science Review, vol. 69, no. 2. Washington: American Political Science Association, 1975.

afirmar que as partes na posição original se utilizariam da regra *maximin* de escolha sob incerteza para definir os princípios de justiça de sua sociedade.

Harsanyi afirma que a regra *maximin* é um critério irracional de decisão, por levar em muitos casos a consequências inaceitáveis. Ele acredita que as partes na posição original, sendo racionais, se utilizariam da regra de maximização da utilidade esperada, da escola de pensamento Bayesiana, na posição original.¹²⁹

Podemos recriar o raciocínio de Harsanyi da seguinte maneira: as partes na posição original, possuindo preferências próprias e sendo racionais e mutuamente desinteressadas, buscariam os princípios de justiça que melhor satisfariam suas preferências ou, em outras palavras, que melhor maximizassem sua utilidade. Isto é sustentado pela noção de racionalidade que foi desenvolvida no capítulo 1 desta monografia.

Na posição original, nenhuma das partes teria como saber que membro da sociedade seria de fato. Além disso, não haveria qualquer motivo para achar mais provável ser algum membro específico ou fazer parte de certa classe social. Portanto, seria atribuída a mesma probabilidade à chance de ser qualquer membro da sociedade. Sendo este o caso, seria buscado o princípio que maximizaria os prospectos de satisfação de preferências de cada cidadão. Este é, por definição, o princípio da utilidade média, que seria escolhido.¹³⁰ Conforme será desenvolvido na segunda seção deste capítulo, entendo que este raciocínio está essencialmente correto.

Após sua introdução, Harsanyi lista algumas implicações paradoxais da regra *maximin* de escolha sob incerteza, que Rawls utiliza ao desenvolver seu princípio da diferença. Esta regra determina que deve ser feita a escolha cujo pior resultado possível seja melhor ou igual ao pior resultado possível das outras escolhas disponíveis. Em outras palavras, o mínimo deve ser maximizado.

Ele se utiliza de um experimento de pensamento para demonstrar como essa regra pode levar a conclusões inaceitáveis. O leitor deve imaginar que é um morador

¹²⁹ Ibid., p. 595.

¹³⁰ Ibid., p. 598.

de Nova York e recebe duas ofertas de emprego: um emprego tedioso e com salário baixo em Nova York e um emprego interessante e bem remunerado em Chicago.

Contudo, para conseguir o emprego de Chicago, seria necessário ir de avião até a cidade, pois o trabalho começaria logo no dia seguinte. Os resultados possíveis das escolhas do leitor podem então ser representados da seguinte maneira:

	O avião para Chicago cai	O avião para Chicago não cai
Escolher o emprego de Nova York	Você terá um emprego ruim, mas estará vivo	Você terá um emprego ruim, mas estará vivo
Escolher o emprego de Chicago	Você morrerá	Você terá um excelente emprego e estará vivo

Como a regra *maximin* determina que se escolha a ação cujo pior resultado seja melhor ou igual ao pior resultado possível das outras escolhas, você deveria escolher ficar com o emprego de Nova Iorque, *independentemente de quão improvável um acidente de avião seja*.¹³¹

Não é difícil utilizar este mesmo raciocínio para demonstrar como este critério de escolha levaria a resultados absurdos constantemente. Não poderíamos atravessar a rua, pois poderíamos ser atropelados. Não poderíamos sair na chuva, pois poderíamos ser atingidos por um raio. Não poderíamos comer para além do mínimo necessário para continuarmos vivos, pois poderíamos morrer devido a uma intoxicação alimentar. Não poderíamos alimentar nossos bebês para além do mínimo para mantê-los vivos, pois eles poderiam se engasgar e morrer. Na verdade, não poderíamos nem ter filhos, pois eles poderiam um dia crescer não gostando de nós e nos matar enquanto dormimos. Como apontado por Ken Binmore,¹³² todos esses raciocínios são exemplos claros de pensamento paranoico, em que uma possibilidade remota de resultados negativos é muito mais valorizada que a possibilidade maior de resultados positivos.

¹³¹ Ibid., p. 595.

¹³² BINMORE, Kenneth George. **Game Theory: A Very Short Introduction**. Oxford: Oxford University Press, 2007, p. 31.

Verdade seja dita, Rawls concorda que a regra *maximin* é um critério de escolha irracional na grande maioria das situações, então simplesmente apontar situações em que ela leva a resultados inaceitáveis não é uma crítica válida de sua teoria de justiça.

Harsanyi afirma, contudo, que as únicas situações em que a regra *maximin* leva a resultados aceitáveis e racionais em políticas públicas são situações em que ela levaria a resultados iguais ou muito similares aos determinados pelo princípio da utilidade. Por outro lado, quando os dois princípios apontam para resultados bastante díspares, a regra *maximin* sempre leva a resultados menos sensatos que o princípio da utilidade. O autor questiona, então, por que se deveria adotar a regra *maximin* como princípio ao invés do princípio da utilidade. Considero este questionamento válido, e o retomarei na segunda seção quando for criticar os requisitos que Rawls afirma serem suficientes para justificar o uso da regra *maximin* na posição original.¹³³

Em sequência, Harsanyi discorre sobre os resultados da adoção do princípio da diferença em uma dada sociedade, devido ao uso da regra *maximin* na posição original, listando algumas consequências inaceitáveis. Ele começa imaginando uma sociedade hipotética composta por um médico e dois indivíduos com pneumonia. O indivíduo A é relativamente saudável, desconsiderando-se a pneumonia, enquanto o indivíduo B é também um paciente em estágio terminal de câncer. Há apenas antibióticos suficientes na sociedade para tratar um dos pacientes.

Segundo o princípio da diferença de Rawls, afirma Harsanyi, o tratamento deveria ir para o paciente com câncer, ainda que isso aumentasse muito pouco sua expectativa de vida e o outro indivíduo perdesse toda uma vida que tinha pela frente. Para Harsanyi, essa conclusão não vai apenas contra a ética utilitarista, mas também contra o senso comum.¹³⁴

Como segundo exemplo de consequência inaceitável, Harsanyi imagina uma sociedade hipotética consistente em dois indivíduos. Ambos têm suas necessidades materiais satisfeitas, mas a sociedade ainda possui recursos extras para investir em

¹³³ HARSANYI, John Charles. **Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory.** The American Political Science Review, vol. 69, no. 2. Washington: American Political Science Association, 1975, pp. 595-596.

¹³⁴ Ibid., p. 596.

educação. O indivíduo A é excepcional em matemática e os recursos poderiam ser utilizados para lhe fornecer ensino superior de qualidade. O indivíduo B possui deficiências mentais e a mesma quantidade de recursos poderia ser utilizada para que ele aprendesse coisas simples, como amarrar os seus sapatos.

Harsanyi aponta que, segundo o princípio da diferença, não é apenas o caso que todos os recursos da sociedade deveriam ser destinados ao indivíduo B, mas também que o aprendizado do indivíduo B possui *absoluta prioridade*. Em outras palavras, não seria possível destinar qualquer recurso à educação do indivíduo A enquanto o indivíduo B fosse menos educado que o indivíduo A.

Ainda que a melhoria da educação do indivíduo B afetasse a satisfação de suas preferências de uma forma mínima, isso justificaria privar de uma educação superior todas as mentes mais brilhantes de uma sociedade, que teriam seus interesses afetados de forma muito maior.¹³⁵

Harsanyi comenta que Rawls está ciente dessas consequências, mas as considera de acordo com a máxima kantiana de “tratar os outros não como meios, mas como fins em si mesmos”, pois dar prioridade à educação de A por questões utilitaristas seria o mesmo que tratar B como um meio. Harsanyi questiona esse raciocínio, alegando que impor grandes sacrifícios a A em nome de pequenas melhorias na condição de B também poderia ser interpretado como um modo de tratar A apenas como meio.¹³⁶

Em seu *paper* “Some Reasons for the Maximin Criterion”, Rawls responde aos contraexemplos de Harsanyi alegando que a regra *maximin* é “um princípio macro, não micro”,¹³⁷ afirmando que ela não deve ser aplicada em situações de pequena escala, como o tratamento por parte de um médico de seus pacientes.

No *post scriptum* de seu artigo, Harsanyi responde a este contra-argumento de duas formas. Primeiramente, ele questiona a ideia de Rawls de que haveria uma diferença especial entre “princípios micro e macro”. Rawls não fornece qualquer razão pela qual princípios de justiça diferentes deveriam ser utilizados em casos que

¹³⁵ Ibid., pp. 596-597.

¹³⁶ Ibid., p. 597.

¹³⁷ RAWLS, John Bordley. **Some Reasons for the Maximin Criterion**. In: FREEMAN, Samuel (org.). **John Rawls: Collected Papers**. Cambridge, Massachusetts: Harvard University Press, 1999.

envolvem poucas pessoas, apenas alegando que este é o caso. Harsanyi considera muito questionável a ideia de que princípios de justiça dependam da escala de uma situação, pois isso implicaria que há um número “X” de pessoas afetadas tal que, de “X” para cima, os princípios de justiça adequados são distintos dos de “X” para baixo.¹³⁸

A segunda resposta de Harsanyi é a de que seus exemplos são facilmente reformuláveis de maneira tal que uma grande quantidade de pessoas seja afetada. O exemplo do médico e dos dois pacientes pode simplesmente ser substituído pela escolha de um Estado entre destinar certa quantidade de recursos a pacientes com expectativa de vida muito baixa ou investir a mesma quantidade em pacientes que serão mais beneficiados pelos tratamentos.¹³⁹ Seria possível criticar esta resposta de Harsanyi afirmando que essas situações não dizem respeito à estrutura básica de uma sociedade, que é o âmbito de aplicação da teoria de justiça de Rawls. Lidarei com este argumento no fim da segunda seção, quando falar sobre a tendência de Rawls e seus defensores de retirar do escopo de sua teoria questões espinhosas.

Em outra seção de seu artigo, Harsanyi aborda a crítica de Rawls ao uso de probabilidades na posição original.¹⁴⁰ Em seu raciocínio a favor do princípio da utilidade média, Harsanyi se utiliza da regra da equiprobabilidade para escolhas sob incerteza, atribuindo a mesma probabilidade à possibilidade de ser qualquer membro de uma dada sociedade. Rawls afirma que, na posição original, as partes não se utilizariam de distribuições probabilísticas sem fundamentos em fatos empíricos, o que impossibilitaria este passo do argumento de Harsanyi.

Harsanyi afirma que diversos autores já trataram da necessidade de se utilizar de probabilidades subjetivas em escolhas¹⁴¹ e menciona dois argumentos simples a favor delas: a) a única alternativa às probabilidades subjetivas seria um critério de escolha da família *maximin*, que pode levar a decisões irracionais em casos relevantes; e b) a teoria de decisão bayesiana mostra que qualquer pessoa que aja

¹³⁸ HARSANYI, John Charles. **Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory**. The American Political Science Review, vol. 69, no. 2. Washington: American Political Science Association, 1975, p. 605.

¹³⁹ Ibid.

¹⁴⁰ Ibid., pp. 598-600.

¹⁴¹ SAVAGE, Leonard Jimmie. **The Foundations of Statistics**. Nova York: John Wiley and Sons, 1954.

de acordo com alguns postulados de racionalidade bastante convincentes não pode deixar de agir *como se* utilizasse probabilidades subjetivas em suas decisões.

Ele dá o exemplo de uma oferta de aposta: um amigo lhe dá a opção de escolha entre receber R\$ 100,00 caso o candidato A vença as eleições e R\$ 0,00 caso ele perca ou receber R\$ 100,00 caso ele perca e receber R\$ 0,00 caso ele ganhe. Se você escolhe alguma das opções, estaria dando uma probabilidade subjetiva à possibilidade do candidato ganhar ou perder de pelo menos 50%.

De modo similar, até mesmo a parte na posição original de Rawls que, segundo ele, seria avessa à utilização de probabilidades subjetivas, ao escolher a regra *maximin*, estaria atribuindo, ao menos implicitamente, uma probabilidade de 100% (ou quase isso) à possibilidade de ser o membro da sociedade na posição menos vantajosa.¹⁴²

Este ponto do artigo de Harsanyi me parece justificar suficientemente a legitimidade do uso de probabilidades subjetivas na posição original. A atribuição de igual probabilidade à chance de ser qualquer membro da sociedade, contudo, possui pressupostos éticos não mencionados no artigo. Tratarei deste ponto na segunda seção.

Na sequência, Harsanyi questiona a tese de Rawls segundo a qual a escolha do princípio da diferença na posição original seria mais facilmente justificável às gerações posteriores. Ele afirma que, como o princípio da diferença dá absoluta prioridade aos membros da sociedade na posição menos vantajosa, todos os outros membros da sociedade que tivessem seus interesses sacrificados em nome desta parcela teriam legitimidade em reclamar, não sendo simples justificar a adoção do princípio para essas pessoas.¹⁴³

Seria difícil, por exemplo, justificar à grande maioria da população a falta de acesso ao ensino superior em nome de ensino pouco efetivo a pessoas com deficiências mentais ou justificar a milhões de pessoas a falta de remédios para

¹⁴² HARSANYI, John Charles. **Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory.** The American Political Science Review, vol. 69, no. 2. Washington: American Political Science Association, 1975, p. 599.

¹⁴³ Ibid., pp. 599-600.

doenças moderadas devido a investimentos massivos em tratamentos paliativos para pacientes em estado terminal.

Embora seja possível alegar que os exemplos de Harsanyi não dizem respeito à estrutura básica da sociedade (alegação esta de que tratarei na segunda seção), entendo que o seu ponto ainda é válido, pois a possível negação de muitos bens à maioria da população devido à valorização excessiva de um estado de coisas em que os menos favorecidos estejam em situação melhor permite que a maioria da população questione a distribuição de bens da sociedade.

Posteriormente, o autor aponta uma má compreensão de Rawls sobre o que as utilidades von Neumann-Morgenstern representam.¹⁴⁴ Rawls alega que elas representariam a atitude de uma pessoa perante o risco, o que seria desconhecido sob o véu da ignorância e não possui lugar na teorização ética.¹⁴⁵

Harsanyi afirma que esta confusão é bastante comum. Embora as funções de utilidade von Neumann-Morgenstern de um indivíduo se baseiem no modo como ele age perante situações de risco, o que elas de fato representam é quanta *importância* este indivíduo atribui aos diferentes resultados possíveis de suas escolhas.

Para ilustrar esta distinção, trago um exemplo de autoria própria. Imaginemos dois indivíduos, João e Maria, que devem ordenar diversas apostas apresentadas para si, em ordem de preferência em participar nelas. Imaginemos também que em seu *ranking* de preferências, ambos atribuam a mesma utilidade cardinal ao ato de participar em cada aposta.

Isto não implica, de maneira alguma, que ambos possuem a mesma atitude perante o risco. É completamente possível que João obtenha prazer ao entrar em situações de risco, enquanto Maria se sente apreensiva e busca evitá-las. Contudo, pode ser o caso que Maria valorize mais os presentes das apostas mais arriscadas do que João os valoriza. Deste modo, teríamos dois indivíduos com atitudes completamente distintas perante situações de risco, mas com a mesma distância entre o ato de participar em cada aposta em suas funções de utilidade cardinal.

¹⁴⁴ Ibid., p. 600.

¹⁴⁵ RAWLS, John Bordley. **A Theory of Justice**. Edição original. Cambridge, Massachusetts: Harvard University Press, 1971, pp. 172 e 323.

Podemos realizar um raciocínio muito similar para o caso de eles atribuírem utilidades cardinais diferentes para as apostas. Novamente, isto não significa que eles possuem atitudes diferentes perante o risco: ambos podem ter a mesma atitude perante o risco, mas atribuir utilidade cardinal diferente às apostas por valorizarem de modo distinto os possíveis resultados destas.

Funções de utilidade von Neumann-Morgenstern, portanto, não formalizam atitudes perante o risco, mas sim a importância que cada pessoa atribui aos diferentes resultados de ações, o que pertence ao campo da teorização ética.

Em outra seção de seu artigo relevante para esta monografia,¹⁴⁶ Harsanyi aborda a crítica de Rawls ao uso de comparações interpessoais de utilidade para definir princípios de justiça.¹⁴⁷ Rawls questiona se a ideia de avaliar a situação de outra pessoa *do ponto de vista dela* é algo possível.

Harsanyi defende esta possibilidade, alegando que comparações interpessoais de utilidade são possíveis porque “*as preferências e funções de utilidade de todos os indivíduos humanos são governadas pelas mesmas leis psicológicas básicas*”.^{148 149}

O que define a função de utilidade de uma pessoa é a sua genética e as experiências que teve ao longo de sua vida. Ainda que estes fatores sejam diferentes para todas as pessoas, as leis que determinam como eles afetam as preferências e as funções de utilidade são universais. Assim, comparações interpessoais de utilidade podem ser encaradas como comparações *intrapessoais*: entre a função de utilidade de uma pessoa “X” no momento atual e a função de utilidade que essa mesma pessoa teria caso possuísse a mesma genética e tivesse passado pelas mesmas experiências de outra pessoa “Y”.

¹⁴⁶ HARSANYI, John Charles. **Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory.** The American Political Science Review, vol. 69, no. 2. Washington: American Political Science Association, 1975, pp. 600-601.

¹⁴⁷ RAWLS, John Bordley. **A Theory of Justice.** Edição original. Cambridge, Massachusetts: Harvard University Press, 1971, p. 173.

¹⁴⁸ HARSANYI, John Charles. **Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory.** The American Political Science Review, vol. 69, no. 2. Washington: American Political Science Association, 1975, p. 600.

¹⁴⁹ HARSANYI, John Charles. **Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility.** Journal of Political Economy, vol. 63, no. 4. Chicago: University of Chicago Press, 1955.

Harsanyi admite que, na prática, temos informações limitadas e não conhecemos estas leis em detalhe, tampouco os genes e a história de vida de todas as pessoas. Contudo, o ponto central é o de que é possível tentar realizar comparações interpessoais aproximadas, e isto é algo que fazemos constantemente no dia a dia, por exemplo, ao imaginar qual presente um de nossos amigos mais gostaria de receber em seu aniversário.

As comparações interpessoais, portanto, são não apenas possíveis, mas também relevantes para a teorização moral, pois são realizadas sempre que um governo pretende definir aonde deve alocar recursos limitados.¹⁵⁰ Isto se reflete em políticas públicas de saúde, educação e segurança, para citar apenas alguns casos. Concordo com esta visão de Harsanyi e irei retomá-la ao criticar na segunda seção a concepção de pessoa utilizada por Rawls em sua argumentação.

Um último ponto do artigo de Harsanyi a ser tratado nesta monografia é a sua abordagem do argumento de Rawls a favor de seus princípios de justiça a partir da ideia de estabilidade de uma sociedade justa.¹⁵¹ Segundo Harsanyi, Rawls sustenta que uma sociedade em que os cidadãos possuam um forte senso de justiça e cujas instituições sejam justas seria mais estável que outras sociedades¹⁵² e, dentre as sociedades com tais características, aquela que seguisse seus princípios de justiça seria mais estável que aquela que seguisse o utilitarismo.¹⁵³

Harsanyi alega que uma sociedade assim não é apenas uma versão melhorada das sociedades conhecidas atualmente; trata-se de uma sociedade muito diferente de qualquer sociedade conhecida por cientistas políticos, historiadores ou demais observadores competentes. Nela, nem cidadãos tampouco legisladores agem por interesses egoístas, e sim por seu senso de justiça. É como se fosse o exato contrário da noção de sociedade desenvolvida por Anthony Downs em seus estudos sobre a democracia.¹⁵⁴

¹⁵⁰ HARSANYI, John Charles. **Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory**. The American Political Science Review, vol. 69, no. 2. Washington: American Political Science Association, 1975, p. 601.

¹⁵¹ Ibid., pp. 603-604.

¹⁵² RAWLS, John Bordley. **A Theory of Justice**. Edição original. Cambridge, Massachusetts: Harvard University Press, 1971, pp. 490-504.

¹⁵³ Ibid., p. 498.

¹⁵⁴ DOWNS, Anthony. **An Economic Theory of Democracy**. Nova York: Harper and Brothers, 1957.

Harsanyi reconhece que a teoria de Downs não representa perfeitamente nossas sociedades, havendo situações em que cidadãos e políticos agem motivados por razões altruísticas. Ela é, contudo, um esquema aproximado que, frequentemente, prediz e explica as ações de políticos e eleitores.

Isto colocaria em questão a validade do argumento de Rawls, segundo Harsanyi. O fato de a maioria das democracias funcionarem aproximadamente conforme aquele esquema pode ser evidência de que sociedades em que um senso de justiça seja onipresente seriam *instáveis*. Uma análise histórica demonstraria, de acordo com o autor, que sociedades só conseguem manter tal estado por um tempo curto, por exemplo, logo após revoluções, e que esse estado pode ter consequências bastante negativas, como uma espécie de fanatismo e intolerância.¹⁵⁵

Harsanyi advoga que sociedades saudáveis devem possuir um misto de motivações egoísticas e altruísticas de seus membros. Embora setores específicos da sociedade não tenham uma visão clara sobre o bem comum, possuem uma visão clara de seus interesses. É possível que milionários altruístas lutando pelos direitos dos trabalhadores de minas não saibam as reais necessidades destes trabalhadores e causem mais problemas no geral do que se não houvessem agido tão altruisticamente. Uma sociedade em que egoísmo e altruísmo tenham seu lugar seria mais estável e feliz que uma sociedade totalmente altruísta.¹⁵⁶

Embora este argumento de Harsanyi seja bastante interessante e digno de menção, não o considero muito convincente. Rawls de fato afirma que uma sociedade em que os cidadãos possuam um forte senso de justiça e cujas instituições sejam justas seria mais estável que aquelas em que este não fosse o caso, e de fato afirma que seus princípios de justiça gerariam a maior estabilidade. Isto não implica, contudo, que esta sociedade seria uma em que todos agem sempre motivados por seu senso de justiça e nunca por razões egoístas, assim como não implica que ela seria similar a uma sociedade estabelecida logo após uma revolução. Este ponto de Harsanyi, portanto, não será retomado na segunda seção.

¹⁵⁵ HARSANYI, John Charles. **Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory**. The American Political Science Review, vol. 69, no. 2. Washington: American Political Science Association, 1975, p. 604.

¹⁵⁶ Ibid.

3.2. CRÍTICAS AUTORAIS

Nesta seção desenvolvo uma série de críticas próprias à teoria de justiça de Rawls. Inicialmente, afirmo que, embora não esteja de acordo com todas as alegações de Harsanyi em seu artigo acima analisado, concordo com o seu raciocínio a favor do princípio da utilidade média na posição original: assim como ele, acredito que as partes na posição original atribuiriam a mesma probabilidade à chance de ser qualquer membro da sociedade e, sendo racionais (conforme o conceito de racionalidade desenvolvido no capítulo 1), escolheriam o princípio da utilidade média como princípio de justiça para maximizar o cumprimento de seus interesses.

Rawls sustenta que, com o cumprimento dos três requisitos mencionados no capítulo anterior,¹⁵⁷ as partes na posição original possuiriam motivação suficiente para escolher utilizando a regra *maximin*, buscando maximizar seus prospectos mínimos. Tenho críticas, contudo, a este raciocínio. Acredito que uma caracterização ideal das partes na posição original faria com que elas atribuíssem a mesma probabilidade à chance de ser qualquer membro da sociedade, não sendo avessas a quaisquer raciocínios probabilísticos. O primeiro requisito, portanto, não seria cumprido. Considero que os dois princípios subsequentes, por sua vez, não dão razões sólidas o suficiente para sustentar os princípios de justiça de Rawls ante o princípio da utilidade média.

Tratemos, inicialmente, do primeiro requisito. Este é o de que a posição original seja caracterizada de modo tal que, não havendo informações empíricas sobre a sua sociedade, as partes não se utilizariam de qualquer raciocínio probabilístico ao decidir os princípios de justiça para regulá-la.

Harsanyi afirma, ao contrário, que as partes, não vendo qualquer motivo para considerar mais provável ser qualquer membro da sociedade, atribuiriam a mesma probabilidade à possibilidade de ser cada membro da sociedade. Isso é uma aplicação do princípio da indiferença de LaPlace, que sugere atribuir a mesma

¹⁵⁷ RAWLS, John Bordley. **A Theory of Justice**. Edição revisada. Cambridge, Massachusetts: Harvard University Press, 1999, p. 134.

probabilidade a diferentes eventos quando não se tem informações para atribuir probabilidades diferentes.¹⁵⁸

Ambos estes raciocínios são passíveis de questionamento. Gerald Gaus e John Thrasher, em seu paper “Rational Choice and the Original Position: The (Many) Models of Rawls and Harsanyi”,¹⁵⁹ analisam ambos os raciocínios e apontam que nenhum deles pode ser justificado simplesmente pela teoria da escolha racional, havendo pressupostos morais na adoção de qualquer um.

Rawls afirma que, na posição original, as partes seriam céticas a atribuições subjetivas de probabilidade. Não há, contudo, nada na teoria Bayesiana de escolha sob incerteza que impeça a atribuição de probabilidades subjetivas a diferentes possibilidades antes que se tenha alguma base empírica para ajustar tais probabilidades.¹⁶⁰ E a teoria Bayesiana de escolha sob incerteza se tornou a maneira padrão de se encarar o modo como as pessoas fazem raciocínios em situações de incerteza a partir dos anos 50,^{161 162} justamente devido aos resultados irracionais a que regras como a *maximin* podem levar. O fato de Rawls rechaçar o uso de probabilidades na posição original é um pressuposto moral de seu raciocínio, tendo-se em vista que não pode ser depreendido apenas da teoria da escolha racional e da incerteza na posição original.

Harsanyi, por outro lado, em seu artigo crítico, afirma que pessoas racionais atribuiriam a mesma probabilidade na posição original à chance de ser qualquer membro da sociedade devido ao princípio da indiferença de LaPlace. O princípio da indiferença, contudo, não é, por si só, capaz de justificar tal atribuição de probabilidades. Assim como seria possível atribuir a mesma probabilidade à chance de ser qualquer membro da sociedade, também seria possível se perguntar “será que eu serei membro da classe menos privilegiada da sociedade?”, e o princípio da

¹⁵⁸ HARSANYI, John Charles. **Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory.** The American Political Science Review, vol. 69, no. 2. Washington: American Political Science Association, 1975, p. 598.

¹⁵⁹ GAUS, Gerald Francis; THRASHER, John. **Rational Choice and the Original Position: The (Many) Models of Rawls and Harsanyi.** In: HINTON, Timothy (org.). **The Original Position.** Cambridge: Cambridge University Press, 2015.

¹⁶⁰ Ibid., p. 56.

¹⁶¹ Ibid., p. 48.

¹⁶² HARSANYI, John Charles. **Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory.** The American Political Science Review, vol. 69, no. 2. Washington: American Political Science Association, 1975, p. 594.

indiferença determinaria que fosse atribuída uma probabilidade de 50% à possibilidade de ser um membro de tal classe e uma probabilidade de 50% à de não sê-lo. Do mesmo modo, o princípio poderia sugerir atribuir uma probabilidade de 50% à possibilidade de ser o membro mais rico da sociedade e uma probabilidade de 50% à possibilidade de não sê-lo.^{163 164}

O próprio Harsanyi reconhece, em um texto posterior, que o princípio da indiferença é bastante problemático. Ele não apenas é muito vago, não deixando claro como deve ser aplicado em casos específicos, como também comete um salto lógico ao tentar extrair uma conclusão positiva (equiprobabilidade) da ausência de informação.¹⁶⁵

De fato, os autores apontam que o modo com que Harsanyi aplica o princípio da indiferença se deve a um pressuposto moral seu, qual seja, o de que as partes na posição original devem ser imparciais e impessoais em sua escolha.¹⁶⁶ Rawls, por sua vez, não compartilha desse pressuposto de Harsanyi, entendendo que escolhas impessoais não respeitariam a separação entre pessoas ao juntar as preferências de todas em uma grande função de utilidade.¹⁶⁷

Percebe-se, portanto, que a escolha entre o cumprimento do primeiro requisito da posição original de Rawls para justificar o uso da regra *maximin* (qual seja, a aversão a qualquer raciocínio com base em probabilidades) e o raciocínio de Harsanyi é, no fim das contas, uma escolha baseada em crenças morais prévias.

Eu considero as noções de imparcialidade e impessoalidade muito relevantes para uma teoria de justiça e, após um raciocínio em equilíbrio reflexivo, acredito que uma teoria de justiça adequada deva tratar da mesma maneira e com o mesmo peso os interesses de todos os membros de uma sociedade, independentemente da classe social a que pertençam. Além disso, não creio que juntar as preferências de

¹⁶³ GAUS, Gerald Francis; THRASHER, John. **Rational Choice and the Original Position: The (Many) Models of Rawls and Harsanyi**. In: HINTON, Timothy (org.). **The Original Position**. Cambridge: Cambridge University Press, 2015, p. 55.

¹⁶⁴ BINMORE, Kenneth George. **Rational Decisions**. Princeton: Princeton University Press, 2009, pp. 128-129.

¹⁶⁵ HARSANYI, John Charles. **Bayesian Decision Theory, Subjective and Objective Probabilities, and Acceptance of Empirical Hypotheses**. Synthese, vol. 57, no. 3. Berlin: Springer, 1983, p. 352.

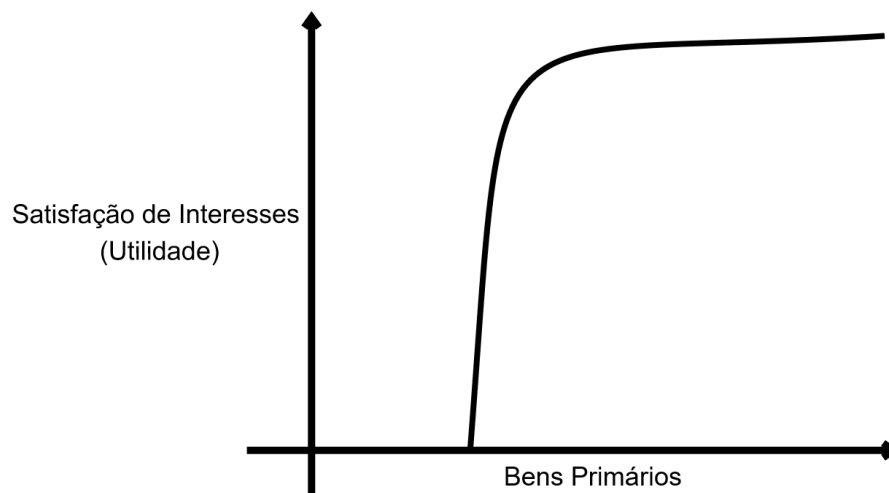
¹⁶⁶ GAUS, Gerald Francis; THRASHER, John. **Rational Choice and the Original Position: The (Many) Models of Rawls and Harsanyi**. In: HINTON, Timothy (org.). **The Original Position**. Cambridge: Cambridge University Press, 2015, p. 56.

¹⁶⁷ Ibid., p. 57.

todas as pessoas em uma função de utilidade “desrespeitaria a separação entre elas”. Entendo, portanto, que uma melhor caracterização da posição original permite a atribuição da mesma probabilidade à possibilidade de ser qualquer membro da sociedade, não restando cumprido o primeiro requisito de Rawls que justificaria o uso da regra *maximin* na posição original.

Os dois outros requisitos são os seguintes: que as partes escolhendo na posição original possuam uma concepção de bem tal que se importam muito pouco com ganhos que poderiam obter acima de um mínimo que poderia ser garantido pelo uso da regra *maximin*; e que situações em que se esteja abaixo deste mínimo que poderia ser garantido seriam tão ruins que não poderiam ser aceitas.¹⁶⁸

Os interesses de pessoas que cumprem ambos estes requisitos poderiam ser representado por um gráfico como este:



Mesmo que estes dois requisitos sejam cumpridos na posição original, não os considero sólidos o suficiente para utilizar a regra *maximin* como critério de escolha no lugar da regra de maximização da utilidade esperada. Isto porque o princípio da utilidade média já leva em consideração tais questões. Se ganhos acima do mínimo que poderia ser garantido aumentam muito pouco a utilidade e perdas que levem uma pessoa para abaixo desse mínimo causem uma perda de utilidade enorme, o

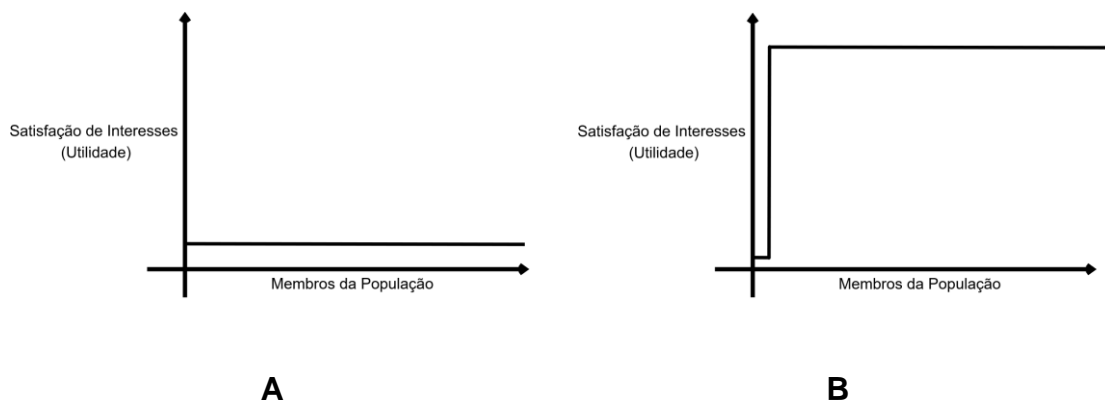
¹⁶⁸ RAWLS, John Bordley. **A Theory of Justice**. Edição revisada. Cambridge, Massachusetts: Harvard University Press, 1999, p. 134.

princípio da utilidade dificilmente recomendaria que alguém ficasse abaixo deste mínimo.

Isso só seria teoricamente possível em situações em que o fato de alguém ficar um pouco abaixo desse mínimo garantisse uma quantidade enorme de ganhos para uma parcela enorme da população, já que, da maneira com que a curva de utilidade das pessoas foi construída, seriam necessários ganhos enormes para muitas pessoas para justificar perdas pequenas de poucas pessoas que estivessem no nível desse mínimo que poderia ser garantido.

Como acredito que as partes na posição original deveriam ser imparciais e impessoais ao decidirem seus princípios de justiça, conforme argumentado acima, elas estariam dispostas a aceitar, em casos raros, perdas nos prospectos de membros da classe menos privilegiada da sociedade, levando-se em consideração que uma quantidade pequena de perdas desse tipo, para ser compensada, exigiria um aumento muito grande dos prospectos dos demais membros da sociedade.

Do contrário, as partes na posição original teriam de preferir o estado de coisas do gráfico A abaixo do que o do gráfico B, o que iria contra os requisitos de imparcialidade e impessoalidade:



A crítica de Harsanyi mencionada na seção anterior é aplicável aqui: considerando-se a maneira com que Rawls caracterizou os interesses das pessoas na posição original em seus dois últimos requisitos para o uso da regra *maximin*, as únicas situações em que o princípio da diferença apontaria para resultados diferentes do princípio da utilidade média são situações em que aquele princípio

levaria a resultados nada razoáveis, como no caso dos gráficos acima. Não seria razoável, portanto, escolhê-lo.¹⁶⁹

Após responder ao raciocínio de Rawls a favor de seus princípios de justiça, faz-se necessário responder às suas críticas ao argumento a favor do princípio da utilidade média a partir da posição original. Ele aponta dois problemas deste raciocínio. O primeiro é o de que, devido à grande importância da escolha na posição original e à necessidade de justificar sua decisão perante terceiros, como descendentes, as partes que nela se encontrassem não se baseariam em probabilidades advindas apenas do princípio da razão insuficiente para embasar sua decisão, exigindo alguma espécie de base empírica para o uso de probabilidades.¹⁷⁰

Este problema pode ser respondido da mesma maneira com que se criticou acima o primeiro requisito de Rawls para tornar o uso da regra *maximin* na posição original algo razoável. Inicialmente, cabe destacar que a teoria Bayesiana de escolha sob incerteza é a teoria padrão para se analisar a escolha de pessoas em situações de incerteza, e ela não rechaça de modo algum a atribuição de probabilidades subjetivas a diferentes situações possíveis.¹⁷¹

O uso do princípio da razão insuficiente, ou princípio da indiferença, por sua vez, é um pressuposto moral, assim como o rechaço a qualquer uso de probabilidades na posição original, conforme desenvolvido acima. É, contudo, um pressuposto moral que considero bastante razoável, pois torna a escolha na posição original imparcial e impessoal, tratando todos os membros da sociedade com igual consideração.

O segundo problema com o raciocínio a favor do princípio da utilidade seria o fato de as pessoas na posição original não possuírem objetivos próprios, mas simplesmente adotarem os objetivos daqueles cujo lugar na sociedade considerassem possível tomar. Rawls afirma que somos capazes de avaliar a situação de outra pessoa a partir do ponto de vista de nossos objetivos últimos, mas

¹⁶⁹ HARSANYI, John Charles. **Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory.** The American Political Science Review, vol. 69, no. 2. Washington: American Political Science Association, 1975, pp. 595-596.

¹⁷⁰ RAWLS, John Bordley. **A Theory of Justice.** Edição revisada. Cambridge, Massachusetts: Harvard University Press, 1999, pp. 145-146.

¹⁷¹ GAUS, Gerald Francis; THRASHER, John. **Rational Choice and the Original Position: The (Many) Models of Rawls and Harsanyi.** In: HINTON, Timothy (org.). **The Original Position.** Cambridge: Cambridge University Press, 2015, p. 56.

questiona se seria possível avaliarmos a situação de outra pessoa a partir do ponto de vista dos objetivos dela.

Para isto, seria necessário que houvesse algum tipo de expectativa unificada. Ele cita como exemplo alguma espécie de regras com que as pessoas concordassem para definir seus fins últimos, que poderiam advir, por exemplo, de regras da psicologia referentes a intensidade de desejos, habilidades naturais e atributos físicos, bens públicos e privados possuídos etc.¹⁷²

Rawls questiona esta pressuposição, afirmando que ela implicaria uma noção de ser humano que não está de acordo com nossos juízos ponderados, pois pessoas assim não teriam um caráter e vontade determinados, variando estes de acordo com circunstâncias contingentes de suas vidas. Ele afirma que estas pessoas deveriam ser consideradas livres, podendo alterar e revisar seus fins últimos livremente.¹⁷³

Assim como Harsanyi o fez ao justificar as comparações interpessoais de utilidade em seu artigo,¹⁷⁴ defendo a tese segundo a qual todos os seres humanos são governados pelas mesmas leis da psicologia. Um mesmo conjunto de regras determina o comportamento e os desejos de todos os seres humanos, e estes variam devido a fatores como carga genética, experiências de vida e contexto social. Do mesmo modo, os fins últimos das pessoas podem ser alterados, mas tais alterações são motivadas por uma interação complexa entre fatores causais biológicos, psicológicos e sociais.

Discordo de Rawls quando ele afirma que esta concepção de ser humano seria uma “pessoa vazia” (“*bare-person*”),¹⁷⁵ o que não estaria de acordo com nossos juízos ponderados. Parece-me que os avanços recentes no campo da psicologia demonstram que as crenças, desejos, emoções e demais atitudes

¹⁷² RAWLS, John Bordley. **A Theory of Justice**. Edição revisada. Cambridge, Massachusetts: Harvard University Press, 1999, pp. 151-152.

¹⁷³ Ibid., pp. 131-132.

¹⁷⁴ HARSANYI, John Charles. **Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory**. The American Political Science Review, vol. 69, no. 2. Washington: American Political Science Association, 1975, p. 600.

¹⁷⁵ RAWLS, John Bordley. **A Theory of Justice**. Edição revisada. Cambridge, Massachusetts: Harvard University Press, 1999, p. 152.

proposicionais dos seres humanos resultam de interações incrivelmente complexas entre neurônios.¹⁷⁶ Estas relações, por sua vez, são físicas e deterministas.

É claro que estas regras da psicologia comuns a todos os seres humanos devem ser incrivelmente complexas, pois há muitos fatores envolvidos, e é questionável se seria possível, em algum momento da história da humanidade, obter um conhecimento profundo sobre elas.¹⁷⁷ Esta complexidade, contudo, não muda o fato de que parece ser o caso que as mesmas regras são aplicáveis a todos os seres humanos, já que conseguimos prever, em alguns experimentos e com razoável acurácia, o comportamento dos seres humanos sob certas condições.¹⁷⁸

De fato, esta noção de ser humano criticada por Rawls me parece estar mais de acordo com nossos juízos ponderados. A questão de como a mente dos seres humanos funciona, incluída aí a questão de o que determina seu caráter e objetivos, é uma questão empírica passível de análise e experimentação. Parece-me muito mais razoável seguir o que apontam estudos científicos recentes de psicologia do que adotar uma ideia de ser humano capaz de mudar autonomamente seus fins últimos que, embora possa parecer razoável durante elucubrações em uma poltrona, não está de acordo com a pesquisa empírica contemporânea.

Posteriormente, Rawls desenvolve mais um argumento a favor de seus princípios de justiça quando comparados ao princípio da utilidade média.¹⁷⁹ Este remete aos esforços do comprometimento, caracterizados no capítulo anterior desta monografia. Como o acordo na posição original é análogo a um contrato entre todas as partes, nenhuma delas poderia realizar um acordo que possivelmente não seria capaz de cumprir. Caso o fizesse, estaria agindo de má-fé.

Para Rawls, o utilitarismo seria inaceitável como princípio de justiça na posição original por esta razão. Ele permite que algumas pessoas tenham suas liberdades básicas sacrificadas em nome de uma utilidade geral maior. Segundo Rawls, muitas pessoas poderiam não ser capazes de aceitar isto de bom grado, não

¹⁷⁶ CHURCHLAND, Paul Montgomery. **Eliminative Materialism and the Propositional Attitudes.** The Journal of Philosophy, vol. 78, no. 2. Nova York: The Journal of Philosophy, Inc., 1981.

¹⁷⁷ ANDRESKI, Stanislav. **Social Sciences as Sorcery.** Londres: André Deutsch, 1972, p. 18.

¹⁷⁸ SOON, Chun Siong; BRASS, Marcel; HEINZE, Hans-Jochen; HAYNES, John-Dylan. **Unconscious Determinants of Free Decisions in the Human Brain.** Nature Neuroscience, vol. 11, no. 5. Londres: Nature Publishing Group, 2008.

¹⁷⁹ RAWLS, John Bordley. **A Theory of Justice.** Edição revisada. Cambridge, Massachusetts: Harvard University Press, 1999, pp. 153-154.

possuindo um altruísmo e empatia grandes a ponto de aceitar sofrer restrições de liberdade em nome do prazer de terceiros.

Considero parte deste argumento questionável. Concordo com Rawls que o utilitarismo poderia levar a resultados que algumas pessoas não aceitariam de bom grado, fazendo-se necessária alguma espécie de coerção para que a utilidade fosse maximizada. Não creio, contudo, que seus dois princípios de justiça se saiam melhor quanto a esta questão.

Como mencionado por Harsanyi, o princípio da diferença poderia levar a enormes sacrifícios nos prospectos de vida de uma grande parte da população em nome de uma pequena melhora na situação dos menos favorecidos.¹⁸⁰ Parece-me ingênuo imaginar que aqueles aceitariam de bom grado ter seus prospectos de vida sacrificados em grande medida em nome de poucos benefícios a terceiros. Quanto a este quesito, acredito que os princípios de justiça de Rawls sejam inferiores ao utilitarismo, pois, embora o utilitarismo possa levar a resultados inaceitáveis para uma minoria em nome de benefícios para uma maioria, os princípios de Rawls podem levar a resultados inaceitáveis para uma maioria em nome de uma minoria. Uma quantidade maior de partes na posição original estaria entrando em um acordo sob má-fé.

Pelo argumento dos esforços do comprometimento, portanto, tanto o utilitarismo quanto os dois princípios de Rawls levariam a resultados que alguns membros do acordo não aceitariam. Ambos necessitariam de alguma espécie de coerção para serem levados a cabo. Entendo, portanto, que o requisito postulado por Rawls de que nenhuma parte poderia aceitar um acordo na posição original que depois possivelmente não cumpriria de bom grado é excessivo.

O argumento seguinte de Rawls a favor de seus princípios de justiça, que considero problemático, faz referência à condição de publicidade.¹⁸¹ Rawls considera um ponto positivo de uma concepção de justiça o fato de ela, sendo publicamente reconhecida pelo povo, desenvolver na população um desejo de agir em seu acordo, criando seu próprio apoio ao longo do tempo.

¹⁸⁰ HARSANYI, John Charles. **Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory.** The American Political Science Review, vol. 69, no. 2. Washington: American Political Science Association, 1975, pp. 596-597 e 605-606.

¹⁸¹ Ibid., pp. 154-155.

Para Rawls, o utilitarismo não geraria este desejo adequadamente, pois exige uma grande capacidade de se identificar com a posição de terceiros para tornar aceitáveis certas desvantagens. Acredito, contudo, que os dois princípios de justiça de Rawls são ainda menos capazes de gerar este desejo na população.

Novamente, concordo com Rawls quando ele afirma que o utilitarismo exige uma alta capacidade de empatia das pessoas para ser seguido voluntariamente. Afinal, algumas pessoas prejudicadas poderiam ter de aceitar sua posição de desvantagem tendo em mente o aumento de utilidade na sociedade. Ainda assim, o princípio da diferença de Rawls poderia exigir que uma grande quantidade de pessoas aceitasse perdas consideráveis de seus prospectos futuros em nome de pequenos benefícios aos menos afortunados. Este princípio, portanto, não exige simplesmente que uma pequena parte da sociedade seja capaz de se colocar no lugar da maioria beneficiada: ele exige que a maior parte da sociedade seja capaz de se colocar no lugar dos menos favorecidos e, posteriormente, valorizar o bem-estar dos menos favorecidos mais do que o seu próprio bem-estar, pois os benefícios aos menos favorecidos seriam menores que os sacrifícios aos mais favorecidos. Em outras palavras, além de um grande nível de empatia, seria necessário que as pessoas fossem capazes de considerar seu bem-estar menos importante que o de um terceiro, e isto seria exigido de uma quantidade maior da população do que no caso do utilitarismo. Por estas razões, considero que os princípios de justiça de Rawls não gerariam um desejo de agir em seu acordo maior que o utilitarismo geraria e, portanto, esta defesa de Rawls de seus princípios de justiça seria falha.

Outra crítica rawlsiana ao princípio da utilidade advém da ideia de que os dois princípios de justiça seriam mais efetivos em nutrir o autorrespeito dos cidadãos caso adotados publicamente, o que seria de suma importância para que eles busquem suas concepções de bem e se sintam realizados ao atingi-las.¹⁸² O utilitarismo não seria efetivo quanto a este aspecto porque todas as pessoas seriam tratadas como um meio para a maximização da utilidade, ao passo que os dois princípios de justiça, ao garantir liberdades mínimas a todos e fazer com que as

¹⁸² Ibid., pp. 155-156.

desigualdades fossem tais que beneficiassem os menos afortunados, desenvolveriam um senso de autorrespeito em todos.

Questiono a alegação de que o utilitarismo impediria o desenvolvimento de um autorrespeito saudável nas pessoas caso adotado como concepção pública de justiça. Segundo Rawls,¹⁸³ o utilitarismo é a teoria ética predominante da modernidade, tendo embasado muitas políticas das democracias modernas. Isto se verifica, por exemplo, ao se analisar a distribuição de recursos em sistemas públicos de saúde. Creio que é do conhecimento da grande maioria das pessoas que a atenção que elas receberão em um sistema de saúde será, em boa parte, definida por princípios utilitaristas. Casos mais severos serão tratados com mais urgência e, em casos terminais, nem todos os recursos serão distribuídos para prolongar a vida do paciente. E, apesar de tudo isto, não me parece que o autorrespeito dos cidadãos é afetado significativamente, deixando-os sem desejo de buscar cumprir sua concepção de bem. Sou cético, portanto, à ideia de que o princípio da utilidade adotado como concepção pública de justiça impactaria significativamente o autorrespeito das pessoas.

Outra resposta utilitarista possível ao argumento de Rawls sobre o autorrespeito, que considero válida, é afirmar que o princípio da utilidade já levaria em conta tais questões. O princípio da utilidade é definido como aquele que considera o justo aquilo que maximize a utilidade das pessoas. Se o autorrespeito por parte dos cidadãos é necessário para que eles busquem seus fins e se sintam realizados ao alcançá-los, o princípio da utilidade determinará que este autorrespeito seja buscado pelo meio mais eficiente. Caso este autorrespeito só seja possível com a garantia de liberdades mínimas e/ou com o princípio da diferença, o utilitarismo determinaria o asseguramento deste(s).

Rawls possui uma resposta a este contra-argumento, que não considero totalmente satisfatória. Ele afirma que a concepção de justiça a ser adotada deve ser pública, e não seria possível, portanto, escolher o princípio da utilidade na posição original e posteriormente se utilizar de princípios não utilitaristas publicamente para a organização da sociedade.¹⁸⁴

¹⁸³ Ibid., p. xvii.

¹⁸⁴ Ibid., pp. 158-159.

Questiono alguns pontos deste argumento. Primeiramente, não me parece que garantir liberdades mínimas ou adotar o princípio da diferença equivaleria a utilizar princípios não utilitaristas publicamente. Não vejo qualquer impeditivo à possibilidade de que a população adote publicamente o princípio da utilidade para estruturar a sociedade, reconheça isto e reconheça também o fato de que garantir liberdades mínimas na constituição, por exemplo, é um bom modo de cumprir sua concepção de justiça. Não me parece que ao fazer isto a população estaria “adotando princípios não utilitaristas”; estaria, na verdade, tomando medidas utilitaristas que também são recomendadas por outras concepções de justiça, como a de Rawls.

Por fim, desejo apresentar outra resposta a este argumento de Rawls, que questiona um de seus pressupostos básicos: a ideia de que a concepção de justiça de uma sociedade deva ser publicamente reconhecida. Rawls pressupõe em sua obra que uma teoria de justiça deve ser pública e conhecida por todos os membros da sociedade.¹⁸⁵ Não parto, contudo, deste pressuposto, e passo a questioná-lo.

Atualmente, a grande maioria das sociedades ao redor do mundo adota uma multitude de valores em suas constituições, muitas vezes conflitantes, e cidadãos com concepções de justiça diametralmente opostas conseguem conviver em relativa harmonia. Parece-me possível que uma sociedade em que a maioria da população não aja sempre reconhecendo uma concepção pública de justiça seja estável e um bom lugar para se viver.

É possível, portanto, que a concepção de justiça de uma sociedade não seja conhecida por todos os seus membros e esta sociedade seja, ainda assim, a melhor sociedade que as partes na posição original poderiam escolher. Logo, a noção de que o mais justo talvez não deva ser conhecido pela maior parte da população não deve ser descartada de antemão.

Esta noção foi bem explicada por Henry Sidgwick, influente filósofo utilitarista do século XIX. O seguinte trecho explicita seu pensamento:

“Portanto, sob princípios Utilitaristas, pode ser certo fazer e recomendar privadamente, sob certas circunstâncias, o que não seria certo advogar abertamente; pode ser correto ensinar abertamente para um grupo de

¹⁸⁵ Ibid., p. 115.

peças o que seria errado ensinar para outras; pode ser concebivelmente certo fazer, se pode ser feito com relativo sigilo, o que seria errado fazer diante do mundo; e até mesmo, se sigilo perfeito pode ser razoavelmente esperado, o que seria errado recomendar por conselho privado ou exemplo. Todas essas conclusões são de um caráter paradoxal: não há dúvida de que a consciência moral de um homem simples repudia amplamente a noção geral de uma moralidade esotérica, diferente da popularmente ensinada; e seria comumente acordado que uma ação que seria má se feita abertamente não se torna boa pelo sigilo. Podemos observar, contudo, que há fortes razões utilitaristas para manter em geral esta última opinião comum.... Portanto, a conclusão Utilitarista, cuidadosamente expressa, pareceria ser isso; que a opinião de que o sigilo pode tornar uma ação certa que não seria tal de outro modo deveria ela mesma ser mantida relativamente secreta; e similarmente parece oportuno que a doutrina de que a moralidade esotérica é oportuna deveria ela mesma ser mantida esotérica. ... um Utilitarista pode razoavelmente desejar, sob princípios utilitaristas, que algumas de suas conclusões deveriam ser rejeitadas pela humanidade em geral; ou mesmo que os vulgares deveriam ser mantidos afastados de seu sistema como um todo, na medida em que a inevitável indefinição e complexidade de seus cálculos tornam provável que leve a maus resultados em suas mãos¹⁸⁶ (tradução livre).¹⁸⁷

A ideia de Rawls de que os princípios de justiça de uma sociedade necessariamente devam ser conhecidos por todos é um pressuposto moral que dificulta, desde o princípio, a escolha do utilitarismo pelas partes na posição original. Contudo, pelas razões acima expostas, não compartilho deste pressuposto rawlsiano e, conseqüentemente, não sou convencido pela argumentação de Rawls a favor de seus princípios de justiça que pressupõe que estes devam ser públicos.

Para finalizar este capítulo, passo a desenvolver uma última crítica à teoria de justiça de Rawls, referente à sua tendência a excluir certos assuntos do seu escopo. Ao longo deste capítulo, foram mencionadas diversas conseqüências inaceitáveis da aplicação dos princípios de justiça de Rawls a certas questões, como a distribuição

¹⁸⁶ SIDGWICK, Henry. **The Methods of Ethics**. 7. ed. Londres: Palgrave Macmillan, 1962, pp. 489-490.

¹⁸⁷ Texto original: "Thus, on Utilitarian principles, it may be right to do and privately recommend, under certain circumstances, what it would not be right to advocate openly; it may be right to teach openly to one set of persons what it would be wrong to teach to others; it may be conceivably right to do, if it can be done with comparative secrecy, what it would be wrong to do in the face of the world; and even, if perfect secrecy can be reasonably expected, what it would be wrong to recommend by private advice or example. These conclusions are all of a paradoxical character: there is no doubt that the moral consciousness of a plain man broadly repudiates the general notion of an esoteric morality, differing from the one popularly taught; and it would be commonly agreed that an action which would be bad if done openly is not rendered good by secrecy. We may observe, however, that there are strong utilitarian reasons for maintaining generally this latter common opinion.... Thus the Utilitarian conclusion, carefully stated, would seem to be this; that the opinion that secrecy may render an action right which would not otherwise be so should itself be kept comparatively secret; and similarly it seems expedient that the doctrine that esoteric morality is expedient should itself be kept esoteric. ... a Utilitarian may reasonably desire, on Utilitarian principles, that some of his conclusions should be rejected by mankind generally; or even that the vulgar should keep aloof from his system as a whole, in so far as the inevitable indefiniteness and complexity of its calculations render it likely to lead to bad results in their hands."

de medicamentos ou recursos para educação. Essas consequências inaceitáveis, inclusive, são a base para a minha resposta ao argumento de Rawls dos esforços do comprometimento.

Uma possível resposta de Rawls e seus defensores a essas alegações seria tentar retirar do âmbito de uma teoria de justiça qualquer situação inaceitável que poderia advir do uso do princípio da diferença. O exemplo de Harsanyi de distribuição de muitos recursos de saúde para tratar casos terminais, por exemplo, seria anulado pela alegação de que “os princípios de justiça devem ser aplicados à estrutura básica da sociedade, logo o sistema de saúde possuiria princípios de justiça próprios, não sendo o princípio da diferença aplicável a ele”. A estrutura básica da sociedade, contudo, é caracterizada de uma maneira um tanto vaga em “Uma Teoria de Justiça”. Não seria possível considerar um sistema público de saúde como parte desta estrutura básica? Ainda assim, um defensor de Rawls poderia argumentar que “ainda que um sistema público de saúde seja considerado parte da estrutura básica da sociedade, ele teria princípios internos próprios para a alocação de recursos”.

Estas respostas são, em minha visão, *ad hoc*; pelo fato de o princípio da diferença levar a casos inaceitáveis se aplicado em certos âmbitos da sociedade, afirma-se que nestes âmbitos outros princípios deveriam ser utilizados. Neste quesito, acredito que Rawls se aproxima um pouco dos intuicionistas que criticou no começo de “Uma Teoria de Justiça” por simplesmente abrirem exceções ao princípio da utilidade quando sua aplicação parecesse inadequada.¹⁸⁸

De fato, Rawls não aborda, em sua teoria de justiça, várias situações em que afirmamos no dia a dia estarem ocorrendo “injustiças”. Sua teoria de justiça não parece tratar adequadamente, por exemplo, de pessoas com deficiência, afirmando o autor em seu paper “Some Reasons for the Maximin Criterion” que assume que as partes na posição original possuem necessidades físicas normais, para evitar o surgimento do problema de assistência de saúde especial.¹⁸⁹

¹⁸⁸ RAWLS, John Bordley. **A Theory of Justice**. Edição revisada. Cambridge, Massachusetts: Harvard University Press, 1999, pp. xvii-xviii.

¹⁸⁹ RAWLS, John Bordley. **Some Reasons for the Maximin Criterion**. In: FREEMAN, Samuel (org.). **John Rawls: Collected Papers**. Cambridge, Massachusetts: Harvard University Press, 1999, p. 227.

Martha Nussbaum desenvolve muito bem esta questão em seu livro “Fronteiras de Justiça”. Ela aponta que Rawls não dá a devida atenção a diversos temas, como pessoas com deficiências físicas e/ou mentais e direitos de animais não humanos. Estes assuntos não deveriam ser tratados como periféricos e ser abordados em outro momento por outros princípios de justiça, mas sim ser levados em consideração desde os fundamentos mais básicos de cada sociedade.

Os seguintes trechos são exemplificativos da crítica da autora:

“Mas o adiamento não é inocente, claramente. É pedido às partes que se imaginem como se representassem cidadãos que são realmente ‘totalmente cooperativos ... ao longo de uma vida completa,’ e, portanto, como se cidadãos não precisassem de cuidados em momentos de extrema dependência. Essa ficção destrói muito do que caracteriza a vida humana e destrói, também, a continuidade entre as chamadas pessoas normais e pessoas com deficiências permanentes. (...) De modo mais geral, o cuidado para com crianças, idosos e pessoas com deficiências físicas e mentais é uma grande parte do trabalho que precisa ser feito em qualquer sociedade e na maioria das sociedades é uma fonte de grande injustiça. Qualquer teoria de justiça precisa pensar sobre esse problema desde o começo, no projeto de sua estrutura institucional básica e, particularmente, em sua teoria de bens primários”¹⁹⁰ (tradução livre).¹⁹¹

“Embora seus comentários sobre os animais não sejam longos, ele não hesita em afirmar que há deveres morais para com os animais, os quais ele chama de “deveres de compaixão e humanidade” (TJ 512). O fato de que animais sentem prazer e dor impõe tais deveres.

Mas para Rawls essas não são questões de justiça e ele afirma explicitamente que a doutrina contratualista não pode ser estendida para lidar com elas “de uma maneira natural”:

(...)

¹⁹⁰ NUSSBAUM, Martha Craven. **Frontiers of Justice**. Cambridge, Massachusetts: Harvard University Press, 2006, p. 127.

¹⁹¹ Texto original: “But the postponement is not innocent, clearly. The parties are being asked to imagine themselves as if they represent citizens who really are “fully cooperating ... over a complete life,” and thus as if citizens have no needs for care in times of extreme dependency. This fiction obliterates much that characterizes human life, and obliterates, as well, the continuity between the so called normal and people with lifelong impairments. (...) More generally, care for children, elderly people, and people with mental and physical disabilities is a major part of the work that needs to be done in any society, and in most societies it is a source of great injustice. Any theory of justice needs to think about the problem from the beginning, in the design of the basic institutional structure, and particularly in its theory of the primary goods.”

O que exatamente significa dizer que essas são questões de justiça, ao invés de questões de “compaixão e humanidade”? (...) Parece que analisar os danos que causamos aos animais apenas em termos de compaixão causa um obscurecimento da importante distinção entre a compaixão que sentimos por um animal que morre de uma doença que não é culpa de ninguém e a resposta que podemos ter aos sofrimentos de um animal que está sendo tratado cruelmente por humanos”¹⁹² (tradução livre).¹⁹³

Este é, portanto, outro problema da teoria de justiça de Rawls: tende a retirar de seu escopo diversas questões relevantes para uma teoria de justiça, por não ser capaz de dar uma resposta adequada a elas. O utilitarismo que Rawls critica, por sua vez, é capaz de tratar teoricamente de todas essas questões, ainda que em alguns casos específicos não seja claro qual política maximiza a utilidade.

Estas são, em suma, minhas principais críticas à teoria de justiça de Rawls. Por estas razões, creio que as partes na posição original não escolheriam seus dois princípios de justiça para fundamentar a estrutura básica de sua sociedade, mas sim o princípio da utilidade média.

¹⁹² NUSSBAUM, Martha Craven. **Frontiers of Justice**. Cambridge, Massachusetts: Harvard University Press, 2006, pp. 331 e 336.

¹⁹³ Texto original: “Although his remarks on animals are not lengthy, he does not hesitate to say that there are moral duties to animals, which he calls “duties of compassion and humanity” (TJ 512). The fact that animals feel pleasure and pain imposes such duties.

But for Rawls these are not issues of justice, and he is explicit that the contract doctrine cannot be extended to deal with them “in a natural way”:

(...)

What exactly does it mean to say that these are issues of justice, rather than issues of “compassion and humanity”? (...) It would seem that analyzing the harms we do to animals in terms of duties of compassion alone entails blurring the important distinction between the compassion we might have for an animal who dies of a disease that is nobody’s fault and the response we might have to the sufferings of an animal who is being cruelly treated by humans.”

CONCLUSÃO

O que pode ser concluído a partir das análises e argumentos desenvolvidos nesta monografia? Inicialmente, que a teoria de justiça de Rawls, apesar de seus inegáveis méritos, é passível de diversas críticas que põem em xeque a noção de que seus princípios de justiça seriam escolhidos pelas partes na posição original descrita em sua obra.

Estas críticas, é claro, não são conclusivas; elas mesmas estão sujeitas a réplicas e tanto a posição a favor da teoria de Rawls quanto a posição contrária a ela são razoavelmente defensáveis. Isso demonstra, contudo, um fato: aqueles que afirmam que a teoria de justiça de Rawls teria “superado” o utilitarismo não estão fazendo uma alegação imparcial e objetiva, já que o utilitarismo continua sendo uma posição ética popular e defensável e possui respostas às críticas rawlsianas. Rawls não fez, em sua obra, alguma espécie de “refutação” do utilitarismo, assim como esta monografia não possui qualquer pretensão de ter “refutado” Rawls.

Outro ponto a ser destacado é o quanto a obra de Rawls ainda é uma fonte rica de análises e desenvolvimento de teses. Há uma imensa literatura acerca da obra de Rawls, apenas uma pequena parte desta tendo sido referenciada nesta monografia. Ainda assim, há pontos e argumentos nesta crítica que não fizeram referência à argumentação de uma obra específica de comentários ou críticas a Rawls. Isto demonstra o quanto a sua teoria de justiça estimula o debate e o pensamento crítico, além do surgimento de novas teorias dentro da ética, da filosofia política e da filosofia do direito.

Também restou demonstrada a relevância da teoria da escolha racional para qualquer análise sistemática da teoria de justiça de Rawls. Sendo as partes na posição original caracterizadas como idealmente racionais, uma análise do que constitui esta racionalidade ideal é indispensável para que se possa alegar que elas escolheriam ou rejeitariam certos princípios de justiça.

Pretende-se, ainda, que a análise aqui feita da teoria da escolha racional mostre que ela não é capaz, por si só, de definir que princípios de justiça seriam escolhidos na posição original. Contudo, somando-se a ela a análise feita dos

argumentos de Rawls e de Harsanyi, sugere-se que a balança de razões pende em favor da escolha do utilitarismo pelas partes na posição original. Esta conclusão, contudo, não é demonstrativa, sendo passível de contestações.

Outra conclusão a ser retirada é a de que, já que a teoria de Rawls, embora bastante abrangente, não aborda muitos temas relevantes para uma teoria de justiça, como os direitos de pessoas com deficiência e de animais não humanos, ela precisa ser alterada para de alguma forma abordar estes temas ou ser rechaçada em nome de uma teoria mais abrangente, como o utilitarismo.

Também foi dado destaque a pressupostos da teoria de Rawls que talvez não fossem percebidos à primeira vista por quem lesse sua obra, como o seu pressuposto, apontado por Gerald Gaus e John Thrasher, de que as partes na posição original rechaçariam atribuições de probabilidade sem bases empíricas à chance de ser algum membro da sociedade, o que não pode ser depreendido da teoria da escolha racional, assim como o seu pressuposto de que uma teoria de justiça deve ser pública. Tornar pressupostos como estes mais claros pode vir a facilitar futuros debates sobre a sua teoria.

A teoria de justiça de Rawls é provavelmente a mais relevante a surgir no último século e deve gerar inúmeras discussões ao longo dos próximos anos. É preciso que se faça uma análise ponderada sobre seus aspectos positivos e negativos, assim como os de teorias diferentes, para que possamos aos poucos avançar no conhecimento ético, jurídico e político e criar uma sociedade melhor para se viver.

REFERÊNCIAS

- ANDRESKI, Stanislav. **Social Sciences as Sorcery**. Londres: André Deutsch, 1972.
- BENTHAM, Jeremy. **An Introduction to the Principles of Morals and Legislation**. Kitchener: Batoche Books, 2000.
- BINMORE, Kenneth George. **Game Theory: A Very Short Introduction**. Oxford: Oxford University Press, 2007.
- BINMORE, Kenneth George. **Rational Decisions**. Princeton: Princeton University Press, 2009.
- CHURCHLAND, Paul Montgomery. **Eliminative Materialism and the Propositional Attitudes**. *The Journal of Philosophy*, vol. 78, no. 2. Nova York: The Journal of Philosophy, Inc., 1981.
- DOWNS, Anthony. **An Economic Theory of Democracy**. Nova York: Harper and Brothers, 1957.
- DREIER, James. **Decision Theory and Morality**. In: MELE, Alfred R.; RAWLING, Piers (orgs.). **The Oxford Handbook of Rationality**. Oxford: Oxford University Press, 2004.
- FREEMAN, Samuel (org.). **The Cambridge Companion to Rawls**. Cambridge: Cambridge University Press, 2003.
- GAUS, Gerald Francis. **On Philosophy, Politics and Economics**. Wadsworth: Cengage Learning, 2008.
- GAUS, Gerald Francis; THRASHER, John. **Rational Choice and the Original Position: The (Many) Models of Rawls and Harsanyi**. In: HINTON, Timothy (org.). **The Original Position**. Cambridge: Cambridge University Press, 2015.
- HARSANYI, John Charles. **Bayesian Decision Theory, Subjective and Objective Probabilities, and Acceptance of Empirical Hypotheses**. *Synthese*, vol. 57, no. 3. Berlin: Springer, 1983.

HARSANYI, John Charles. **Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory.** The American Political Science Review, vol. 69, no. 2. Washington: American Political Science Association, 1975.

HARSANYI, John Charles. **Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility.** Journal of Political Economy, vol. 63, no. 4. Chicago: University of Chicago Press, 1955.

HUME, David. **A Treatise of Human Nature.** Oxford: Clarendon Press, 1960.

KAHNEMAN, Daniel; TVERSKY, Amos Nathan. **Choices, Values, and Frames.** Nova York: Cambridge University Press, 2000.

KAHNEMAN, Daniel; TVERSKY, Amos Nathan. **On the Psychology of Prediction.** Psychological Review, vol. 80, no. 4. Washington: American Psychological Association, 1973.

KANT, Immanuel. **Metafísica dos Costumes.** São Paulo: Editora Vozes, 2013.

LUCE, Robert Duncan; RAIFFA, Howard. **Games and Decisions: Introduction and Critical Survey.** Nova York: John Wiley and Sons, 1957.

NEUMANN, John von; MORGENSTERN, Oskar. **Theory of Games and Economic Behavior.** 3. ed. Princeton: Princeton University Press, 1953.

NISBETT, Richard Eugene; ROSS, Lee. **Human Inference: Strategies and Shortcomings of Social Judgement.** Englewood Cliffs: Prentice-Hall, INC., 1980.

NUSSBAUM, Martha Craven. **Frontiers of Justice.** Cambridge, Massachusetts: Harvard University Press, 2006.

POLLOCK, John Leslie. **Thinking about Acting: Logical Foundations for Rational Decision Making.** Oxford: Oxford University Press, 2006.

RAWLS, John Bordley. **A Theory of Justice.** Edição original. Cambridge, Massachusetts: Harvard University Press, 1971.

RAWLS, John Bordley. **A Theory of Justice.** Edição revisada. Cambridge, Massachusetts: Harvard University Press, 1999.

RAWLS, John Bordley. **Outline of a Decision Procedure for Ethics**. The Philosophical Review, vol. 60, no. 2. Durham: Duke University Press, 1951.

RAWLS, John Bordley. **Political Liberalism**. Nova York: Columbia University Press, 1996.

RAWLS, John Bordley. **Some Reasons for the Maximin Criterion**. In: FREEMAN, Samuel (org.). **John Rawls: Collected Papers**. Cambridge, Massachusetts: Harvard University Press, 1999.

SAMUELSON, Paul Anthony. **Consumption Theory in Terms of Revealed Preference**. *Economica*, vol. 15, no. 60. Londres: Wiley-Blackwell, 1948.

SAVAGE, Leonard Jimmie. **The Foundations of Statistics**. Nova York: John Wiley and Sons, 1954.

SEN, Amartya. **Maximization and the Act of Choice**. *Econometrica*, vol. 65, no. 4. Nova York: The Econometric Society, 1997.

SIDGWICK, Henry. **The Methods of Ethics**. 7. ed. Londres: Palgrave Macmillan, 1962.

SOON, Chun Siong; BRASS, Marcel; HEINZE, Hans-Jochen; HAYNES, John-Dylan. **Unconscious Determinants of Free Decisions in the Human Brain**. *Nature Neuroscience*, vol. 11, no. 5. Londres: Nature Publishing Group, 2008.

URBINATI, Nadia. **On John Rawls' Representational Judgement**. *Contemporary Readings in Law and Social Justice*, vol. 1, no. 1. Nova York: Addleton Academic Publishers, 2009.