

**UNIVERSIDADE FEDERAL DE PELOTAS
INSTITUTO DE FILOSOFIA, SOCIOLOGIA E POLÍTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA**



Equilíbrio Reflexivo Amplo e Justificação Pessoal

TIARAJU MOLINA ANDREAZZA

PELOTAS, 2013

TIARAJU MOLINA ANDREAZZA

EQUILÍBRIO REFLEXIVO AMPLO E JUSTIFICAÇÃO PESSOAL

Dissertação apresentada como requisito parcial para a obtenção do grau de mestre em filosofia no programa de pós-graduação da Universidade Federal de Pelotas.

Orientador: Prof. Dr. Denis Coitinho Silveira

Pelotas, 2013

Dados Internacionais de Publicação (CIP)
Catálogo na Fonte: Maria Fernanda Monte Borges CRB:10/1011
Universidade Federal de Pelotas

A557e Andreazza, Tiaraju Molina
 Equilíbrio reflexivo amplo e justificação pessoal / Tiaraju
Molina Andreazza; Denis Coitinho Silveira, orientador. – Pelotas,
2013.

157 f.

Dissertação (Mestrado em Filosofia), Instituto de Filosofia,
Sociologia e Política, Universidade Federal de Pelotas. Pelotas,
2013.

1.Equilíbrio reflexivo estreito. 2.Equilíbrio reflexivo amplo.
3.Justificação pessoal. I. Silveira, Denis Coitinho , orient. II.
Título.

CDD: 100

Banca examinadora:

Dr. Denis Coitinho Silveira (orientador)

Dr. Carlos Adriano Ferraz

Dr. Darlei Dall'Agnol

Agradecimentos

À CAPES, pelo financiamento.

Aos professores do Programa de Pós-Graduação em Filosofia da UFPEL com os quais tive a oportunidade de cursar créditos e aprimorar os meus conhecimentos em filosofia.

Ao Prof. Dr. João Hobuss, por gentilmente ter lido praticamente toda esta dissertação na ocasião do exame de qualificação e encorajado o seu desenvolvimento e também ao Prof. Dr. Carlos Adriano Ferraz, que além de ter participado da banca de qualificação e fornecido sugestões, comentários e críticas pertinentes e construtivas, cordialmente também aceitou fazer parte da banca de defesa.

Ao Prof. Dr. Darlei Dall'Agnol, pela sua franca disposição em aceitar o convite para compor a banca de defesa.

Aos colegas da graduação e do mestrado que ao longo dos últimos anos contribuíram diversamente para esta dissertação.

Aos amigos Carlos e Arthur, por me auxiliarem a cultivar certas características imprescindíveis àqueles que gostam de se ocupar com questões de filosofia.

Ao meu tio, Paulo, por ter de tão bom grado se disposto à ingrata tarefa de corrigir os meus deslizes gramaticais.

Ao meu orientador, Prof. Dr. Denis Coitinho Silveira, não apenas por sua orientação esplêndida no mestrado, mas, sobretudo, pelo seu comprometimento com a minha formação acadêmica nesses últimos cinco anos. Por ter posto todo o seu conhecimento filosófico, a sua seriedade e o seu profissionalismo ao meu dispor, a sua importância para esta dissertação é simplesmente inestimável.

Aos meus pais, Antônio e Josi, e aos meus irmãos, pelo suporte próximo e necessário, indispensável para que os meus estudos progredissem tranquilamente.

À Tania, por tudo.

*Talk your own nonsense and I'll
kiss you for it.*

Razumikhin, em “Crime e
Castigo” (Fiódor Dostoiévski)

Equilíbrio Reflexivo Amplo e Justificação Pessoal

RESUMO: Desde a sua nominal formulação por John Rawls em *A Theory of Justice* (1971), e posterior popularização por uma série de ensaios de Norman Daniels (1979-1983), o método do equilíbrio reflexivo amplo tornou-se parte relevante do debate metaético e epistemológico. Não obstante esse reconhecimento, desde então a compreensão que se possui do método avançou pouco além dos ensaios do Daniels. O objetivo desta dissertação é preencher essa lacuna argumentando que a filosofia de Rawls contém pressupostos a partir dos quais é possível articular uma versão alternativa do método. Contrariamente à leitura padrão favorecida por Daniels, é defendido que o equilíbrio reflexivo amplo é utilizado por Rawls para mostrar que a justiça como equidade é a resposta para a seguinte questão: se uma pessoa razoável estiver pessoalmente justificada em sustentar as suas crenças morais, que concepção de justiça ela sustentaria em reflexão? Partindo dessa caracterização, o equilíbrio reflexivo amplo é apresentado como o esboço de uma concepção moderadamente fundacionalista de justificação pessoal. Com base nessa imagem do método são discutidas algumas das objeções tradicionalmente endereçadas contra ele, como as objeções do descritivismo e do conservantismo. Conclui-se que essas críticas, se devidamente interpretadas, apontam acertadamente algumas limitações do ERA, embora seja contestável se essas limitações representem falhas genuínas no método proposto por Rawls.

PALAVRAS-CHAVE: Equilíbrio Reflexivo Estreito. Equilíbrio Reflexivo Amplo. Justificação Pessoal.

Wide Reflective Equilibrium and Personal Justification

ABSTRACT: Ever since its nominal formulation by John Rawls in *A Theory of Justice* (1971), and later popularization by a series of papers published by Norman Daniels (1979-1983), the method of wide reflective equilibrium has remained as an important part of the epistemological and metaethical debates. In spite of this role, since then the understanding of the method has not progressed substantially beyond Daniels' papers. The aim of this dissertation is to fill this gap by arguing that Rawls's philosophy has presuppositions from which it is possible to articulate an alternative version of the method. Contrary to the standard reading put forward by Daniels, it is defended that the wide reflective equilibrium is laid out to inform that justice as fairness is the answer to the following question: if a reasonable person were personally justified in believing her moral beliefs, what conception of justice she would believe upon reflection? The wide reflective equilibrium is presented as the outline for a model of personal justification that is moderately foundationalist. Based on this image of the method are examined some objections traditionally addressed against it, such as the objections of descriptivism and conservatism. It is concluded that those objections correctly indicates some of its limitations, although it is uncertain whether such limitations are genuine flaws in Rawls's method or not.

KEYWORDS: Narrow Reflective Equilibrium. Wide Reflective Equilibrium. Personal Justification.

Lista de abreviaturas e orientações para o leitor

Para o equilíbrio reflexivo:

- *Narrow Reflective Equilibrium*: ERE
- *Wide Reflective Equilibrium*: ERA
- *General Reflective Equilibrium*: ERG
- *Full Reflective Equilibrium*: ERP

Para as obras e artigos de Rawls citados:

- *Outline for a Decision Procedure for Ethics*: Outline
- *A Theory of Justice*: TJ
- *Fairness to Godness*: FG
- *Reply to Alexander and Musgrave*: RAM
- *A Kantian Conception of Equality*: KCE
- *The Independence of Moral Theory*: IMT
- *Kantian Constructivism in Moral Theory*: KC
- *Justice as Fairness: Political, Not Metaphysical*: JPFNM
- *Political Liberalism*: PL
- *Reply to Habermas*: RH
- *Justice as Fairness: A Restatement*: JF
- *The Domain of the Political and Overlapping Consensus*: DPOC
- *The Idea of Public Reason Revisited*: PRR

Designo o *wide reflective equilibrium* com a sigla ERA, que é a abreviatura para a expressão traduzida como *equilíbrio reflexivo amplo*. O *narrow reflective equilibrium* é referido com a sigla ERE, em alusão à expressão traduzida em português por *equilíbrio reflexivo estreito*. O *general reflective equilibrium* é citado como ERG como tradução para *equilíbrio reflexivo geral*. A expressão *full reflective equilibrium* é traduzida por *equilíbrio reflexivo pleno* e mencionada através da sigla ERP.

Haverá ocasiões em que utilizarei simplesmente a expressão *equilíbrio reflexivo*. O contexto especificará se estou me referindo ao estreito, amplo, geral ou pleno ou a todas as versões ao mesmo tempo.

Referir-me-ei às obras e artigos de Rawls com as abreviaturas listadas acima. Para obras e artigos de outros autores, seguirei o padrão AUTOR, ano, p.x. No caso de *A Theory of Justice*, utilizarei tanto a paginação da edição original quanto da edição revisada, utilizando o prefixo rev. para indicar a referência para esta última. Quando as duas paginações não estiverem acompanhadas, é porque a referência em questão consta apenas em uma das edições.

A paginação dos ensaios de Rawls remete à coleção *Collected Papers* (Edited by Samuel Freeman. Cambridge: Harvard University Press, 1999), exceto pelo artigo *Reply to Habermas*, cuja citação remete à sua republicação em *Political Liberalism Expanded Edition* (New York: Columbia University Press, 2005). Adotarei as seguintes edições das demais obras: *A Theory of Justice* (Cambridge: Harvard University Press, 1971), *A Theory of Justice: Revised Edition* (Cambridge: Harvard University Press, 1999), *Justice as Fairness: A Restatement* (Edited by Erin Kelly. Cambridge: Harvard University Press, 2001).

Para os artigos de Norman Daniels, farei uso da sua coleção *Justice and Justification: Reflective Equilibrium in Theory and Practice* (Cambridge, Cambridge University Press, 1996), onde eles encontram-se integralmente republicados.

Todas as traduções das citações encontradas neste trabalho são originais e de minha inteira responsabilidade.

Sumário

Introdução	12
1. O Equilíbrio Reflexivo Estreito	19
1.1 O <i>rationale</i> em Nelson Goodman.....	19
1.1.1 ERE e justificação	25
1.2 A aplicação à filosofia moral no <i>Outline</i>	30
1.3 Duas objeções	38
1.3.1 Conservantismo epistêmico	38
1.3.2 Descritivismo.....	50
2. O Equilíbrio Reflexivo Amplo	53
2.1 A concepção de teoria moral	53
2.2 A relação com a posição original	62
2.3 Primeira pessoa vs terceira pessoa.....	69
2.4 ERA no <i>Political Liberalism</i> e <i>Reply to Habermas</i>	75
2.4.1 O internalismo de razões	82
2.5 Amplo versus estreito	94
3. Justificação em Equilíbrio Reflexivo Amplo.....	98
3.1 A posição epistemológica.....	98
3.1.1 O conceito geral de justificação.....	98
3.1.2 Fundacionalismo moderado ou coerentismo holista?.....	102
3.1.3 A ideia de justificação pessoal.....	119
3.1.4 A ideia de justificação pública.....	125
3.2 Duas objeções reconsideradas	132
3.2.1 Conservantismo epistêmico	132
3.2.2 Descritivismo.....	136
Considerações finais	147
Referências	152

Introdução

Antes de começar a escrever eu contava com um plano claro: analisar como Rawls justifica a sua teoria da justiça, a justiça como equidade, e avaliar essa justificação a partir do diálogo com os seus críticos. Como a justificação é uma questão de equilíbrio reflexivo, a minha intenção inicial era escrever uma dissertação sobre o problema da justificação no método do equilíbrio reflexivo. Esse foi apenas um plano inicial, contudo. Quanto mais pensava seriamente sobre o assunto, tanto mais me convencia da sua impraticabilidade. Dois foram os problemas: o que é justificação e o que é equilíbrio reflexivo? Durante os seus escritos, Rawls formulou três versões do equilíbrio reflexivo: a estreita, a ampla e a geral (há também uma que é denominada de plena, grosso modo, ela designa uma situação em que as condições de justificação impostas pelas versões ampla e geral são satisfeitas ao mesmo tempo)¹. A ampla foi desenvolvida como uma reformulação da estreita, e a geral (e plena) como um complemento à ampla. Paralelamente, Rawls concebeu três noções de justificação: *pro tanto*, plena e pública². Esses três conceitos respondem a, ou são derivados de, um quarto *conceito geral* apresentado já no apagar das luzes em *A Theory of Justice*: justificar é “convencer os outros, ou nós mesmos, da razoabilidade dos princípios”³.

Eu sempre soube da existência de toda essa maquinaria conceitual. O que eu não sabia é que seria extremamente difícil tanto compreender exatamente o que vem a ser cada uma das versões do equilíbrio reflexivo, quanto relacionar cada um daqueles conceitos de justificação com cada uma dessas versões. A questão “o que é justificação para o equilíbrio reflexivo?” logo me pareceu ingênua e improdutiva. Na sua maioria, os artigos que a abordavam me abatiam como imprecisos ou incompletos. Ou o autor típico escrevia sobre justificação sem precisar o que entendia por esse conceito, ou não o associava a uma versão definida do equilíbrio reflexivo, ou não definia nenhuma versão particular do método, ou simplesmente escrevia sobre todos os conceitos de justificação e todas as versões do equilíbrio reflexivo ao mesmo tempo.

No percurso de esclarecer as coisas, cheguei à conclusão de que a questão original – o que é justificação para o equilíbrio reflexivo? - exigiria mais do que uma dissertação, de modo que, para não cometer os mesmos erros usuais, eu deveria reparti-la em partes, assim abandonando o plano inicial de escrever sobre o equilíbrio reflexivo como um todo. Como a

¹ *JF*, p.29-32.

² *RH*, p.386-387.

³ *TJ*, p.580-581/rev, p.508-509.

versão estreita é corretamente abandonada por Rawls após o *Outline*, e a versão geral não pode ser compreendida sem uma sólida compreensão da ampla, concluí que o equilíbrio reflexivo amplo (ERA) seria a escolha natural. Ele compõe um método autossuficiente e, como sustentarei, encerra em si mesmo um conceito específico de justificação que independe dos demais, um conceito que é denominado nesta dissertação de justificação pessoal. Além disso, discutir o ERA é uma oportunidade para dialogar mais diretamente com a versão que para muitos é sinônimo de equilíbrio reflexivo, sendo por isso a versão mais presente no cenário da metaética e epistemologia moral das últimas décadas. Assim, a questão original foi substituída pela seguinte: o que é justificação para o ERA?

Ao lado de outras, essa escolha trouxe consequências que se não forem devidamente esclarecidas poderão gerar falsas expectativas. Para adiantar o que se deve e o que não se deve esperar encontrar nesta dissertação, abaixo menciono algumas dessas consequências.

(i) Como esta dissertação não versa sobre o equilíbrio reflexivo como um todo, ela não analisa todo o argumento justificacional que Rawls oferece em favor da justiça como equidade. Como o próprio Rawls se esforçou em tornar claro, o seu argumento possui dois estágios: um que se concentra na posição original e no ERA, e outro, posterior, que depende das ideias de consenso sobreposto (*overlapping consensus*) e estabilidade (*stability*)⁴. Fazendo uma série de movimentos que não convém aqui detalhar, podemos distinguir um primeiro estágio caracterizado pela versão ampla e pela ideia de justificação plena (que reinterpreto em termos de justificação pessoal) e um estágio subsequente em que a versão geral e as ideias de justificação *pro tanto* e pública operam. Excetuando-se algumas poucas referências superficiais, esta dissertação nada dirá a respeito do segundo estágio.

(ii) Ela não pretende apresentar a filosofia de Rawls. Assumo do leitor um conhecimento básico da conceituografia de *A Theory of Justice* e do *Political Liberalism*. Porém, como também se busca apresentar o ERA tomando a filosofia de Rawls como ponto de partida, será inevitável tratar de outros elementos conhecidos da sua teoria. Tratar desses outros elementos, mais do que contribuir para um entendimento da interpretação que será proposta do ERA, permitirá mostrar como essa interpretação se harmoniza com o restante do argumento de Rawls e como ela permite entendê-lo melhor.

(iii) Ela não é uma dissertação que trata de temas de filosofia política, sendo mais bem classificada como uma análise metaética ou epistemológica do ERA. Muitos leitores podem perplexar-se imediatamente ao ler isto. Eles podem acusar-me de estar traindo o aspecto mais

⁴ DPOC, p.486n27.

fundamental e original da proposta de Rawls, que é defender que a filosofia política deve se manter “filosoficamente na superfície”, sendo independente do restante da filosofia, especialmente da epistemologia. Não posso responder agora, mas adianto o que direi: tudo depende o que queremos significar com “epistemologia”. Quando Rawls afirma que a filosofia política deve ser independente da epistemologia, ele supõe uma noção bastante limitada e restrita do que seja essa disciplina.

(v) Ela ignora uma série de questões relevantes para a avaliação crítica do ERA mesmo em Rawls, questões que muitos autores plausivelmente mobilizaram como objeções. Por exemplo, Rawls afirma que os nossos juízos morais ponderados mais básicos (“pontos fixos provisórios”, na sua definição) são o de tolerância religiosa e de repúdio à escravidão, e que podemos extrair desses juízos certas ideias morais neles implícitas, como as concepções de pessoa como livre e igual e de sociedade como um sistema cooperativo. Como a justiça como equidade descreveria essas ideias, ele argumentou que aqueles que seguissem o ERA acabariam por endossá-la em detrimentos de teorias rivais, como a utilitarista. Essas três teses são extremamente disputáveis. É duvidoso se esses são realmente os nossos pontos fixos provisórios, se aquelas ideias de fato podem ser extraídas daquelas desses pontos fixos em detrimento de ideais conflitantes e, por fim, não é óbvio que alguém poderia aceitar a justiça como equidade mesmo admitindo que ela descreve os pontos fixos provisórios reais e extrai corretamente as ideais realmente subjacente a eles. Esta dissertação ignorará essas questões. Elas são desnecessárias para o projeto de considerar o que significa estar justificado em ERA. Assim, se a justiça como equidade fosse realmente afirmada em ERA, que tipo de credencial epistêmica ela teria? Esta é a questão desta dissertação.

(vi) Como consequência de ser uma dissertação limitada ao ERA, ela não avaliará críticas epistemológicas fundamentais ao projeto de Rawls (como a crítica endereçada por Habermas à ideia de justificação pública por consenso sobreposto⁵). A consequência do que será aqui discutido para esses problemas mais amplos também não será objeto de investigação.

(vii) Por vezes defenderei que Rawls deveria dizer o que não disse, ou que ele não deveria dizer o que disse, ou que ele caracterizou erradamente o ERA e a sua própria posição. Esses são comentários de natureza crítica, pois, em suma, direi como Rawls deve conceber o ERA. Porém, essas sentenças serão sempre circunscritas: o que eu realmente estarei dizendo é que, dados outros elementos seus, Rawls poderia ter concluído diferentemente.

⁵ HABERMAS, 1995, p.109-131.

(viii) Embora em muitos momentos ela tenha a forma de um comentário sobre Rawls, o seu objetivo final é articular uma versão autossuficiente do ERA. Será efetivamente defendido que essa é a melhor interpretação possível do que Rawls escreveu, e que ele poderia aceitá-la sem abrir mão de nenhum elemento substancial da sua posição. A interpretação que será apresentada não se ajusta perfeitamente a todo o texto de Rawls. Em verdade, duvido que haja qualquer interpretação que consiga tal ajuste. Se por seu mérito ou demérito, uma peculiaridade notável a respeito da obra de Rawls é que ela foi interpretada (e utilizada) divergentemente por várias pessoas. Não foram apenas divergências menores, mas desacordos realmente profundos. Neo-kantianos (e alguns kantianos mais tradicionais), pragmatistas, céticos e até naturalistas viram em Rawls um amparo para as suas respectivas posições. Cada uma dessas leituras têm de ignorar certos pontos e destacar outros. Ao oferecer a presente interpretação, evidentemente penso que ela ignora o que pode ser ignorado, e destaca o que deve ser destacado, com isso sendo, se não a posição de Rawls, uma posição “rawlsiana”. Contudo, não nutro a expectativa de oferecer a única leitura viável do pensamento de Rawls.

O que esta dissertação pretende então é caracterizar o ERA a partir de Rawls e investigar o conceito de justificação a ele vinculado. Para tanto, ela é estruturada em três capítulos: o primeiro versa sobre o equilíbrio reflexivo estreito (ERE), o segundo apresenta os traços fundamentais do ERA no contexto da obra de Rawls e o terceiro analisa a sua posição epistemológica e apresenta a ideia de justificação pessoal como a sua melhor interpretação.

Por que começar pelo ERE? Em poucas palavras, porque o próprio Rawls indica que o ERA é um refinamento (para melhor) do ERE. Quando o ERA é anunciado pela primeira vez em *A Theory of Justice*, Rawls indica que está seguindo tanto a concepção de justificação de Goodman, em *Fact, Fiction and Forecast* (1953), quanto a noção de teoria moral do seu antigo ensaio, *Outline for a Decision Procedure for Ethics* (1951) (uma versão resumida da sua tese de doutoramento)⁶.

O primeiro capítulo é dividido em três seções principais. Na primeira, começo descrevendo brevemente como Goodman definiu e utilizou o ERE para a resolução de problemas no âmbito da lógica indutiva, buscando caracterizar alguns de seus pressupostos epistemológicos. O que é primeiramente digno de nota nessa história incipiente do equilíbrio reflexivo, é que já na sua gênese ele é desenhado como um mecanismo de justificação para regras ou princípios regulativos intencionalmente designado para situar-se no que poderíamos chamar de uma posição de “deflacionamento epistêmico”. Via ERE, Goodman sustenta que

⁶ TJ, p.20n7/rev, p.18n7; TJ, p.46n24/rev, p.40n24

podemos nos considerar justificados indutivamente mesmo não dispondo de meios para saber se a nossa indução é um caso de conhecimento e pode vir a ser verificada como verdadeira ou falsa. Goodman propõe uma desconexão forte entre justificação e verdade, e essa desconexão é um traço comum a todas as versões do equilíbrio reflexivo.

O outro elemento do ERA antecipado por Goodman é a rejeição da dicotomia entre descrição e normatividade. Goodman argumenta que uma regra válida é justificada pela sua conformidade com raciocínios indutivos efetivamente realizados na prática indutiva de uma determinada comunidade e, estes, por sua vez, são justificados quando conformes a regras válidas. Como Goodman reconhece, o argumento é “flagrantemente circular” porque tanto regras (princípios normativos) quanto inferências particulares (descrições) são justificadas por um processo de ajuste mútuo segundo o qual uma regra deve “ser corrigida se ela gera uma inferência que nós estamos indispostos a aceitar” e uma inferência deve “ser rejeitada se ela viola uma regra que nós estamos indispostos a corrigir”⁷. O ERE nada mais é do que esse processo de justificação por ajuste mútuo para regras e inferências indutivas.

As duas seções restantes do primeiro capítulo são destinadas ao *Outline*. Na segunda seção apresento a posição de Rawls nele defendida e, na última seção, passo a argumentar em favor de duas objeções que impõem a necessidade de rejeitá-la. O problema principal do *Outline* é o seguinte: Rawls defende que um princípio moral é um caso de conhecimento ético quando ele é resultado de uma sistematização de comprometimentos morais de um agente idealizado, ou quando ele está em uma relação de ajuste mútuo com esses comprometimentos, mas não oferece razões convincentes para mostrar que esses comprometimentos morais são epistemicamente confiáveis e credíveis. Como consequência, não há razão para pensar que os princípios morais resultantes do ERE não são mais do que meras descrições de comprometimentos morais.

Depois do primeiro capítulo, passo a dedicar-me efetivamente à versão ampla do equilíbrio reflexivo. Analisando em retrospectiva o restante da dissertação, passei a considerá-la uma reflexão sobre uma conferência proferida por Rawls no outono de 1995, na *Santa Clara Conference on Rawls' Work*. Infelizmente, o conteúdo da mencionada conferência nunca veio a ser oficialmente publicado e, tanto quanto sei, tudo o que dele nos resta é um breve excerto citado por Daniels no seu último ensaio acerca do equilíbrio reflexivo. No trecho preservado por Daniels, Rawls diz que *A Theory of Justice* foi “apresentada como um projeto que indivíduos podem estudar na sua tentativa de alcançar o auto-entendimento do ERA”, um

⁷ GOODMAN, 1983, p.63-64.

projeto “solitário porque eles refletem sobre os seus próprios juízos morais ponderados”⁸. Essas afirmações não são de fácil interpretação. No esforço de entendê-las identifiquei três características *próprias* do ERA que são aparentemente subentendidas: (i) ele é o ponto de vista de uma pessoa tomada individualmente, é (ii) designado para especificar metodologicamente como se deve refletir e o (iii) seu resultado é um senso de justiça que o indivíduo pode ver como aceitável em reflexão ou, como direi, pessoalmente justificado. Nos capítulos segundo e terceiro, esforço-me para dar sentido a (i), (ii) e (iii). Essas são as alegações decisivas que esta dissertação pretende analisar e explicar.

O capítulo segundo é destinado à caracterização do ERA seguindo mais de perto o texto de Rawls. O que primeiramente desafia uma abordagem do ERA é a notória dificuldade em compreender como ele se relaciona com os outros elementos da justiça como equidade. Para superar esse obstáculo, o capítulo foi dividido em cinco partes. As três primeiras partes apresentam a estrutura do ERA e mostram como o método se relaciona com a posição original e a noção de teoria moral de Rawls. A quarta seção do capítulo é dedicada ao estudo do ERA no ambiente conceitual do *Political Liberalism*. O objetivo é sugerir que a tese de Rawls de que uma teoria da justiça deve ser *freestanding* e *política* é compatível com uma interpretação possível do internalismo de razões de Bernard Williams. Tento mostrar como esse internalismo permite explicar, por uma nova perspectiva, por que Rawls julgou que a filosofia política não pode mobilizar valores abrangentes na estipulação do seu conteúdo. Na quinta parte encerro o capítulo comparando o ERA com ERE para indicar que Rawls e outros identificam erradamente quais são as diferenças entre as duas versões.

Algo que a todo o tempo busco tornar coerente no segundo capítulo é a ideia de que o ERA, diferentemente do ERE, não é o que poderíamos chamar de um *método em terceira pessoa* que um teórico moral pode fazer uso a fim de identificar um conjunto de princípios que explica um conjunto de juízos morais, mas um *método em primeira pessoa* que especifica como devemos deliberar ou raciocinar moralmente. Partindo dessa característica de primeira pessoa, no terceiro capítulo sustento a conclusão de que o *desideratum* epistêmico do ERA é oferecer, para a pessoa que o emprega, *justificação pessoal*, e que essa justificação não precisa ser coerentista, mas pode ser mais bem compreendida como um tipo moderado de fundacionalismo.

⁸ “*TJ* não endereça pessoas como cidadãos, mas como indivíduos tentando desenvolver sua própria concepção de justiça e sua aplicação às instituições políticas e sociais básicas da sociedade democrática. Na maior parte do tempo a sua tarefa é solitária porque eles refletem sobre os seus próprios juízos ponderados com os seus pontos fixos e vários primeiros princípios e conceitos intermediários e ideais que eles afirmam. *TJ* é apresentada como um projeto que indivíduos podem estudar na sua tentativa – nunca completamente alcançada e que se está sempre em busca – de alcançar o auto-entendimento do ERA” (Rawls via DANIELS, 1996, p.145. Os itálicos são meus.)

A tese fundamental de Rawls com o ERA, como eu então defendo, é a de que se um indivíduo razoável de um regime democrático contemporâneo seguir o ERA, ele apoiará a justiça como equidade e estará justificado em sustentar o senso de justiça que professa. O que a justiça como equidade faz é interpretar nossas “certezas morais” e organizá-las, por meio da posição original, nos termos de uma teoria, assim transmitindo a confiança com que sustentamos certos juízos morais para outros juízos e princípios justificados a partir deles. Proponho que, naquilo que concerne ao ERA, o objetivo de Rawls é responder à seguinte questão: dado o que eu acredito ser o correto, o que mais eu devo sustentar em questões de justiça para estar *pessoalmente justificado* no senso de justiça que professo?

Mas o que é o ERA? Essa questão é não exatamente fácil de responder e não há uma definição incontroversa. Neste trabalho é defendido que o ERA é um conjunto de recomendações sobre como um indivíduo deve raciocinar quando ele está decidindo o que fazer ou o que pensar. Na medida em que esse conjunto estipula um ideal de reflexão moral ou de deliberação ética, o ERA pode ser então definido como um curso hipotético de reflexão ideal. É claro, o ERA é uma espécie de critério justificacional por meio do qual um indivíduo pode distinguir entre juízos morais (pessoalmente) justificados e juízos injustificados. Pode-se dizer que uma crença p é pessoalmente justificável para uma pessoa S se e somente se S poderia aceitar p em ERA. As crenças que são pessoalmente justificáveis para S são aquelas crenças que ele teria *se* seguisse o ERA.

De posse desses elementos caracterizados, na parte final do capítulo eu retorno às objeções analisadas no capítulo primeiro. Apoiando-me no fundacionalismo moderado e na ideia de justificação pessoal, argumento que a conservantismo epistêmico não constitui uma séria ameaça ao ERA. Por outro lado, defendo que o ERA não dispõe de elementos para afastar a objeção do descritivismo, e discuto as razões pelas quais considero que seja discutível se essa sua limitação de fato representa uma falha.

1. O Equilíbrio Reflexivo Estreito

O ERA não foi uma criação inédita de *TJ*. Por meio de duas notas de rodapé, nessa obra Rawls indica aos seus leitores que as origens do método remontam, de um lado, a Nelson Goodman em seu livro *Fact, Fiction and Forecast*⁹ e, de outro, a um antigo artigo de sua própria autoria, intitulado *Outline for a Decision Procedure for Ethics*¹⁰. Nessas duas referências está presente o que depois se convencionou chamar de versão estreita do equilíbrio reflexivo. Como o ERA, a versão do método efetivamente empregada em *TJ* e no *PL*, é resultado de um aperfeiçoamento do ERE, e como os críticos em geral rejeitam o ERA denunciando a sua incapacidade de representar um genuíno avanço em relação às falhas do ERE, uma dissertação como esta não pode deixar de investigar as suas origens. O objetivo deste capítulo é caracterizar o ERE e, seguindo os críticos, destacar as suas insuficiências e fraquezas. Para tanto, ele é dividido em três partes. Na primeira é analisado como o ERE é utilizado por Goodman e que tipo de pressupostos epistêmicos pode ser identificado. Na segunda parte é exposto como o ERE é aplicado à filosofia moral no *Outline* e quais são as diferenças básicas entre Goodman e Rawls. De posse dessa visão geral, a parte final do capítulo avalia duas objeções que podem ser direcionadas especificamente contra o ERE tal como ele é concebido no *Outline*.

1.1 O *rationale* em Nelson Goodman

Goodman entende que desde Hume se sabe que o raciocínio indutivo não pode ser validado a partir da experiência, posto que envolveria uma petição de princípio, nem pode ser logicamente inferido dela, pois não haveria conexão lógica entre eventos distintos. Goodman parte do pressuposto de que as reservas de Hume quanto à possibilidade de se estender o conhecimento a questões gerais ou a questões sobre o futuro não foram respondidas. Não foram não em virtude da inabilidade dos filósofos, mas em razão da natureza insolúvel do próprio problema.

Goodman é da opinião de que a única maneira de afastar o tradicional problema da indução é dissolvendo-o, isto é, é ignorando a questão de saber se nossas predições são verdadeiras ou falsas (mesmo provavelmente verdadeiras) e se há qualquer tipo de garantia indutiva¹¹. Para Goodman o problema real, que de fato pode ser resolvido, é o problema de definir projeções *válidas*, definir a relação adequada entre a evidência e a hipótese. Nas suas palavras, “o que nós queremos é um modo geral e preciso de dizer quais hipóteses são confirmadas por, ou que projeções são validamente feitas a partir de, uma dada evidência”¹². Por exemplo, considere a hipótese “todos os cisnes são brancos” e a evidência “todos os cisnes observados até hoje são brancos”. Para Goodman simplesmente não há como saber se o enunciado geral expresso nessa hipótese é verdadeiro ou falso, mas nós podemos saber se ele é válido ou inválido. Ele será válido se a evidência confirmar a hipótese.

⁹ *TJ*, p.20n7/rev, p.18n7.

¹⁰ *TJ*, p.46n24/rev, p.40n24.

¹¹ GOODMAN, 1983, p.64.

¹² GOODMAN, 1983, p. 84.

O problema então é o de distinguir entre induções válidas, em que a hipótese é inferida de uma evidência adequada, e induções inválidas, em que a generalização de uma hipótese a partir de uma dada evidência não é legítima. Esse problema é resolvido, segundo Goodman, formulando regras de indução que permitam fazer essa distinção. O ERE é utilizado para encontrar e mostrar a validade *dessas* regras e para solucionar *esse* problema.

O ERE é inserido por Goodman quando ele pede que se considere mais seriamente o que Hume pretendeu defender com a ideia de hábito. Para Hume nós fazemos certas projeções específicas porque observamos que do evento T_1 sempre decorre o evento T_2 , e que toda vez que verificamos T_2 ele está associado de alguma forma relevante com T_1 . Da observação dessa regularidade a mente forma um hábito a partir do qual sempre fazemos a inferência de T_2 quando confrontados com T_1 . Segundo o entender de Goodman, a tradição sempre compreendeu essa explicação de Hume como a exposição de uma objeção ao conhecimento indutivo (uma vez que ele seria meramente um hábito) e também como uma descrição psicológica de como nós raciocinamos indutivamente. O problema normativo, concernente à validade da indução, sequer teria estado entre as preocupações de Hume. Goodman discorda dessa interpretação. Para ele, esse viés interpretativo tradicionalmente adotado parte da premissa equivocada de que a questão da validação da indução é distinta daquela sobre como a indução é realizada efetivamente. Para Goodman não é possível traçar uma dicotomia entre o projeto de descrever e o projeto de mostrar a validade, pois ambos são complementares. Desse modo, quando Hume descreve como juízos indutivos são realizados ele estava lidando, argumenta Goodman, com a questão da validade indutiva¹³.

Essa leitura de Goodman é notadamente polêmica, mas o interessante dela é que essa característica (supostamente) humeana de colocar em relação complementar esferas aparentemente distintas, como a descritiva e a normativa, define a natureza do equilíbrio reflexivo e, como poderá ser visto ao longo desta dissertação, é a ela que se deve tanto as suas principais virtudes quanto as suas principais fraquezas. Com essa inspiração humeana, Goodman defende que a questão de definir regras indutivas para distinguir inferências indutivas válidas de inválidas requer, necessariamente, referência a um componente descritivo. Especificamente, as normas para regular o raciocínio indutivo devem ser encontradas na própria prática indutiva ordinária, e não em axiomas autoevidentes, na natureza da mente humana ou em argumentos transcendentais¹⁴.

¹³ GOODMAN, 1983, p.64-65.

¹⁴ GOODMAN, 1983, p.63.

Então sabemos que para Goodman as normas devem ser encontrada na prática, mas também sabemos que a prática nem sempre está de acordo com as normas – é por isso que buscamos normas, afinal. O próximo passo consiste em saber como se dá essa relação entre prática e norma, e como é possível extrair regras a partir de descrições. É para cumprir essa função que o ERE é inserido. Sem mais prefácios, cito a bem conhecida definição do procedimento oferecida por Goodman:

Princípios de inferência dedutiva são justificados pela sua conformidade com a prática dedutiva aceita. A sua validade depende da concordância com as inferências dedutivas particulares que nós realmente fazemos e sancionamos. Se uma regra gera inferências inaceitáveis, nós a descartamos como inválida. Assim, a justificação de regras gerais deriva do juízo que rejeita ou aceita inferências dedutivas particulares.

Isso parece flagrantemente circular. Eu disse que inferências dedutivas são justificadas pela sua conformidade com regras gerais válidas, e que regras gerais são justificadas pela sua conformidade com inferências válidas. Mas esse círculo é virtuoso. O ponto é que tanto regras quanto inferências particulares são justificadas por estar em acordo uma com a outra. *Uma regra é corrigida se ela gera uma inferência que nós estamos indispostos a aceitar; uma inferência é rejeitada se ela viola uma regra que nós estamos indispostos a corrigir.* É um processo delicado de justificação que consiste em fazer ajustes mútuos entre regras e inferências aceitas; e no acordo alcançado reside a única justificação necessária para ambas¹⁵.

Os dois parágrafos reproduzidos acima são absurdamente compreensivos. Não é exagero dizer que tudo o que será discutido ao longo desta dissertação poderá ser remetido de alguma maneira a eles. No momento o que basta verificar é que para Goodman esse processo de ajuste mútuo se aplica igualmente ao raciocínio indutivo. Uma inferência indutiva é justificada pela sua conformidade a regras indutivas válidas, e regras indutivas são justificadas em virtude da sua conformidade a inferências indutivas. O que Goodman está dizendo é que a regra válida, para usar o seu vocabulário, “codifica” adequadamente a prática indutiva¹⁶, mas que pelo processo de ajuste mútuo, ou seja, pelo ERE, essa codificação não é uma relação assimétrica. Parte-se da prática, formulam-se regras que a codificam, e, de posse dessas regras, volta-se à prática para revisá-la¹⁷. Nesse processo, que depois Rawls denomina de um processo

¹⁵ “Principles of deductive inference are justified by their conformity with accepted deductive practice. Their validity depends upon accordance with the particular deductive inferences we actually make and sanction. If a rule yields unacceptable inferences, we drop it as invalid. Justification of general rules thus derives from judgment rejecting or accepting particular deductive inferences.

This looks flagrantly circular. I have said that deductive inferences are justified by their conformity to valid general rules, and that general rules are justified by their conformity to valid inferences. But this circle is a virtuous one. The point is that rules and particular inferences alike are justified by being brought into agreement with each other. A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend. The process of justification is the delicate one of making mutual adjustments between rules and accept inferences; and in the agreement achieved lies the only justification needed for either” (GOODMAN, 1983, p.63-64).

¹⁶ GOODMAN, 1983, p.64.

¹⁷ A assunção fundamental é que os partícipes dessa prática indutiva raciocinam segundo regras que não estão explicitamente formuladas. Goodman expressa esse aspecto quando afirma que definir regras que discriminam

de avanços e recuos, tanto a prática quanto as regras são revisadas reciprocamente. O ponto ideal consiste em um equilíbrio entre ambas. O que é descrito é o resultado final do ERA. Aí está o elemento descritivo presente na ideia de codificação, mas também o elemento normativo, pois a regra formulada serve para regular a própria prática, que por sua vez é válida apenas quando está de acordo com regras válidas. O argumento é “flagrantemente circular”, Goodman reconhece, mas isso não é um problema.

Goodman adverte que o quadro acima é bastante geral porque apenas mostra como as regras indutivas são justificadas, ele, nas suas palavras, “limpa o ar, mas deixa muito por ser feito”¹⁸. Limpa o ar porque esclarece o que é possível esperar como resolução do problema e como essa solução deve ser formulada. O ERE esclarece que não é possível esperar pela certeza, por garantias ou por fundações *a priori* para regras, e que a resolução do problema deve proceder em um processo de ajuste mútuo entre prática e regras, processo em que nenhum dos dois lados possui um status epistêmico independente. O que resta para se fazer é formular quais seriam essas regras e como o problema é afinal resolvido. Fugindo um pouco da opção usual daqueles que abordam o ERE em *Fact, Fiction and Forecast*, pretendo no que se segue identificar brevemente a solução de Goodman porque entendo que ela fornece elementos decisivos para a caracterização do ERE.

De uma forma geral, o “problema da indução” para Goodman é o de definir regras para identificar uma relação adequada, *R*, entre a evidência, *e*, e a hipótese, *h*. Isso é usualmente expresso como o problema de formular uma teoria da confirmação que nos permita diferenciar entre uma *R(h,e)* válida de uma *R(h,e)* inválida, ou, simplesmente, que determine quando uma dada evidência confirma uma certa hipótese. Para alertar da dificuldade e da dramaticidade envolvidas nessa tarefa, Goodman formula o seu clássico paradoxo do *grue*¹⁹. Ele pede para considerar o seguinte conjunto de evidência mais hipótese:

e – todas as esmeraldas examinadas até *t* são verdes;

*h*₁ - todas as esmeraldas são verdes;

inferências indutivas válidas de inválidas se assemelha à tarefa de definir qualquer termo da linguagem com um uso estabelecido (GOODMAN, 1983, p. 66). Por exemplo, todos nós sabemos que “relâmpago” requer acento circunflexo, portanto sabemos como usar a nossa língua, ainda que nem sempre sejamos capazes de dizer que a palavra “relâmpago” deve ser acentuada porque é uma proparoxítona, e todas proparoxítonas têm de ser acentuadas. O ERE codifica a prática quando formula explicitamente as regras que a regulam. E ao formular essas regras nós podemos adquirir conhecimento que não tínhamos, como quando uma nova palavra é introduzida no vocabulário, ou podemos revisar a prática, vindo a saber que a mesma regra determina que devemos acentuar “esdrúxula” e “ácaro” (supondo que já não soubéssemos disso).

¹⁸ GOODMAN, 1983, p.65.

¹⁹ GOODMAN, 1983, p.71-81.

h_1 projeta o predicado verde a esmeraldas ainda não observadas e também prediz que todas as esmeraldas encontradas depois de t serão verdes. Desde que não se busque por garantias, não há até aqui nenhuma dificuldade. Goodman então inventa um predicado novo, o predicado *grue*. *Grue* se aplica a todas as coisas observadas até t no caso de serem verdes, e a todas as outras coisas no caso de serem azuis (como coisas observadas após t). Se uma esmeralda é verde, então ela também é *grue*. Assim a evidência e acima confirma tanto que todas as esmeraldas observadas são verdes, confirma h_1 , quanto confirma uma segunda hipótese:

h_2 - todas as esmeraldas são *grue*.

Assim como h_1 , h_2 projeta o predicado *grue* a esmeraldas ainda não observadas e prediz que todas as esmeraldas observadas depois de t serão *grue*. Eis o problema: h_1 e h_2 são projeções mutuamente exclusivas. Dada a definição de *grue*, se após t uma esmeralda encontrada é *grue*, então ela é azul, e não verde. Dispomos de uma mesma evidência que oferece igual suporte para duas projeções contraditórias. É óbvio que h_2 é uma projeção inadequada a partir de e , nós intuitivamente sabemos isso, mas não é óbvio por quê. A teoria de projeção de Goodman busca oferecer uma regra que permita determinar por que h_1 é a única inferência legítima a partir de e . Se o problema tradicional da indução, o problema formulado por Hume, questionava por que uma evidência e que confirma uma hipótese h fornece razões para predizer h , o novo problema é mais fundamental porque quer saber quando e confirma h . Goodman o denomina de *the new riddle of induction*²⁰. Diferentemente do velho problema da indução, o novo possui solução. Passemos a ela.

Quando nós precisamos avaliar uma relação $R(h,e)$ nós não enfrentamos o problema de “cabeça vazia, mas com algum estoque de conhecimento” que pode ser utilizado para determinar ou não a sua validade²¹. Como especifica o ERE, isto implica dizer que nós devemos buscar a solução na prática, precisamente nas hipóteses precedentes que foram de fato projetadas. Com alguma simplificação, pode-se dizer que uma hipótese é projetada quando um

²⁰ Hume teria, segundo Goodman, sido incapaz de identificar o problema confirmatório porque ele não percebeu que nem todas as regularidades geram um hábito indutivo (GOODMAN, 1983, p.81). Por exemplo, a regularidade “o sol sempre nasceu” efetivamente gera um hábito indutivo que nos leva a projetar a hipótese “o sol nascerá amanhã”; por outro lado, a regularidade “toda a vez que eu ouvi a voz do Frank Sinatra ele estava cantando” não nos leva a generalizar que “o Frank Sinatra apenas fala quando canta”. Com o ERE Goodman julga ser possível mostrar por que o primeiro raciocínio indutivo é válido (ele chama de *law-like*) e o segundo é inválido (ou é uma generalização acidental).

²¹ GOODMAN, 1983, p.86.

indivíduo a vê como “suficientemente mais credível (*credible*) que hipóteses alternativas”²², em que o credível denota a confiabilidade da hipótese do ponto de vista de quem a adota. Assim, podemos definir que h_x é credível para um sujeito, S , se e somente se S acredita firmemente na validade de h_x . É preciso destacar que o material para o ERE são projeções realmente feitas, não projeções verdadeiras ou projeções que deveriam ser realizadas²³. Nesse sentido S não precisa ter certeza da verdade de h_x , mas apenas acreditar que, em t , h_x é mais confiável que uma hipótese alternativa, h_y . Isso significa, outras coisas sendo iguais, que h_x se ajusta melhor ao sistema de crenças de S , em t , que h_y ; ou que a projeção de h_x é mais coerente com outras projeções que S sente confiança em fazer em t . Note que a projeção de h_x por S em t , a despeito do fato de h_x ser uma projeção coerente com outras projeções que S sente confiança em fazer em t , não implica a validade de h_x . Não implica porque Goodman afirma que devemos buscar a validade na prática da comunidade, não apenas no raciocínio indutivo de S . Se assim como S outros indivíduos dessa prática projetarem h_x em t , então h_x será uma projeção válida e justificada em t para C (comunidade)²⁴.

Voltemos às esmeradas. A solução de Goodman consiste em recorrer à regra ou ao princípio do entrincheiramento (*entrenchment*). Para arbitrar entre a projeção dos dois predicados (verde e *grue*) nós devemos recorrer ao registro de projeções passadas desses predicados e verificar qual dos dois é mais entrincheirado na linguagem e nesse registro²⁵. Goodman parece ter visto como uma questão de menor relevância a de oferecer uma definição precisa de entrincheirado, pois ele não a oferece. Indubitavelmente, um predicado dito bem entrincheirado é um predicado frequentemente usado nas práticas indutivas de C em t , e é um predicado que os membros de C veem como credível em t ²⁶. Embora Goodman não declare, penso que é possível introduzir, como decorrência, o elemento do consenso: um predicado é bem entrincheirado para C em t quando C consente, pelo uso, na sua projeção. Como é óbvio (se a comunidade em questão for próxima a nossa), o verde possui uma biografia mais favorável do que *grue*, e por isso a única projeção válida a partir de e é h_1 ²⁷.

²² “Something like affirmation as sufficiently more credible than alternative hypotheses” (GOODMAN, 1983, p.88).

²³ GOODMAN, 1983, p.88.

²⁴ Presumo que esse deve ser o raciocínio de Goodman porque ele define uma regra válida como uma regra com um uso bem estabelecido (GOODMAN, 1983, p. 66). Putnam enfatiza exatamente essa leitura do argumento de Goodman (PUTNAM, 1983, p.xx-xv).

²⁵ GOODMAN, 1983, p.101.

²⁶ GOODMAN, 1983, p.97.

²⁷ Estritamente falando, Goodman não defende que h_1 é válida, mas que ela é mais entrincheirada, isto é, mais válida do que a hipótese h_2 alternativa, portanto uma relação $R(e, h_1)$ é justificada quando em confronto com $R(e, h_2)$. Teller chama atenção para a necessidade de entender a validade ou a confirmação de uma hipótese em termos de níveis de projetabilidade. Assim h_1 não é projetável enquanto h_2 é não-projetável, mas h_1 é mais projetável que

Outra maneira de expressar a regra do entrincheiramento é afirmar que h_2 é inválida porque, caso ela seja aceita, implicará a rejeição de uma outra hipótese, h_1 , mais entrincheirada. Para verificar esse caso considere outro predicado, *grund*, que se aplica a todas as coisas observadas antes de t no caso de serem verdes, e a todas as outras coisas no caso de serem redondas. Agora suponha h_3 : todas as esmeraldas são *grund*. Embora não haja propriamente um conflito entre h_1 e h_3 , pois uma esmeralda pode ser verde e redonda ao mesmo tempo, h_3 conflita com outra hipótese, h_4 (“todas as esmeraldas possuem formato irregular”), bem entrincheirada e que julgamos credível²⁸. Penso que isso conduz à conclusão geral, embora Goodman não a estabeleça nesses termos, é preciso destacar, de que $R(h,e)$ é uma relação válida quando h é uma hipótese coerente com um sistema coerente composto por hipóteses bem entrincheiradas. $R(h,e)$ será inválida quando a hipótese em questão for incoerente com esse sistema. Uma hipótese bem entrincheirada é uma hipótese cuja projeção não implica a rejeição de outras hipóteses projetadas que a comunidade está indisposta a rejeitar.

Parece pertinente perguntar, adotando um raciocínio pela melhor explicação, se os predicados válidos não foram tão frequentemente projetados justamente porque são válidos, ao invés de serem válidos por terem sido projetados. Essa pergunta típica pressupõe o que o equilíbrio reflexivo nega, a saber, a existência de um critério de validade externo à prática indutiva a partir do qual seja possível determinar a sua validade. Como Goodman elegantemente afirma, os predicados válidos são frequentemente projetados simplesmente porque os predicados frequentemente projetados são válidos²⁹. Simplesmente a questão não lhe é relevante.

1.1.1 ERE e justificação

Em *Fact, Fiction and Forecast* Goodman pouco recorre ao uso da palavra “justificação”, presume-se, para afastar a abordagem tradicional ao problema da indução. Agora, é inegável que se uma projeção é válida, então eu posso estar, de algum modo relevante, epistemologicamente justificado em fazê-la. Não há razão para pensar que Goodman não veja a questão assim. É necessário então especificar o que significa estar justificado; em outras palavras, caracterizar a concepção de justificação que subjaz ao uso do ERE em *Fact, Fiction*

h_2 , que é menos projetável que h_1 (TELLER, 1969, p.224). Como não há garantias *a priori*, h_1 pode vir a ser projetável (válida) em t_1 e improjetável (inválida) em t_2 , mas a mudança de status é gradual relativamente ao uso efetivo de h_1 e h_2 na prática indutiva de uma comunidade. A ideia de validade comparativa é uma característica do equilíbrio reflexivo presente também em Rawls.

²⁸ GOODMAN, 1983, p.102-103.

²⁹ GOODMAN, 1983, p.98.

and Forecast. Creio que a ideia de coerência entre projeções, utilizada para afastar a projeção de predicados do tipo *grund*, indica parte da posição de Goodman.

Em *Sense and Certainty*, um ensaio de 1952, Goodman oferece elementos para que possamos delinear a sua posição epistemológica³⁰. No referido ensaio ele discute e refuta a posição segundo a qual não há espaço para dúvida entre a apresentação de um elemento na experiência e o meu juízo imediato a respeito do que é apresentado, de modo que a percepção poderia ser fonte de conhecimento direto, imediato e incorrigível. Essa pode ser lida com uma típica tese fundacionalista clássica. Contra ela Goodman afirma que qualquer enunciado, mesmo aqueles que dizem respeito à experiência imediata, pode ser rejeitado caso se mostre incompatível com um conjunto de enunciados coerente³¹. Em suas palavras, um enunciado deve ser rejeitado se “a sua retenção resultar em um sistema que não satisfaz tanto quanto possível a totalidade de afirmações apresentadas por todos os enunciados relevantes”³².

Positivamente formulada, essa tese implica que um enunciado é justificado se e somente se ele é coerente com um sistema coerente de enunciados. Essa é a tese epistemológica conhecida como holismo. Para visualizá-la com mais clareza, lembremo-nos que para Goodman as hipóteses “todas as esmeraldas são *grue*” e “todas as esmeraldas são redondas” não são projeções válidas porque elas são incompatíveis com, ou implicam a rejeição de, outras hipóteses relevantes que são mais entrincheiradas, como as hipóteses “todas as esmeraldas são verdes” e “todas as esmeraldas possuem formato irregular”. Se nós pensarmos que todas as projeções realizadas compõem um sistema coerente de projeções, podemos dizer que estou justificado na projeção que realizo se e somente se a participação dessa projeção no sistema não culmina em inconsistências no interior do sistema e maximiza a coerência interna do sistema. O holismo, pode-se dizer, estabelece uma condição necessária de justificação: a justificação de uma hipótese depende sempre do seu ajuste com um sistema coerente de projeções aceito e nenhuma hipótese é justificada independentemente da sua relação com outras hipóteses.

Agora, como se sabe, contos de fada são histórias coerentes, porém falsas. Essa é a objeção *standard* ao coerentismo, conhecida como a objeção dos sistemas alternativos. Muito poderia ser dito sobre ela, mas para os presentes propósitos basta a seguinte formulação: se a

³⁰ Ao fazer essa aproximação entre o equilíbrio reflexivo no *Fact, Fiction and Forecast* e a concepção epistemológica desenvolvida em *Sense and Certainty* eu sigo o que foi implícita ou explicitamente assumido por Stich (1988), Stich e Nisbett (1980) e Siegel (1992).

³¹ GOODMAN, 1952, p.161.

³² “A statement will be dropped if its retention – along with consequent adjustment in the interest of coherence – results in a system that does not satisfy as well as possible the totality of claims presented by all relevant statements. In the ‘search for truth’ we deal with the clamoring demands of conflicting statements by trying, so to speak, to realize the greatest happiness of the greatest number of them” (GOODMAN, 1953, p.163)

coerência é suficiente para a justificação então qualquer crença pode estar justificada, já que qualquer crença sempre coere com qualquer sistema. Como o holismo supõe que a justificação depende de uma relação entre crenças que é interna ao sistema, segue a objeção, então não haveria como arbitrar entre sistemas alternativos que são igualmente coerentes, porém conflitantes entre si³³. Goodman está ciente desse tipo de problema e por isso não está propondo uma “teoria coerentista pura”³⁴. Ele reconhece que o problema dos sistemas alternativos requer um critério epistêmico para arbitrar entre sistemas conflitantes que seja *independente* da coerência, de modo que a coerência deve ser necessária, mas não *suficiente* para a justificação. O critério suplementar que Goodman propõe é o da “credibilidade inicial” (*initial credibility*): para um sistema de crença ser justificado, algumas de suas crenças, além de internamente coerentes, têm de possuir uma credibilidade inicial baseada em outros fatores que não a sua coerência com as outras crenças do sistema³⁵. Entre sistemas alternativos, o justificado é aquele que coere com o maior número de crenças inicialmente credíveis, ou que implicar a rejeição do menor número delas.

Mas o que significa essa credibilidade inicial? Indubitavelmente significa que, outras coisas sendo iguais, um enunciado com credibilidade inicial está em uma situação epistêmica superior a de um enunciado que não possui credibilidade inicial. Com uso da palavra “inicial” nós podemos também inferir que a credibilidade de um enunciado independe de outras considerações, ou da sua relação com outros enunciados, e que essa credibilidade não conta toda a história necessária para a justificação plena do enunciado. Porém, mais do que isso não é possível extrair do breve ensaio de Goodman. Felizmente, outros autores simpáticos às ideias do *Sense and Certainty* as desenvolveram para além da formulação original. Israel Scheffler, para quem a proposta de *Sense and Certainty* pode ser estendida à justificação em geral, inclusive a justificação moral, é um desses autores³⁶. Vejo os pontos destacados por Scheffler em seu ensaio *On Justification and Commitment* como acertados para a caracterização do

³³ Alguns autores consideram que um dos graves problemas de qualquer versão do equilíbrio reflexivo é justamente a indeterminação entre sistemas alternativos (HASLETT, 1987, p.305-311; LITTLE, 1984, p.373-387). Um dos motivos dessa indeterminação seria a indeterminação das próprias escolhas de ajuste no interior do sistema. Considere um sistema, *c*, que não é coerente em razão de uma de suas crenças, *p*, ser incompatível com um subconjunto de crenças, *b*. Nesse caso há vários modos de fazer com que *c* seja coerente e justificado: pode-se revisar *p*, mantendo *b*, pode-se descartar *b* em favor de *p* ou mesmo é possível abandonar tanto *p* quanto *b*. Dependendo da escolha de ajuste que se adotar, haverá um sistema diferente e incompatível como resultado. Como o equilíbrio reflexivo não determinaria como a coerência interna deve ser alcançada e apenas especifica que *c* deve ser coerente, então há sistemas alternativos igualmente coerentes, porém incompatíveis entre si. A ideia de credibilidade inicial tem a função de ser esse critério para determinar a escolha de ajuste e, conseqüentemente, a determinação entre sistemas alternativos.

³⁴ GOODMAN, 1952, p.162.

³⁵ GOODMAN, 1952, p.163.

³⁶ SCHEFFLER, 1954, p.181.

equilíbrio reflexivo, tal como Goodman ou Rawls (mesmo em *TJ*, como defenderei no capítulo seguinte) o utilizam.

O primeiro ponto a se destacar da abordagem de Scheffler é a que a ideia de credibilidade inicial especifica um critério epistêmico “correspondentista” que um sistema justificado tem de satisfazer caso se pretenda justificado. Ele enfatiza que essa correspondência não deve ser vista como uma descrição fixa de qualquer fato da experiência, tampouco como um critério totalmente independente do sistema. Por exemplo, considere a proposição p (“Há agora uma folha de papel na minha frente”): por mais *prima facie* credível que p seja para um sujeito S , Scheffler afirma que ela pode ser substituída por uma sentença rival com menor credibilidade inicial caso haja um ganho para a credibilidade total de c (sistema ou conjunto total de crenças de S) em t para S ³⁷. Esse critério correspondentista, como definido por Goodman, afirma que um sistema está justificado quando ele maximiza a credibilidade inicial atribuída a certas crenças.

Um segundo e crucial ponto a se ter mente é que a ideia de credibilidade remete a uma perspectiva de primeira pessoa. Scheffler afirma que classificar uma sentença de acordo com a sua credibilidade inicial é classificá-la, ao mesmo tempo, de acordo com o comprometimento inicial com que ela é sustentada³⁸. Percebe-se que Scheffler entende essa ideia na primeira pessoa: a credibilidade de uma sentença é igual ao comprometimento (*commitment*) com que S sustenta a sentença. Para Scheffler S está comprometido com sentenças “em todos os níveis de generalidade”³⁹, de modo que o seu objetivo deve ser o de harmonizar todos os seus comprometimentos selecionando, a cada momento, um conjunto deles que pode servir como um padrão de justificação para julgar sistemas alternativos igualmente coerentes, porém incompatíveis entre si. Scheffler não caracteriza o que significa “comprometimento”, mas a ideia não pode ser outra senão uma compreensão, por parte de S , de que aquela sentença é *prima facie* justificada, em t . E uma crença é credível para S quando ela coere com outras crenças que S vê como credíveis.

Assumindo os pressupostos de Goodman, Scheffler equaciona credibilidade com comprometimento, assim esclarecendo que S é justificado nas projeções que realiza quando a aceitação de uma projeção particular não exige de S que ele faça modificações nos seus comprometimentos iniciais (ou quando ela exige menos modificações do que projeções rivais).

³⁷ SCHEFFLER, 1954, p.181-182.

³⁸ SCHEFFLER, 1954, p.187.

³⁹ “At various levels of generality” (SCHEFFLER, 1954, p.187). Essa é a clássica expressão utilizada por Rawls em um sentido idêntico ao aqui utilizado por Scheffler. Não há razão para supor, penso, que Rawls não estivesse familiarizado com esse ensaio, apesar de jamais tê-lo citado.

Essa é uma lei de menor ação: a justificação é uma questão de manter, tanto quanto possível, inalteradas aquelas crenças que julgamos *prima facie* justificadas. E essas crenças são aquelas que nós sustentamos com mais comprometimento ou confiança. Pode-se concluir, assim, que o equilíbrio entre prática e regras tem força justificacional para *S* porque *S* pode ver esse equilíbrio como algo que maximiza ou preserva o comprometimento inicial que ele possui com respeito a alguns dos componentes desse equilíbrio. A ideia de credibilidade inicial, portanto, é de crucial importância.

Antes de finalizar esta seção, é útil destacar duas consequências importantes do quadro acima. Em primeiro lugar, a justificação fornecida pelo ERE é relativa aos comprometimentos iniciais, logo, “os desacordos sobre comprometimentos iniciais não são racionalmente solúveis”⁴⁰. Como é razoável assumir, as pessoas diferem naquilo a que atribuem credibilidade inicial, principalmente se pensarmos em pessoas de culturas distintas. Scheffler entende que esse desacordo nas posições iniciais está além da possibilidade de adjudicação racional e conclui que por isso a ética é subjetiva, ainda que não irracional⁴¹. De acordo com esse raciocínio, o objetivo do ERE seria assegurar a racionalidade da ética e não a fundamentação da sua objetividade (se por objetividade entende-se o seu estatuto enquanto conhecimento). E se é assim, a equiparação entre credibilidade e comprometimento cumpre para Scheffler, e presumivelmente para Goodman, a função de justificar quando é racional para um sujeito particular sustentar uma dada crença, e não pode ser aplicada à justificação epistêmica que visa à garantia de conhecimento. Esse ponto é absolutamente decisivo e ele voltará a ser discutido ao longo desta dissertação.

Em segundo lugar, ao equiparar o nível de credibilidade epistêmica de uma sentença com o nível de comprometimento com que a sustentamos, Scheffler está atribuindo uma credencial epistêmica para uma sentença meramente porque é uma sentença que nós vemos como justificada. Esse movimento é visto por Scheffler e Goodman como necessário para se afastar do puro coerentismo e da objeção dos sistemas alternativos. Eles defendem que nós podemos julgar sistemas alternativos avaliando qual maximiza o comprometimento inicial que possuímos com relação a algumas de nossas crenças. Agora, essa estratégia apenas foi utilizada porque, como Scheffler afirma, o objetivo é apenas descrever o significado da ideia de

⁴⁰ SCHEFFLER, 1954, p.189. Alguém pode protestar contra esse “logo”, negando que a conclusão sugerida seja óbvia. No capítulo final desta dissertação eu avaliarei essa questão e lá argumentarei em favor dessa tese de Scheffler.

⁴¹ SCHEFFLER, 1954, p.189.

*justificação racional*⁴². O pressuposto é o de que há aqui um limite sobre o que podemos justificar seja na ética como nos outros domínios de conhecimento⁴³. Aparentemente, não devemos ser pretensiosos no uso que fazemos do termo justificação. Como eu mostrarei, o problema para Rawls no *Outline* é que ele discorda de Scheffler e Goodman neste ponto.

1.2 A aplicação à filosofia moral no *Outline*

A tarefa a que me proponho nesta seção é delinear como o ERE foi aplicado por Rawls ao problema da justificação moral em *Outline for a Decision Procedure for Ethics*⁴⁴. Na discussão precedente sobre o *Fact, Fiction and Forecast* foi visto que uma maneira de entender essa obra é vê-la como propondo um novo direcionamento para a epistemologia. Parece um escopo demasiado amplo para uma obra preocupada em lidar com um problema que nem é especificamente epistêmico, mas a solução oferecida por Goodman apenas funciona se o projeto tradicional da filosofia for desconsiderado. O mesmo pode ser dito a respeito do equilíbrio reflexivo de uma forma geral: a viabilidade da sua solução ao problema da justificação moral (independentemente do que for compreendido como “justificação”) está intimamente conectada com o afastamento de um projeto tradicional. Em *Outline* Rawls pretende seguir, *de certa forma*, esse raciocínio de Goodman.

No seu entender, na lógica indutiva o objetivo é encontrar um critério razoável que, quando dada uma teoria e a evidência empírica para ela, nos permita determinar a verdade da teoria⁴⁵. Em outras palavras, é o objetivo do *Fact, Fiction and Forecast* que eu expus em termos de especificar uma $R(h,e)$ adequada (exceto por Rawls usar aqui a ideia de “verdade”, ideia que ele infelizmente nunca define). A ética faria exatamente a mesma coisa: buscaria por princípios razoáveis a partir dos quais nós possamos especificar uma relação adequada entre, de um lado, uma linha de conduta (hipótese) e , de outro, a situação na qual ela é adotada e os interesses relevantes que ela mobiliza (evidência)⁴⁶. Em outros termos, dada a situação x , a ação y foi correta? O problema é então o de especificar uma $R(x,y)$ adequada. A questão não é a de saber se valores morais existem (realismo vs antirrealismo) ou se os juízos morais expressam emoções (cognitivismo vs não-cognitivismo) e tampouco importa a variedade de códigos morais

⁴² “Posso ser questionado como eu justifico a maximização do comprometimento. A minha resposta pode apenas ser que eu estou tentando descrever o que eu considero ser o significado da ideia de justificação racional” (SCHEFFLER, 1953, p.189).

⁴³ SCHEFFLER, 1953, p.190.

⁴⁴ Estritamente, o ERE é utilizado por Rawls na sua tese de doutoramento *A Study in the Grounds of Ethical Knowledge: Considered with Reference to Judgments on the Moral Worth of Character*, de 1950. O *Outline* é uma versão sintetizada da sua tese.

⁴⁵ *Outline*, p.2.

⁴⁶ *Outline*, p.2

(relativismo vs não-relativismo). No *Outline* Rawls acredita que a questão da objetividade moral não precisa ser colocada nesses termos tradicionais. Essa é uma convicção de Rawls que perpassa toda a sua carreira. O ponto então torna-se o seguinte: “existe um método razoável para validar e invalidar regras morais dadas ou propostas e as escolhas realizadas com base nelas?”⁴⁷. O que é interessante é que Rawls afirma que a objetividade da ética dependerá da confiabilidade desse método, e não de ela oferecer “garantias” ou a “verdade moral”. Rawls também assume que a ética ergue a mesma pretensão de conhecimento da ciência, pois as regras e princípios que regulam tanto uma quanto a outra são validadas a partir do mesmo método, a saber, a “lógica indutiva” ou simplesmente, como estou chamando, o ERE. Está claro que diferentemente de Scheffler e Goodman, Rawls está buscando, com o seu uso da noção de justificação por ERE, algum tipo de conhecimento e não apenas justificação racional.

Naturalmente, o método do *Outline* consiste em partir de inferências morais ou indutivas reais professados por indivíduos reais em circunstâncias reais. Esse componente descritivista constitui a essência de qualquer versão do equilíbrio reflexivo. Goodman tomava a totalidade de projeções realmente feitas como o seu material, sem restringir o método a projeções corretas ou razoáveis e nem a projeções realizadas por *experts* em lógica indutiva⁴⁸. Neste aspecto Rawls também adota uma posição consideravelmente distinta. Em primeiro lugar, os membros da comunidade cujos juízos serão tomados como material precisam satisfazer certas restrições; aqueles que satisfazem as condições estipuladas compõem a classe dos juízes morais competentes (*competent moral judges*). Em segundo lugar, nem todas as crenças desses juízes são tomadas em consideração; apenas os juízos morais ponderados (*considered moral judgments*) devem estar no interior do ERE. Passo a considerar essas categorias.

Rawls estipula quatro critérios que caracterizam um juiz moral competente⁴⁹. Podemos dividir esses critérios em um grupo que estabelece algumas capacidades cognitivas básicas e outro que determina algumas virtudes especificamente morais. Dentre os critérios do primeiro grupo, Rawls menciona (1) um certo grau mínimo de inteligência e um (2) conhecimento básico sobre os fatos envolvidos naquelas ações que são objeto de julgamento. Em acréscimo, Rawls defende que ele (3) deve ser razoável (*reasonable*), em que o razoável supõe quatro características: (i) disposição para usar os critérios da lógica indutiva a fim de determinar o que

⁴⁷ “The objectivity or the subjectivity of moral knowledge turns, not on the question whether ideal value entities exist or whether moral judgments are caused by emotions or whether there is a variety of moral codes the world over, but simply on the question: does there exist a reasonable method for validating and invalidating given or proposed moral rules and those decisions made on the basis of them?” (*Outline*, p.2).

⁴⁸ GOODMAN, 1983, p.87-88.

⁴⁹ *Outline*, p.2-5.

é adequado sustentar; (ii) sempre pesar razões prós e contra em cada caso sob deliberação; (iii) disposição para sempre reconsiderar a própria opinião à luz dos fatos relevantes; (iv) conhecer e reconhecer a influência que as suas particularidades emocionais, intelectuais e morais exercem no seu juízo. No grupo das propriedades morais do caráter Rawls inclui a (4) simpatia aos interesses do outro, em que a simpatia é caracterizada a partir de três traços: (i) o agente deve estar diretamente familiarizado com esses interesses por ter experienciado o bem que eles representam; (ii) quando não está diretamente familiarizado, o agente deve ter a capacidade para, via imaginação, considerar o valor desse interesses e (iii) o agente deve atribuir ao interesse do outro o mesmo peso que atribui ao seu. Em suma, um juiz moral competente é um homem intelectualmente racional e razoável e moralmente imparcial.

Depois de delimitada a classe de juízes morais competentes, o próximo passo consiste em selecionar, dentre os juízos morais desses agentes, apenas os ponderados. Rawls entende os juízos morais ponderados como resultado da deliberação em situações cotidianas reais. Um juízo ponderado deve satisfazer sete critérios⁵⁰: (1) o juízo deve ser realizado sob condições imunes à influência de possíveis consequências (como punição) e (2) e sob condições em que o agente não se beneficiará imediata e pessoalmente da sua decisão. Esses dois critérios excluem, respectivamente, o medo e a parcialidade. O juízo também deve (3) ser resultado de uma deliberação quando há um conflito real de interesses (casos hipotéticos são excluídos). O juízo tem de levar em consideração (4) os fatos relevantes em questão, assim como os argumentos de todos os concernidos. O agente moral (5) deve professar o juízo com um sentimento (*felt*) de *certitude*. Esse quinto critério nos remete diretamente à definição de hipótese projetada de Goodman, segundo a qual uma hipótese é projetada quando o indivíduo a vê como mais credível que hipóteses alternativas, em que o credível é equacionado com o comprometimento ou com a confiança com que o indivíduo crê na hipótese. O sexto critério (6) assevera que o juízo deve ser estável: ele deve ser feito sempre que as circunstâncias relevantes forem as mesmas. O último critério (7) estabelece que o juízo não deve ser determinado pelo uso sistemático e consciencioso de algum princípio ético. O objetivo do *Outline* é justamente o de descobrir quais princípios estão implícitos em juízos morais ponderados, por isso esse último critério. A ideia é que um princípio não-consciosamente expresso representa com mais acuidade o que um juiz moral competente vê como intuitivamente correto. Nesta dissertação me remeterei às restrições e às condições acima como *restrições do ponderado*.

⁵⁰ *Outline*, p.5-7.

Antes de entrar no funcionamento do ERE no *Outline* é preciso abordar um aspecto central desse processo de filtragem dos juízos morais iniciais. O ponto a se ter em mente sobre o recorte oferecido por esse processo é que ele alega que os juízos que satisfazem as restrições do ponderado estão mais próximos da correção do que juízos que não satisfazem e, como consequência, que o agente deve satisfazê-las⁵¹. Como Rawls entende que o ERE é capaz de produzir algo como conhecimento moral, tem-se como implicação que as restrições do ponderado são padrões epistêmicos e que os juízos que satisfazem esses padrões são mais confiáveis que juízos que não os satisfazem. DePaul entende que caso qualificações não sejam oferecidas, as restrições do ponderado constituem um caso de arbitrariedade epistêmica. Não é totalmente claro que juízos morais feitos em um estado emocional tempestuoso sejam menos confiáveis, de um ponto de vista epistêmico, que juízos realizados segundo as condições descritas por Rawls⁵². Entretanto, o problema maior, como destaca DePaul, é que a estipulação de critérios epistêmicos arbitrários viola a própria ideia de justificação holista. Quero me focar nesse segundo ponto.

Para relembrar, segundo o ERE uma regra é justificada pela sua conformidade com inferências aceitas, que são justificadas pela sua conformidade com regras válidas, do que se segue que não há qualquer regra, critério ou inferência cuja validade seja estabelecida independentemente da sua relação com outros elementos no interior desse sistema de regras, nem qualquer regra irrevogável. Se é nisso que consiste a justificação epistêmica, então a estipulação desses padrões, para adotar a expressão utilizada por DePaul, “peca contra o espírito coerentista do método”⁵³ em virtude de atribuir credibilidade epistêmica a certos padrões epistêmicos independentemente da sua conformidade com práticas aceitas. Ademais, com essa arbitrariedade Rawls poderia estar assumindo que o padrão da imparcialidade, por exemplo, poderia ser aceito incontestavelmente, o que seria contrário ao holismo e constituiria uma petição de princípio contra o cético moral que nega que o agente deve atribuir ao interesse do outro o mesmo peso que atribui ao seu.

No *Outline* ou em qualquer uma das suas obras Rawls não demonstra estar ciente dessa objeção ou considerá-la um obstáculo. Para ele os critérios não são arbitrariamente escolhidos posto que são estipulações bastante razoáveis para o raciocínio em geral. Essa é, porém, apenas uma aposta no caráter incontroverso, não uma justificativa da sua razoabilidade. Para DePaul esse problema pode ser solucionado considerando os critérios epistêmicos envolvidos como

⁵¹ *Outline*, p.7 expressa bem essa suposição.

⁵² Para uma extensa discussão deste ponto ver Sencerz (1986, p.76-95).

⁵³ DEPAUL, 1993, p.17.

princípios epistêmicos firmemente sustentados e implicitamente aceitos nas práticas cognitivas por aqueles que se engajam no método do equilíbrio reflexivo⁵⁴. Os padrões epistêmicos não seriam arbitrários porque refletiriam esses princípios epistêmicos sustentados com confiança.

O ponto que tem de ser destacado nessa solução é que ela só funciona se o equilíbrio reflexivo for um método para pessoas se engajarem reflexivamente a fim de alcançar um equilíbrio entre as suas crenças. Em outros termos, a solução a partir desses princípios epistêmicos supõe que o equilíbrio reflexivo é um método na primeira pessoa. Assim como DePaul eu realmente creio que essa é a melhor interpretação do equilíbrio reflexivo. Também entendo que o equilíbrio reflexivo em *TJ* é claramente um método que funciona na primeira pessoa. Todavia, receio que no *Outline* Rawls tenha construído o ERE de tal modo que ele opera na terceira pessoa, como uma espécie de ferramenta para o teórico moral fazer uso. O equilíbrio entre a prática e os princípios seria buscado pelo teórico, que seria então o observador e avaliador das crenças alheias. A noção de explicação desenvolvida por Rawls, que será abordada a seguir, confirma essa leitura. E se esse é o caso, a inclusão de princípios epistêmicos simplesmente não funcionaria. Para que ela pudesse se aplicar, o ERE precisaria ser revisto para incluir não apenas juízos morais ponderados, mas juízos ponderados em geral, como juízos epistêmicos ponderados. O teórico moral deveria ser mais do que um teórico moral, ele deveria avaliar todas as crenças que um agente professa para buscar também por uma explicação sistemática dos seus juízos ponderados não-morais. Por isso a solução do DePaul não se aplica totalmente ao ERE a não ser que ele seja qualificado do modo sugerido. Contudo, e este ponto é importante, ela afasta a arbitrariedade epistêmica que poderia estar contida na estipulação dos padrões ou virtudes *morais* (como a imparcialidade) porque claramente está aberta ao defensor do ERE a alegação de que a imparcialidade é boa porque essa crença é coerente com outras crenças morais também sustentadas⁵⁵.

Mas voltemos às etapas do ERE. De posse do material delimitado, a próxima etapa descrita por Rawls consiste em oferecer uma explicação (*explication*) desses juízos. O processo explicativo é entendido como um “dispositivo heurístico” cujo propósito é produzir na forma de “princípios razoáveis e justificáveis” uma caracterização da “gama total” dos juízos morais

⁵⁴ DEPAUL, 1993, p.18.

⁵⁵ É crucial reconhecer que as restrições não são apenas de natureza epistêmica, mas também são morais. É por isso que elas precisam ser justificadas *no interior* da teoria e é por isso que é completamente circular assumi-las como legitimadoras da própria teoria. Dentre outros motivos, é por não ter reconhecido este ponto que Mikhail considera a distinção entre juízos morais e juízos morais ponderados suficiente para assegurar a confiabilidade do ERE (MIKHAIL, 2011, p.247). Segundo a sua apresentação, temos de erradamente concluir que as restrições envolvidas seriam justificadas independentemente da teoria.

ponderados⁵⁶ (para usar o vocabulário de Goodman, uma explicação busca codificar juízos por meio de princípios). Enquanto que afirma que a explicação é uma “investigação empírica”⁵⁷, pois se trata de descrever um fenômeno natural, Rawls afasta a interpretação de que ela seria um estudo sociológico ou psicológico por duas razões: (i) uma explicação dos juízos morais ponderados não busca mostrar as causas desses juízos e (ii) a explicação incide sobre uma classe delimitada de juízos morais (os ponderados)⁵⁸. Rawls não diz exatamente por que (ii) afastaria o ERE de um estudo sociológico ou psicológico, mas é plausível entender que a ética faz um corte normativo no que vai contar como os seus “dados brutos” (que já não seriam brutos, portanto) enquanto a ciência não faria. O objeto de investigação da ética não seriam fatos valorativamente neutros, mas um estado ideal de fatos (os juízos morais ponderados de juízes morais competentes). Sobre (i), Rawls afirma que uma explicação tem de ser oferecida na forma de princípios. Princípios, ele nos conta, são diretivas gerais expressas em juízos morais ponderados. Os princípios fornecidos em uma explicação bem sucedida, ele também afirma, devem poder ser aceitos por um juiz moral competente. Essas duas características nos autorizam a inferir que um agente, quando adota os princípios oferecidos em uma explicação, deve poder vê-los como refletindo as máximas adotadas na sua ação ou como as *causas justificatórias* da sua ação⁵⁹. Os princípios nada dizem sobre as causas psicológicas, sociológicas ou biológicas dos juízos morais.

Um último ponto relevante sobre a ideia de explicação é que ela admite que os princípios que caracterizam o senso de certo e errado de um determinado agente não refletem a opinião desse agente. Por isso a explicação busca não por princípios expressos ou explícitos, não busca caracterizar a opinião ética de um indivíduo, mas identificar princípios que regem a sua

⁵⁶ *Outline*, p.7.

⁵⁷ *Outline*, p.8.

⁵⁸ *Outline*, p.8.

⁵⁹ Claramente nós podemos entender a ideia de causa de formas distintas. Suponha que eu pratique uma ação qualquer, como doar sangue. Por um lado é possível remeter a causa dessa ação a fatores sociológicos (a doação de sangue pode ser um ato elogiável no meu contexto social), psicológicos (eu me sinto bem comigo mesmo quando faço bem aos outros), além de outros fatores factuais em geral (como o fato de alguém estar precisando de sangue naquele momento). Por outro lado, pode também ser causa dessa minha ação a minha compreensão de que eu devo praticar ações universalizáveis. Esse outro tipo de causa se diferencia do primeiro porque ele é justificatório: é o tipo de resposta que eu daria a quem me indagasse a respeito da correção da minha ação. Mas isso é apenas uma forma elucidativa de apresentar a questão. Uma razão justificatória não é a razão que eu efetivamente ofereço para justificar a minha ação. Ela de fato pode ser, mas o princípio que regula a minha ação é identificado, pelo uso do ERE, independentemente de qualquer referência à minha própria opinião sobre a minha ação (para lembrar, a explicação dos juízos morais ponderados não é oferecida pelo próprio agente moral que os professa, mas pelo teórico moral). Podemos dizer que uma causa justificatória é uma razão (expressável na forma de um princípio) que justifica moralmente a ação do ponto de vista de quem age. Quando Rawls afirma que o objetivo da ética é oferecer uma explicação dos juízos morais ponderados, ele está justamente postulando que a ética deve descobrir, formular e sistematizar na forma de uma teoria essas causas justificatórias. O método para chegar a esses princípios é o ERE.

capacidade de formar opiniões éticas e que não são conscientemente utilizados⁶⁰. É preciso destacar que em nenhum momento Rawls afirma que esses princípios são inacessíveis ao agente. Pelo contrário, para que possam ser vistos como justificados, eles têm de mostrar “uma capacidade para serem aceitos por juízes morais competentes após eles livremente pesarem os seus méritos por meio de criticismo e discussão aberta”⁶¹. Nesta passagem Rawls parece entender que após deliberação e criticismo, o indivíduo é capaz de introspectivamente ver os princípios oferecidos como seus princípios. Consequentemente, esse acesso introspectivo via reflexão assegura que sempre que um princípio for justificado de acordo com o ERE, os indivíduos (competentes) disporão de uma razão para adotá-lo. Os princípios não precisam (nem poderiam) ser impostos coercitivamente.

O que nós temos de nos perguntar agora é por que os princípios oferecidos como explicação dos nossos juízos morais ponderados são princípios justificados. Essa pergunta é levantada pelo próprio Rawls⁶². Em resposta ele oferece quatro razões: em primeiro lugar, (i) porque eles constituem uma explicação abrangente dos juízos morais ponderados de juízes morais competentes. Como visto, isso assegura que os princípios são imunes a várias fontes de erro e, portanto, confiáveis. Em segundo lugar, (ii) porque eles podem ser aceitos livremente pelos juízes morais competentes após eles pensarem crítica e reflexivamente sobre os princípios apresentados (os princípios poderiam exibir uma capacidade para alcançar “convergência” por meios não-coercitivos). Em terceiro lugar, (iii) os princípios são justificados porque quando há um novo caso a ser decidido o resultado que eles produzem é um resultado que seria aceito por todos os juízes morais competentes concernidos (o princípio seria capaz de predizer que escolhas seriam realizadas no futuro, assim como leis científicas válidas podem “predizer” fenômenos que ainda não ocorreram). Em quarto lugar, (iv) o princípio é justificado porque, embora ele seja originado de juízos morais ponderados, ele é forte o suficiente para promover revisões em subclasses de juízos morais ponderados incompatíveis com ele⁶³. Como foi abordado em Goodman, o equilíbrio reflexivo não toma a prática indutiva ou a prática moral como incorrigível. Rawls afirma que um juízo é justificado quando ele é capaz de ser explicado por um princípio justificado⁶⁴, e como vimos um princípio é justificado quando ele é capaz de explicar um juízo moral ponderado. É isso o que está expresso na ideia de justificação por reciprocidade. Em nenhum momento no *Outline* Rawls descreve um procedimento de avanços

⁶⁰ *Outline*, p.8.

⁶¹ *Outline*, p.11.

⁶² *Outline*, p.10-11.

⁶³ *Outline*, p.11.

⁶⁴ *Outline*, p.11.

e recuos ou de ajuste mútuo (diferentemente de Goodman), mas ele tem isso claramente em mente ao postular esse “quarto teste” que um princípio justificado deve corresponder.

Agora quero destacar a crucial diferença entre Goodman e Rawls. O ponto chave da proposta de Rawls é a sua defesa de que tanto na ética quanto na lógica indutiva se parte da “pressuposição de que os homens têm capacidade para saber o que é certo e o que é errado, assim como têm capacidade para saber o que é verdadeiro e falso”⁶⁵ e, por conseguinte, o ERE especifica quando “possuímos conhecimento” e delimita uma “distinção objetiva” entre o certo e o errado a partir da caracterização, via explicação, dessa capacidade:

Se nós fizermos a assunção de que o homem tem a capacidade para conhecer o que é certo e errado, assim como ele tem a capacidade de conhecer o que é verdadeiro e falso, então o presente método é um modo de desenvolver um procedimento para determinar quando nós possuímos esse conhecimento... Assim como o desenvolvimento da ciência e o método da ciência evidenciam a capacidade para saber o que é verdadeiro e falso, assim a efetiva formulação de princípios éticos e o método por meio do qual eles podem ser testados evidenciará a capacidade para saber o que é certo e errado como também a validade da distinção objetiva entre os dois⁶⁶.

Percebe-se que ao atribuir essa condição de adequação epistêmica ao ERE Rawls se afasta da minha caracterização da proposta de Goodman e Scheffler. O raciocínio de Rawls parece ser o seguinte: a concepção de justificação por ajuste mútuo entre juízo e regra apenas poderá legitimar epistemicamente as regras se o juízo em questão for resultado da deliberação moral de uma capacidade cognitiva que, é assumido, distingue acuradamente entre o certo e o errado. Sendo assim, justificação para Rawls é conducente ao conhecimento. Como eu vejo, para Goodman (e Scheffler) o ERE não é capaz de gerar conhecimento indutivo e distinguir objetivamente entre o verdadeiro e o falso; antes, a sua finalidade seria a de oferecer um critério que permita escolher (racionalmente, eu acrescento) entre hipóteses conflitantes naquelas situações de indeterminação em que mais de uma hipótese é confirmada pela mesma evidência. O objetivo de Goodman com o ERE é assegurar a racionalidade da escolha realizada nessas circunstâncias ou oferecer alguma orientação nesses casos⁶⁷. Por outro lado, Rawls quer mais

⁶⁵ *Outline*, p.12.

⁶⁶ “If we make the assumption that men have a capacity for knowing right and wrong, as they have for knowing what is true and false, then the present method is likely way of developing a procedure for determining when we possess that knowledge...just as the development of science and the method of science evidences the capacity to know what is true and false, so the actual formulation of ethical principles and the method whereby they can be tested, as this formulation is shown in the existence of satisfactory and reasonable explications, will evidence the capacity to know what is right and wrong as well as the validity of the objective distinction between the two” (*Outline*, p.12-13).

⁶⁷ De acordo com sua a formulação do problema da indução, nós não temos como *saber* (nem probabilisticamente) se o sol surgirá amanhã. Tudo o que está ao nosso alcance é a regra que determina que a projeção de que o sol

do que isso. Ele deseja afirmar que o resultado do ERE, além de racional, pode ser confiável de um ponto de vista objetivo.

Dada a apresentação da noção de explicação, é possível ver como se configura a relação entre as esferas descritiva (explicação) e normativa (justificação) no *Outline*. Elas confluem no seguinte sentido: a solução para um problema descritivo (caracterizar a capacidade moral) acarreta também a solução para um problema normativo (saber o que conta como justificado)⁶⁸. É preciso salientar que o componente descritivo é primário e o normativo é secundário em razão da solução do problema normativo depender da solução para o problema descritivo. E, aparentemente esse é o argumento decisivo de Rawls, a solução para o problema descritivo acarreta uma solução para o problema normativo porque o que é descrito não são fatos brutos ou o simples desempenho moral de indivíduos comuns, mas a capacidade de deliberar moralmente de um juiz moral competente⁶⁹. O que deve ser destacado aqui é a relevância e a proeminência da tarefa descritiva do ERE, e o conseqüente peso das restrições do ponderado.

1.3 Duas objeções

1.3.1 Conservantismo epistêmico

No início da subseção passada foi afirmado que o movimento fundamental do ERE é o de deflacionar as exigências de justificação. Porém, essa não é uma estratégia plausível a menos que se possa oferecer alguma razão cogente pela qual, dada essa deflacionada concepção de justificação, um princípio justificado é mais confiável que um princípio alternativo não-justificado. Por que os princípios oferecidos como explicação dos nossos juízos morais ponderados são princípios justificados e confiáveis? E há razão para pensar que efetivamente o são?

Rawls aborda essa questão diretamente, oferecendo aquelas quatro razões identificadas a pouco. Porém, a sua resposta é completamente circular: ela apenas mostra que um princípio oferecido em uma explicação dos juízos morais ponderados de um juiz moral competente é justificado porque ele é um princípio oferecido em uma explicação dos juízos morais ponderados de um juiz moral competente. Rawls parece desconfortável com esse tipo de

surgirá amanhã é válida (isto é, é comparativamente mais entrincheirada que a projeção de que ele não surgirá), mas isso de nenhuma forma vem a constituir conhecimento indutivo sobre o comportamento do sol. Como ele afirma, “não devemos buscar por um conhecimento que não temos” (GOODMAN, 1983, p.64).

⁶⁸ MIKHAIL, 2011, p.27-32.

⁶⁹ “Embora a explicação seja uma investigação empírica, julga-se que ela é provavelmente um modo de encontrar princípios razoáveis e justificados *em vista da natureza da classe de juízos morais ponderados que constituem o seu alcance*”. (While explication is an empirical inquiry, it is felt that it is likely to be a way of finding reasonable and justifiable principles *in view of the nature of the class of considered moral judgments which make up its range*) (*Outline*, p.8, meu itálico).

constatação. Ao contrário de Goodman ele jamais se refere ao seu argumento como circular. Por estar preocupado apenas em assegurar a justificação racional de hipóteses projetivas, Goodman de bom grado pôde reconhecer que as projeções entrincheiradas são válidas porque projeções válidas são projeções entrincheiradas⁷⁰.

Poder-se-ia objetar que eu estou dramatizando essa circularidade. Um objetor poderia lembrar que os princípios são confiáveis porque eles incorporam as restrições do ponderado. Em acréscimo, segue o crítico, há também o uso do ERE para alcançar os princípios, o que implica que eles são o resultado de um processo de ajuste mútuo que supõe a revisibilidade. Então o crítico conclui que a minha dramatização é exagerada⁷¹. Creio que há duas razões que bloqueiam essa conclusão do crítico. Em primeiro lugar, é preciso ter em mente que a ideia básica do ERE para Rawls é a de que a capacidade moral ou senso de certo e errado de juízes morais competentes é confiável. Isso nos remete a Scheffler. Para Scheffler, a credibilidade de uma sentença particular é determinada pelo nível de comprometimento com que ela é asseverada. Da mesma forma que Scheffler, porém com um objetivo notadamente distinto, Rawls está equacionando o fato de que certas crenças são sustentadas com comprometimento (pois é isso que os juízos morais ponderados são) por juízes morais competentes com a credibilidade das mesmas. Verifica-se com clareza que esse é o caso considerando no que consiste afinal, o ERE. A definição oferecida por Daniels serve a esse propósito:

Um equilíbrio reflexivo *estrito* consiste em um par ordenado de (a) um conjunto de juízos morais ponderados aceitáveis para uma dada pessoa *P* em dado tempo, e um (b) conjunto de princípios morais gerais que economicamente sistematiza (a). O conjunto de juízos morais ponderados (a) é delimitado a partir de um conjunto de juízos iniciais em dois estágios. Primeiro ele é delimitado para eliminar juízos em que *P* não está confiante, fez sem a informação adequada sobre a situação ou fez em um estado mental conducente ao erro moral. Segundo, os juízos ponderados resultantes são posteriormente ajustados para eliminar irregularidades que podem bloquear o seu ajuste com o conjunto de princípios mais desejável. Tais princípios não devem apenas sistematizar economicamente os juízos ponderados que resultam do primeiro estágio de delimitação, mas se possível devem de alguma forma estender o conjunto de juízos ponderados aceitáveis para incluir alguns sobre os quais a pessoa não estava confiante ou via como indeterminados. O conjunto resultante (b) pode então ser considerado como caracterizando a visão moral por *P*⁷².

Como eu penso que está suficientemente claro nesta passagem, a justificação do que Daniels chama de conjunto (b) dependerá de quão bem sucedido em caracterizar a visão moral de um juiz moral competente ele for. Se o método do ERE é confiável, e Rawls afirma que a objetividade da ética dependerá da confiabilidade desse método, então é necessário oferecer um

⁷⁰ GOODMAN, 1983, p.98

⁷¹ Essas críticas podem ser encontradas em Mikhail (2011, p.228-265) e Brink (1989, p.100-143).

⁷² DANIELS, 1996, p.67.

suporte para alegação de que princípios que codificam (ou explicam) a prática são mais confiáveis que princípios que não a codificam. Em outros termos, é preciso mostrar que a adequação descritiva dos princípios implica uma adequação normativa assim entendida. O recurso às restrições do ponderado não é suficiente para esse propósito porque a legitimidade das próprias restrições está ancorada na pressuposição de que há uma capacidade moral confiável.

Em segundo lugar, a crítica acima não é válida mesmo concedendo que a capacidade humana de deliberar moralmente é confiável. O apelo às restrições do ponderado é insuficiente para responder por que um princípio oferecido em uma explicação é confiável porque as próprias restrições são justificadas em razão de serem resultado de uma explicação. Considere novamente DePaul. Para DePaul essas restrições não constituem um caso de arbitrariedade epistêmica apenas se elas forem justificadas em virtude da sua coerência com princípios epistêmicos. No caso das restrições morais (como a imparcialidade), podemos afirmar que elas são justificadas porque a crença de que a imparcialidade é boa estaria presente em uma explicação de juízos morais ponderados. Alternativamente, são justificadas porque seriam crenças que um juiz moral competente aceitaria como uma caracterização do seu senso de certo e errado após criticismo e reflexão. Sendo assim, não é possível apelar para essas restrições para legitimar o status epistêmico de uma explicação porque as próprias restrições dependem da legitimidade da noção de explicação. Se não é esse o caso, se se argumentar que as restrições são válidas independentemente de serem resultado de uma explicação, então qual a legitimidade da crença de que a imparcialidade é boa? Ela parece assim uma mera pressuposição dogmaticamente estipulada para o procedimento funcionar. Em qualquer um dos casos o apelo a essas restrições não funciona⁷³. Por que explicação implica justificação e confiabilidade?

Que o modelo holista suposto pelo ERE tem de atribuir alguma credibilidade inicial para certas crenças, e que essa credibilidade inicial confere uma credencial epistêmica para uma dada crença independentemente da sua relação com outras crenças, não são por si só alegações problemáticas. O que nós temos razão para questionar é a propriedade que diferencia uma

⁷³ Estou aqui deliberadamente evitando o expediente de refutar diretamente as restrições. Tal expediente é adotado por Sencerz. Ele analisa cada uma das restrições relevantes com o intuito de defender que elas são incapazes de credenciar epistemicamente os juízos morais ponderados. Como resultado ele rejeita o próprio método do equilíbrio reflexivo (ver SENCERZ, 1985, p.77-95). Para os meus propósitos não é necessário questionar aquelas restrições (mesmo porque parece mais na alçada das ciências ou da metodologia empírica determinar a relação entre a confiabilidade do aparato cognitivo e os variados estados de ânimo, por exemplo), bastando destacar que (i) o apelo a elas não fornece um argumento não-circular para a confiabilidade do ERE e (ii) elas contêm restrições morais que precisam ser justificadas no interior da teoria. De qualquer forma, se Sencerz está correto nas suas alegações, e eu também, então nós podemos alcançar a mesma conclusão de que o apelo aos juízos morais ponderados por si só não assegura a confiabilidade do ERE.

crença credível de uma crença não-credível⁷⁴. O ERE defende que essa propriedade é o nível de comprometimento com que a crença é sustentada pelo indivíduo. E por que uma crença sustentada com comprometimento é credível? Rawls não responde essa questão. Porém, eu imagino que ele poderia adotar a seguinte linha de raciocínio: (i) na medida em que a crença é professada com comprometimento pela comunidade, ela é aceita intersubjetivamente por todos os agentes (competentes) dessa comunidade. Por sua vez, esse aceite intersubjetivo implica que a crença está imune à influência de interesses ou pontos de vista particulares, pois é justamente por isso que ela pode ser aceita por todos, e (ii) que a crença é coerente com outras crenças também sustentadas com comprometimento, pois, em condições normais, uma crença é firmemente sustentada desde que ela não esteja em conflito com outras crenças também firmemente sustentadas⁷⁵. Então no mero fato de ser sustentada implica que a crença satisfaz tanto uma restrição de “publicidade”, pois ela é validada no uso, quanto de “coerência”.

Sobre (i), creio que ela apenas desloca a pergunta para outro nível ao invés de respondê-la. A questão que estou levantando – por que uma crença sustentada com comprometimento é credível? – é mais fundamental. Com ela eu quero saber por que nós devemos confiar nas crenças indutivas ou morais correntes na comunidade e por que uma crença validada no uso (ou pela coerência assim especificada) é mais confiável que uma crença que a comunidade não professa. Quanto à restrição de coerência (ii), creio que ela seja, no sentido acima especificado, igualmente insuficiente. Tudo o que ela diz é que p possui algum valor epistêmico positivo porque p é logicamente consistente com outras crenças, isto é, p não contradiz outras crenças ou um subconjunto de crenças. Porém, isto está longe de abranger todo o significado do conceito de coerência. Considere um sistema de crenças muito pequeno, constituído por três crenças apenas. A crença p_1 afirma que “todas as pedras são sólidas”, a crença p_2 asseve que “o Brasil é um país da América Latina” e a crença p_3 sustenta que “o filme *Beautiful Mind* é um bom filme”. Esse é um sistema coerente? Se o critério de coerência for consistência lógica, então ele é. Nenhuma crença nega ou implica a rejeição de outra, e cada uma pode ser sustentada sem contradição. Porém, esse sistema não pode, obviamente, ser considerado coerente no

⁷⁴ O ERE sofreu muitas críticas por acabar equacionando credibilidade com comprometimento. Siegel afirma que não há razão para supor que a codificação de uma prática seja suficiente para a justificação de uma regra indutiva porque a própria prática precisa ser justificada (SIEGEL, 1992, p.36). De acordo com ele o proponente do ERE não ofereceria nenhum argumento para suportar a sua alegação principal de que a codificação da prática assegura a justificação dos princípios. Para Siegel “a nossa medida de credibilidade não pode ser ela mesma meramente uma questão de nossa confiança, do que nós tomamos por ser justificado” (SIEGEL, 1992, p.43). Ele defende que nós devemos distinguir o ser justificado do ser tomado como justificado a fim de evitar uma postura acrítica. Eu subscrevo os comentários de Siegel apenas quando direcionados a Rawls.

⁷⁵ Esse argumento é utilizado por Timmons para sustentar a sua ideia de crença justificada enquanto crença responsavelmente sustentada (TIMMONS, 1999, p.206).

sentido relevante do conceito. No que consiste esse sentido relevante é uma questão complexa que não cabe aqui abordar. Para mencionar, seguindo Bonjour pode-se destacar que a relação justificatória de coerência tem de incluir, dentre outras coisas e além da consistência lógica, uma relação de explanação ou implicação inferencial. Assim, p é coerente com outras crenças se e somente se p implicar ou explicar outras crenças⁷⁶. O sistema imaginário acima não é explanatoriamente coerente.

Se esses dois possíveis argumentos não são convincentes, ainda faz sentido perguntar por que um princípio obtido por meio de uma explicação é confiável. Aparentemente, Rawls está apenas contando com a confiabilidade da própria capacidade cognitiva humana. Os críticos do equilíbrio reflexivo geralmente se apoiam nessa constatação para criticar a ideia de que o ajuste com a prática é suficiente para a justificação de regras. Eles argumentam, a partir de estudos empíricos que demonstram falhas sistemáticas no raciocínio inferencial humano, que essa pressuposição não pode ser aceita⁷⁷. Um dos estudos preferidos é o de Tversky e Kahneman (daqui pra frente TK). TK buscaram mostrar que ao avaliar a probabilidade de um evento incerto ocorrer as pessoas contam com um número limitado de “princípios heurísticos” a partir dos quais o cálculo de probabilidade é feito. Embora esses heurísticos se mostrem úteis, TK argumentaram que eles conduzem a erros sistemáticos⁷⁸. Para fins de conveniência, ignorarei esses princípios e abordarei apenas um erro sistemático dentre os diversos identificados por TK.

TK mostraram que as pessoas raciocinam, dentre outros modos, seguindo o que é conhecido por falácia do apostador. Essa falácia é recorrente nas críticas ao equilíbrio reflexivo como evidência para a inconfiabilidade do raciocínio indutivo ou projetivo (que TK identificaram ser um raciocínio basicamente probabilístico). A falácia é exemplificada recorrendo-se ao jogo de roleta, popular em cassinos. Nesse jogo uma bola é lançada em uma roleta para cair em casas vermelhas ou pretas. A probabilidade de a bola cair em uma casa vermelha ou preta é idêntica, pois o processo de lançamento da bola é completamente randômico e o número de casas vermelhas (V) e pretas (P) é o mesmo. Porém, as pessoas intuitivamente julgam que a sequência V-V-V-V-V-V é menos provável que a sequência V-P-V-P-P-V. Se a bola cair seguidamente na casa vermelha, as pessoas inferem que é (e apostam dinheiro nisso) mais provável que na próxima vez a bola cairá na casa preta. Para TK a sorte é

⁷⁶ BONJOUR, 1985, p.93-101.

⁷⁷ Siegel (1992), Stich; Nisbett (1980), Stich (1988, 1994), Harman (2005), por exemplo.

⁷⁸ Como Nisbett e Ross salientam, esses heurísticos identificados por TK não conduzem sempre ao erro. Em verdade, eles podem produzir mais juízos corretos do que juízos incorretos na maior parte dos casos. O ponto é que esses princípios são frequentemente mal utilizados, seja porque as pessoas ignoram informações relevantes ou os aplicam em situações inadequadas (NISBETT;ROSS, 1994, p.290).

vista pelo raciocínio comum como uma espécie de processo autocorretivo que busca o equilíbrio, o que é absurdo⁷⁹.

Segundo Stich e Nisbett, podemos dizer que a prática inferencial envolvida na falácia do apostador supõe o seguinte princípio: em um jogo de azar justo, a probabilidade de um dado tipo de resultado ocorrer após $n+1$ casos consecutivos de não-ocorrência é maior do que a probabilidade da sua ocorrência após n casos consecutivos de sua não-ocorrência. Para eles, se aplicarmos o equilíbrio reflexivo então esse princípio tem de ser visto como justificado, uma vez que ele codifica corretamente uma prática indutiva vista como correta⁸⁰. Stich e Nisbett alertam que não adianta afirmar que os indivíduos, caso pressionados, rejeitariam a validade desse princípio (para lembrar, a possibilidade de ser aceito é um dos requisitos que um princípio justificado tem de satisfazer para Rawls) ou que esse princípio é aceito apenas por pessoas pouco instruídas ou não suficientemente inteligentes. Eles citam passagens em manuais de lógica para ilustrar que mesmo pessoas esclarecidas e reflexivas endossam o princípio acima⁸¹. TK afirmam que erros como os envolvidos na concepção equivocada da sorte são sistematicamente cometidos mesmo por pesquisadores experientes quando eles raciocinam intuitivamente⁸² e que esses defeitos cognitivos não podem ser atribuídos a efeitos motivacionais particulares ou a outras circunstâncias do gênero, pois os erros ocorreram apesar do fato de que os voluntários tenham sido encorajados e até recompensados pelas respostas corretas⁸³. Para TK, um dos resultados mais surpreendentes a que eles chegaram foi a constatação da resistência das pessoas em aceitar o erro nos seus raciocínios ou a simples incapacidade em percebê-los e corrigi-los autonomamente ao longo da vida⁸⁴. Em um ensaio seu mais recente, Stich chama a atenção para outros experimentos que demonstram o que ele denomina de irracionalidade humana. Em um desses experimentos evidenciou-se que depois de formar um juízo com base em uma evidência, as pessoas mantêm o juízo mesmo após convencidas de que a evidência na qual ele estava baseado era falsa. Segundo Stich essas falhas apontam para uma certa irracionalidade humana e são encontradas não somente em experimentos controlados em laboratório, mas também na vida cotidiana⁸⁵.

⁷⁹ TK, 1974, p.1126.

⁸⁰ STICH;NISBETT, 1980, p.192.

⁸¹ Eles citam uma passagem de Henry Coppée's, no seu livro *Elements of Logic*. A curiosa passagem é a seguinte: "Assim, ao lançar cubos, não podemos estar certos de que lado único ou que combinação de lados surgirá; mas se em vários lances algum lado particular não surgir, as chances de sua vinda em cada lance futuro são cada vez mais fortes, até que cheguem muito próximas à certeza. Ele tem de vir; e enquanto cada lance é realizado e ele não surge, a certeza de seu surgimento a aproxima mais e mais" (STICH;NISBETT, 1980, p.193).

⁸² TK, 1974, p.1127.

⁸³ TK, 1974, p.1130.

⁸⁴ TK, 1974, p.1130-1131.

⁸⁵ STICH, 1994, p.341-342.

Se pesquisadores experientes comentem erros sistemáticos no raciocínio (não são apenas apostadores apostando que cometem a falácia do apostador) é possível perceber porque seria inútil apelar aqui para as restrições do ponderado. E se for levado em consideração o fenômeno identificado por Stich e TK de que as pessoas se mostram dogmáticas quanto a alterar o seu modo de raciocinar, nós temos uma razão adicional para desconfiar da prática inferencial ou moral corrente. Todos esses elementos – esses fatos empíricos, como os críticos preferem – apontam para a necessidade de se oferecer um argumento em favor da ideia de que a relação de reciprocidade entre prática, de um lado, e princípios, de outro, é justificatória no sentido pretendido.

Quando se requisita um argumento cogente em favor da confiabilidade da prática (tal como ela é exemplificada nos juízos morais ponderados de juízes morais competentes), aparentemente parece estar pressuposto que é possível reivindicar um critério externo a ela com o qual a tese de que ela é inconfiável poderia ser fundamentada. A premissa de que a prática poderia ser validada de um ponto de vista externo à relação de reciprocidade entre inferência/juízo e princípio é, como se sabe, rejeitada logo de início pelo ERE. Se a prática não pode ser validada, já que esse critério não existe, pela mesma razão segue-se que é impossível que ela seja invalidada. Então, seria razoável concluir que a minha questão simplesmente não pode ser posta. Creio que Jonathan L. Cohen adota uma estratégia argumentativa assim para defender o equilíbrio reflexivo. Embora o seu objetivo não fosse o de justificar regras ou princípios epistemologicamente, ele utilizou o equilíbrio reflexivo para responder a experimentos empíricos tais como os de TK e Stich que ofereceriam evidências para a tese de que o homem é irracional por não raciocinar segundo regras normativas válidas. Creio que os argumentos de Cohen são pertinentes para o presente propósito e merecem consideração.

O tópico da racionalidade ou irracionalidade humana é abrangente, contém inúmeras conexões e complicações, mas mesmo assim o raciocínio de Cohen é surpreendentemente simples. Ele parte da assunção de que uma teoria normativa da racionalidade apenas é aceitável quando “está conforme, em pontos cruciais, com a evidência da intuição não-refinada (*untutored*)”, e então esclarece que essa é a tese de Goodman em *Fact, Fiction and Forecast*. É uma tese que segundo ele precisa ser expandida e aprimorada para dar conta de objeções tais como as de TK e Stich⁸⁶. A expansão relevante consiste em uma distinção entre competência cognitiva e desempenho cognitivo (*performance-competence distinction*): erros tais como os evidenciados nos experimentos de TK apenas mostram erros de desempenho que precisam ser

⁸⁶ COHEN, 1981, p.317.

devidamente explicados, mas não autorizam inferências sobre a competência cognitiva humana⁸⁷. Por competência cognitiva Cohen entende uma habilidade cultural ou geneticamente herdada que estabelece o que as pessoas podem fazer, e não o que elas fazem ou farão⁸⁸. Assim, todo e qualquer erro evidenciado em qualquer experimento, como os experimentos de TK, apenas será um indicativo de falha de desempenho, mas não de competência. A inconfiabilidade do nosso desempenho, defende Cohen, não atesta contra a confiabilidade da nossa capacidade.

Mas nós temos de nos perguntar por que essa distinção invalida a própria possibilidade de se demonstrar uma falha de competência. Como resposta Cohen argumenta no sentido de colocar o objetor do equilíbrio reflexivo em uma espécie de dilema inescapável. Como visto ele defende que uma teoria normativa, que distingue raciocínio válido de inválido, apenas pode ter intuições como base. Caso se queira defender a irracionalidade humana é preciso assumir alguma teoria normativa como correta, isto é, é preciso afirmar que aquele comportamento cognitivo não está de acordo com critérios normativos válidos. Até aqui, tudo bem. Porém, se se assume que alguma teoria normativa determina critérios válidos, então compromete-se, com isso, com a tese de que o homem é racional. Por outro lado, se se afirma que o homem é irracional, isto é, de que suas intuições são irracionais, então não há uma teoria normativa válida com a qual justificar essa proposição. Do argumento de Cohen resulta a completa impossibilidade de se demonstrar a irracionalidade humana porque qualquer demonstração terá de contar com a própria racionalidade humana. Nas suas palavras, “nós não podemos atribuir racionalidade inferior àqueles que são eles mesmos os próprios árbitros canônicos da racionalidade”⁸⁹.

Considero o argumento de Cohen não convincente. Ele afirma que pela experiência nós não podemos extrair a tese de que os seres humanos são irracionais, mas, se é assim, como a tese oposta, a tese da racionalidade, é inferida? Qualquer intuição que Cohen reivindicar como racional será derivada da observação do desempenho cognitivo. Se o seu dilema é correto, se realmente não houver um critério outro que o ajuste entre prática e princípios para constituir uma teoria normativa, então a questão – se os homens são racionais ou irracionais – é simplesmente indeterminada. A indeterminação da questão, antes que a racionalidade humana, parece ser a conclusão mais natural do argumento de Cohen.

⁸⁷ Se a distinção entre desempenho e competência é considerada por Cohen um avanço em relação a Goodman, não se trata definitivamente de um avanço em relação a Rawls. As noções de juiz moral competente e juízo moral ponderado cumprem para Rawls o papel de mostrar que nós podemos confiar na capacidade cognitiva a despeito dos erros flagrantes que cotidianamente podem ser identificados.

⁸⁸ COHEN, 1981, p.321.

⁸⁹ COHEN, 1981, p.322.

Talvez haja aqui uma inversão do ônus da prova. Quando o crítico requisita uma razão pela qual confiar na prática (e aqui podemos estender à prática exemplificada nos juízos morais ponderados de juízes morais competentes), e consequentemente no ERE, Cohen responde exigindo uma razão satisfatória para não confiar. Mas o ponto é que a crítica, ao menos na forma que esta dissertação está apresentando, não exclui previamente que a questão normativa seja inteiramente indeterminada, isto é, que não seja possível afirmar tanto a confiabilidade da prática quanto a sua inconfiabilidade. Cohen arbitrariamente exclui essa possibilidade. É aqueles que afirmam que um princípio que a codifica é mais confiável que um princípio alternativo que não a codifica (ou que afirmam que uma crença sustentada com comprometimento é confiavelmente superior a uma crença alternativa que não é sustentada) que devem fundamentar o seu ponto de vista.

Poderia ser dito que a crítica exige demais e não compreende adequadamente o dilema proposto por Cohen. Uma das características fundamentais do ERE é que ele defende um novo programa para a epistemologia que permite, dentre outras coisas, que ela abdique da certeza, de garantias ou de fundações e reduza as exigências da justificação até o ponto em que, por exemplo, determinados argumentos circulares não são viciosos. Nesse sentido, a crítica exigiria de Cohen o que com o ERE ele defende ser impossível oferecer. Esse tipo de réplica compreende mal a objeção que vem sendo desenvolvida. A objeção não pede ao proponente do ERE que ofereça uma justificação não-circular da prática, portanto ela reconhece que o ERE está situado em um novo programa epistemológico. O que ela quer saber é: por que uma crença sustentada com comprometimento é mais confiável que uma crença que não é professada com comprometimento ou que é sequer professada? Se argumentos não-circulares para justificar a prática cognitiva não estão à disposição, como alguns defensores do ERE afirmam, isso não constitui um argumento em favor do método, apenas um argumento contra aqueles que buscam por uma justificação desse tipo.

Até aqui a objeção foi ocasionalmente dirigida ao nível particular de crenças tomadas isoladamente, o que torna as coisas um tanto quanto mais fáceis para o proponente do ERE. Invocando o holismo ele pode defender que a equação entre credibilidade e comprometimento é legítima porque o simples fato de uma crença não ser sustentada por *S* evidencia que ela é incompatível com um sistema de crenças que *S* sustenta com comprometimento. Esse é um bom movimento? Em parte sim, pois apela para o holismo. Porém, ela apenas desloca a relevância da questão do nível particular de crenças para o nível holista do conjunto total de crenças. Considerando isso a objeção pode ser facilmente alterada para questionar essa totalidade. A sua formulação pode ser expressa como então se segue: *por que o sistema de crenças que uma*

pessoa sustenta com comprometimento é mais confiável que um sistema de crenças que a pessoa não sustenta? Não vejo qualquer tipo de argumento à disposição de Rawls no *Outline* para responder a essa formulação da objeção⁹⁰.

Se eu estou certo até o momento, então estou autorizado a caracterizar a proposta do *Outline* da seguinte forma. Segundo o ERE, uma crença credível epistemicamente é uma crença sustentada com comprometimento por uma pessoa ou por uma comunidade. Essa credibilidade, embora credencie epistemicamente uma crença, não fornece toda a sua justificação. De acordo com o ERE, para uma crença credível ser justificada ela tem de ser coerente com um sistema de crenças credível, ou a sua inclusão em um sistema deve maximizar a credibilidade total desse sistema. Esse é o holismo epistemológico subentendido no ERE. Porém, um sistema de crenças credível é justificado puramente porque ele é sustentado por uma pessoa ou por uma comunidade. Essa última sentença legitima a conclusão de que o holismo defendido pelo ERE é um caso de conservantismo epistêmico (*epistemic conservatism*).

O conservantismo epistêmico é definido por Foley como se segue: a visão segundo a qual uma proposição adquire um status epistêmico favorável para uma pessoa simplesmente por ser sustentada (*believed*) por ela⁹¹. Não significa que a crença é racional ou justificada meramente porque ela é sustentada, mas que a crença tem alguma presunção de racionalidade em razão de ser sustentada⁹². O ponto essencial do conservantismo é o de que uma crença é devidamente sustentada se e somente se ela não contradiz outra crença também sustentada⁹³.

⁹⁰ Para Stich o ERE se mostra problemático uma vez que nós tenhamos em mente o fato da diversidade cognitiva entre culturas diferentes. Ele escreve: “Se diferentes pessoas têm competências cognitivas muito distintas, então a explicação de Cohen a respeito da justificação de uma teoria *normativa* do raciocínio enfrenta algum embaraço. Para recordar, em sua explicação uma teoria normativa do raciocínio é idêntica a uma teoria descritiva acerca da competência cognitiva; ambas são construídas a partir do mesmo material e idealizadas da mesma forma. Então se há muitas competências cognitivas na sociedade e em outras, há muitas teorias normativas da cognição. Mas se há muitas teorias normativas da cognição, qual é a correta?... Equilíbrio reflexivo produziria diferentes teorias normativas para diferentes pessoas, e nós ficamos com o problema de dizer qual teoria normativa é a correta” (STICH, 1994, p.353). Não quero aqui endossar toda a crítica de Stich. A meu ver ela depende de certas premissas que o proponente do ERE não precisaria aceitar. Porém, o ponto básico para o qual ela aponta (por que a teoria normativa que codifica a nossa prática é melhor do que a teoria normativa que codifica outra prática?) é análogo ao que eu venho questionando, a saber, por que um sistema de crenças sustentado com comprometimento (por nós, membros dessa comunidade) é epistemicamente superior (credível) a um sistema alternativo que nós não sustentamos (mas que outras pessoas, de outra comunidade, poderiam muito bem professar)?

⁹¹ FOLEY, 1983, p.165.

⁹² Christensen oferece a seguinte definição da ideia básica do conservantismo epistêmico: “que um agente está em alguma medida justificado em sustentar uma crença simplesmente em virtude do fato de que o agente tem aquela crença. Assim um agente pode, de acordo com o princípio conservantista, corretamente dizer ‘Eu acredito nela – e isto é parte da minha justificação para continuar acreditando nela’” (CHRISTENSEN, 1994, p.69).

⁹³ O que seguindo Foley eu denomino de conservantismo foi definido por Pollock como teoria coerentista negativa: “Há teorias coerentistas que tomam todas as proposições como *prima facie* justificadas. De acordo com essas teorias, se alguém acredita em uma proposição, p, alguém está automaticamente justificado em fazê-lo ao menos que ele tenha alguma razão para rejeitar aquela crença. De acordo com teorias desse tipo, razões funcionam primariamente de um modo negativo, levando-nos a rejeitar crenças, mas não sendo requeridas para a aquisição justificada de crenças” (POLLOCK, 1979, p.101).

Logo, uma crença sustentada possui uma presunção de racionalidade porque ela é coerente com um sistema coerente de crenças. Com base nessa restrição holista, podemos refinar a definição acima. Conforme Foley, o princípio conservantista afirma que quando uma pessoa sustenta alguma crença que não está em conflito com o sistema de crenças que ela sustenta, então independentemente da sua situação do que mais ela tem razão para sustentar, a proposição sustentada tem alguma presunção de racionalidade em seu favor para aquela pessoa. A tese central subentendida no princípio conservantista é que uma crença possui alguma presunção epistêmica em seu favor desde que ela não seja epistemicamente suspeita, isto é, desde que não haja, no interior do sistema de crenças ao qual ela pertence, um subconjunto de proposições que confirma a sua negação⁹⁴.

É preciso destacar duas características do conservantismo epistêmico. Em primeiro lugar, há uma íntima conexão entre holismo e conservantismo⁹⁵. Para Foley, o holismo é conservantista justamente porque ele utiliza as proposições que as pessoas de fato sustentam como material para determinar o que é racional endossar⁹⁶. Tal estratégia é indefensável se a lista original de proposições sustentadas não tem algum status epistêmico favorável, ou seja, se o próprio sistema não possuir um status epistêmico favorável. Esse status epistêmico seria assegurado por meio do conservantismo.

Em segundo lugar, enquanto posição epistemológica o conservantismo é geralmente utilizado para assegurar a racionalidade de um ato doxástico de sustentar uma crença em

⁹⁴ FOLEY, 1983, p.172.

⁹⁵ Para Sklar a conexão é de indissociabilidade. Se o conhecimento possui a forma holista, ele argumenta, então aceitar o conservantismo é a única forma plausível de se falar em conhecimento. O conservantismo é, nas palavras de Sklar, “transcendentalmente necessário” uma vez que se assume o pressuposto holista de que não existe um critério epistêmico último e “básico” no sentido fundacionalista de crença auto-justificada (SKLAR, 1975, p.399). Em seu ensaio Sklar defende o que ele denomina de conservantismo metodológico, porém o seu conservantismo é comparativamente mais plausível do que o que é defendido por Rawls por duas razões principais. Em primeiro lugar, Sklar oferece uma razão “utilitária” pela qual devemos dar prioridade epistêmica àquelas crenças que já sustentamos. Ele entende que é “uma perda de tempo, esforço e energia” mudar de opinião sem boas razões para tal. Em segundo lugar, Sklar sustenta a racionalidade de um sistema de crenças, não o seu estatuto enquanto conhecimento (SKLAR, 1975, p.390). Creio que devidamente compreendido o holismo não precisa implicar conservantismo, como abordarei no capítulo final desta dissertação.

⁹⁶ Considere o que Foley tem a dizer sobre o holismo e conservantismo. Entendo que o que ele oferece se aplica como uma descrição do ERE: “O ponto que eu estou elucidando aqui se aplica a qualquer posição que afirma que as proposições que são racionais para uma pessoa sustentar dependem de um processo que começa com todas as proposições que as pessoas sustentam e, de alguma forma holística, opera através desse conjunto total até alcançar um conjunto final de proposições aprovadas. Tais visões terão que responder questões tais como as seguintes: o que é tão especial nas proposições que as pessoas de fato sustentam? Por que pensar que o que é racional para uma pessoa sustentar é depende de algumas considerações holísticas aplicadas àquelas proposições?...O único modo remotamente plausível de responder essas questões, eu tenho sugerido, é aceitar um princípio de conservantismo epistêmico” (FOLEY, 1983, 1970).

situações de indeterminação da teoria (ou das projeções) pela evidência⁹⁷. Portanto, as credenciais epistêmicas atribuídas a crenças meramente sustentadas (ou da maximização de um sistema de crenças meramente sustentado, como é o caso do ERE) são credenciais de racionalidade epistêmica e não de conhecimento proposicional. Porém, Rawls utiliza o ERE para assegurar conhecimento e não apenas justificação racional. Sendo assim é preciso ter em mente que ele compreende diferentemente a ideia de credencial epistêmica atribuída a um sistema de crenças meramente sustentado. O conservantismo que eu estou objetando neste capítulo é o de Rawls⁹⁸.

A crítica de que o equilíbrio reflexivo assumiria algo como uma justificação conservantista é bastante recorrente⁹⁹. Nos anos imediatamente seguintes à publicação de *TJ* uma crítica bastante parecida foi remetida por autores como Hare, Lyons e Singer¹⁰⁰. Hare critica o fato de que a teoria da justiça seria vista por Rawls como uma “ciência empírica” por lidar com aquilo que as pessoas efetivamente dizem, com os juízos morais ponderados. Segundo

⁹⁷ Ver Christensen (1994), Foley (1983) e Sklar (1975). Christensen e Foley defendem que o conservantismo epistêmico não é um pressuposto aceitável sequer para aqueles que almejam sustentar a normatividade da ideia de racionalidade epistêmica.

⁹⁸ Princípios conservantistas podem ser mais plausíveis em outros contextos que não o de determinar o que conta como conhecimento proposicional. No tópico das teorias do raciocínio (*reasoning*) ou das teorias da revisão de crença (*belief-revision*), o tópico de investigar como as pessoas devem raciocinar quando precisam alterar suas crenças, planos ou intenções, Harman defendeu uma “teoria coerentista da revisão de crença” ou uma “teoria conservantista” (HARMAN, 1994). De acordo com essa teoria, alguém não deve revisar as suas crenças na ausência de razões especiais para alterá-las (o que Harman chama de *principle of positive undermining*); as mudanças são prescritas se e somente se elas maximizarem a coerência entre as crenças que a pessoa já sustenta. Essa teoria é conservantista, segundo Harman, porque ela prescreve que devemos manter as crenças que possuímos mesmo se não dispormos de evidência para justificá-las como razoáveis (é a “lei da menor ação” ou da “inércia psicológica” que Scheffler cita e utiliza; ver SCHEFFLER, 1954, p.187). Harman argumentou que essa teoria é imediatamente contrária à “teoria fundacionalista da revisão de crenças”, de acordo com a qual alguém deve sustentar uma crença se e somente se possuir razões para aceitar essa crença (*principle of negative undermining*). Harman rejeita a teoria fundacionalista porque, por uma questão de fato, nós somos incapazes de rastrear a justificação para as crenças que sustentamos porque não conseguimos rastrear as evidências nas quais elas estavam originalmente baseadas. Por isso se adotássemos a teoria fundacionalista deveríamos rejeitar grande parte das nossas crenças como injustificadas ou irracionais. Harman então argumenta que uma teoria coerentista é mais palatável, pois há boas “razões práticas” para não rastreamos a justificação de cada crença sustentada. O argumento parece ser o de que é, de um ponto de vista pragmático de eficiência cognitiva, simplesmente melhor raciocinar segundo as prescrições conservantistas. Essa teoria de Harman é interessante para os presentes propósitos porque ela demonstra que há (i) uma conexão íntima entre conservantismo e holismo e (ii) que o conservantismo pode ser plausível se utilizado em conjunto com considerações pragmáticas ou em outros contextos que não o da justificação do conhecimento proposicional. Além de Harman, Timmons também se vale do conservantismo epistêmico a partir de considerações pragmáticas e para defender quando uma crença pode ser sustentada responsavelmente, de um ponto de vista epistêmico (TIMMONS, 1999, p.206-208).

⁹⁹ Goodman foi caracterizado de conservantista por Harman (2005), Kvanvig (1989), Sklar (1975) e Tersman (1993).

¹⁰⁰ Embora essas críticas tenham sido direcionadas ao equilíbrio reflexivo tal como ele é apresentado em *TJ*, é seguro afirmar que os críticos não perceberam que nessa obra Rawls estava utilizando o ERA e não o ERE. Esse ponto é observado pelo próprio Rawls quando ele se defende das objeções (*IMT*, p.289) e também por Daniels (1996, p.43). Creio que essas primeiras críticas podem ser vistas como críticas contra o ERE. Se elas se aplicam ao argumento de *TJ* será discutido no capítulo seguinte. O ERA foi considerado conservantista por autores como Brandt (1998, capítulo 1) e DePaul (1993). Diferentemente das outras objeções, Brandt e DePaul apresentam críticas direcionadas contra a versão ampla do método, e por isso eu as abordarei apenas no capítulo seguinte.

ele, ao fazer isso Rawls defende um tipo de subjetivismo em que a resposta à questão “Eu estou certo no que eu digo sobre questões morais?” depende da resposta à questão “Você e eu concordamos no que eu digo?”. Para Hare, e também Singer, isso é um absurdo porque a opinião das pessoas não é decisiva para determinar a validade de uma teoria¹⁰¹. Muito poderia ser dito contra essa interpretação do equilíbrio reflexivo, mas ela parece compartilhar um dos pontos relevantes para uma caracterização conservantista: o apelo não sustentado à capacidade humana de julgar moralmente e às opiniões morais ponderadas que dela resultam. Lyons apresentou a mesma objeção. Ele duvidou da força justificatória do “argumento coerentista” de Rawls, pois argumentos “puramente coerentistas fazem-nos mover em círculos entre as nossas atitudes correntes e os princípios que elas supostamente manifestam”. Lyons pergunta por que nós não devemos ver esses princípios como comprometimentos ou sentimentos arbitrários meramente compartilhados. Como Rawls silenciaria sobre essa questão, argumenta Lyons, é legítimo inferir que ele assume ou um “convencionalismo moral complacente” ou um “intuicionismo misterioso”¹⁰².

Diferentemente das críticas acima, a presente objeção não precisa se apresentar contrariamente à concepção de justificação por codificação ou à pressuposição de que os juízos morais ponderados são, de alguma forma, autoritativos sobre aquilo que conta como correto ou justificado. O seu ponto é mais modesto: defende que o ERE deve ser rejeitado porque ele não consegue oferecer uma razão pela qual o sistema de crenças que uma pessoa sustenta com comprometimento é mais confiável que um sistema de crenças que a pessoa não sustenta, e também porque não consegue afastar a pertinência dessa pergunta. E por não oferecer essa razão, ela conclui que o ERE é insuficientemente desenvolvido.

1.3.2 Descritivismo

Agora quero voltar à crítica de Hare, Lyons e Singer para alertar para uma implicação indesejada de se admitir esse conservantismo epistêmico. Para recordar, a questão central para a avaliação do ERE consiste em saber por que a solução para o problema descritivo de caracterizar a competência moral de juízes morais competentes acarreta também uma solução para o problema normativo de saber como os princípios são justificados. Se a objeção do conservantismo é sólida, implica que Rawls não dispõe de meios para fazer a transição da descrição para a normatividade. Dessa forma, o conservantismo traz como consequência o descritivismo. A *objeção do descritivismo* é geralmente oferecida contra todas as versões do

¹⁰¹ HARE, 1989, p.82-83; SINGER, 2002, p.47.

¹⁰² LYONS, 1989, p.148-149.

equilíbrio reflexivo, mas a sua versão mais poderosa é direcionada contra o ERE. Cito a objeção feita e exemplificada por Daniels:

Em filosofia moral nós estamos interessados em entender como justificar princípios morais e escolher entre conjuntos de princípios rivais. Quando o linguista produz uma gramática que parece caracterizar uma competência sintática... a sua tarefa descritiva está feita. Ademais, ele não tem nenhuma tarefa *justificatória* e *prescritiva* de tentar mostrar por que uma gramática descritivamente (e, nós podemos supor, explanatoriamente) adequada deve ser preferida sobre outra. O equilíbrio reflexivo estreito nos deixa na posição do linguista descritivo¹⁰³.

O contexto em que Daniels apresenta essa crítica é o de *TJ*, em que Rawls define a natureza da teoria moral a partir de uma analogia com o papel de um gramático: assim como um gramático tenta caracterizar a habilidade dos nativos reconhecer sentenças corretamente compostas formulando um conjunto de princípios que façam as mesmas discriminações que o falante nativo, o teórico moral teria a tarefa de caracterizar uma capacidade moral formulando princípios que sistematize as regras implicitamente utilizadas pelas pessoas quando fazem juízos morais ponderados¹⁰⁴. Goodman também adota analogias com o papel do linguista, como já mencionado. Para Daniels essa analogia com a gramática, caso fosse levada às últimas consequências, seria incompatível com o papel normativo da filosofia moral. Por motivos que não convém tratar agora, ele argumenta que em *TJ*, apesar de ter formulado a analogia, Rawls não se compromete com essas decorrências, mas que o ERE é claramente um método apenas descritivista. E a razão pela qual o ERE é descritivista é a mesma identificada pela objeção do conservantismo: o ERE não determina por que um princípio que codifica a prática, um princípio justificado de acordo com a sua concepção de justificação, dispõe, por ser assim justificado, de alguma propriedade outra que ser apenas uma descrição adequada. Sendo assim, um princípio proposto pelo método do ERE será tão somente uma descrição da prática ou uma reprodução de comprometimentos cuja autoridade normativa está por ser assegurada¹⁰⁵. A única diferença

¹⁰³ DANIELS, 1996, p.69.

¹⁰⁴ *TJ*, p.47/rev, p.41.

¹⁰⁵ Goodman também sofreu essa crítica do descritivismo. Resnik defende que o ERE não faz jus ao caráter normativo da lógica (RESNIK, 1982, p.221-238). Também é argumentado que Goodman viola a natureza normativa da epistemologia. Por vezes se entende que o objetivo de Goodman é o de oferecer uma explicação do conceito de justificação (STICH;NISBETT, 1980, p.188-202), e então não é incomum relacionar esse projeto com o de Alvin Goldman. Para Goldman a epistemologia tem de definir o que é crença justificada tomando como metodologia uma abordagem descritivista que identifica o que é, na prática, entendido como “justificado” (GOLDMAN, 1994, p.292). Kornblith critica essa postura afirmando que ela não consegue dar conta da natureza normativa da epistemologia. Segundo ele, “independentemente do que eu quero dizer com o termo ‘justificado’, e independentemente do que eu correntemente aprovo, há uma questão substantiva para ser feita sobre por que eu deveria aprovar certos tipos de crenças, e essa questão não é respondida destacando o que eu quero dizer pelo termo ‘justificado’” (KORNBLITH, 1993, p.363).

entre um princípio justificado de um princípio não-justificado é a *adequação descritiva* de ambos. Há um salto inferencial misterioso na defesa do ERE em *Outline* (um *gap*): Rawls passa da premissa de que o princípio codifica (ou explica) a prática para a conclusão de que o princípio é justificado e/ou normativo, sem considerar que essa relação entre prática, de um lado, e regras ou princípios, de outro, precisa ser qualificada. Desse modo, enquanto é verdade que a essência de qualquer versão do equilíbrio reflexivo é apontar para a relação harmoniosa entre as esferas descritiva e normativa, segundo as objeções do conservantismo e descritivismo o ERE no *Outline* não é capaz de fazer jus a essa pretensão porque a esfera descritiva se sobrepõe à normativa, anulando-a.

2. O Equilíbrio Reflexivo Amplo

Diferentemente do ERE, uma caracterização do ERA tem de enfrentar dificuldades extras. Essas dificuldades se devem todas ao fato de que agora o equilíbrio reflexivo está imerso em uma pluralidade conceitual complexa, e é designado para operar conjuntamente com outros mecanismos (como, por exemplo, a posição original) e compor com eles uma teoria da justiça. O objetivo deste capítulo é esclarecer o que é o ERA tal como Rawls o formula em suas obras e artigos publicados, e como ele funciona no interior da sua teoria. Para tanto, o capítulo é particionado em cinco partes. Na primeira parte o método é caracterizado principalmente a partir das seções 4 e 9 de *TJ*, onde ele é pela primeira vez introduzido por Rawls. Na segunda parte é defendido que a posição original e o ERA não são métodos distintos, mas mecanismos que funcionam conjuntamente. Na terceira parte a interpretação de que o ERA é um método na primeira pessoa é defendida e uma interpretação rival, proposta por Mikhail, é rejeitada. Na quarta seção o ERA é abordado no contexto do *PL* com a finalidade de defender que ele supõe uma determinada compreensão de racionalidade prática que pode ser chamada de internalismo de razões. Por fim, o capítulo encerra com algumas notas conclusivas sobre a relação entre o ERE e o ERA. De posse de uma visão clara do método, a expectativa é que o caminho esteja livre para que no próximo capítulo seja possível finalmente caracterizar o que significa falar em justificação em ERA.

2.1 A concepção de teoria moral

Das 87 seções que compõem *TJ*, apenas a 9 é inteiramente dedicada ao equilíbrio reflexivo. Nessa seção, em que Rawls revela estar seguindo “o ponto de vista geral do *Outline*”¹⁰⁶, ele apresenta a sua concepção de teoria moral e o equilíbrio reflexivo como a técnica adequada para atingir os objetivos dessa concepção. É a isso que passarei agora.

A concepção de teoria moral de Rawls parte do pressuposto de que cada pessoa além de uma certa idade e com uma capacidade intelectual mínima desenvolve sob circunstâncias sociais normais um senso de justiça. É esse senso de justiça ou essa capacidade moral que explica como podemos proferir um potencialmente infinito número de juízos morais nas mais variadas circunstâncias. Para Rawls a teoria moral pode ser vista “provisoriamente” como a “tentativa de descrever nossa capacidade moral”, “senso de justiça” ou “sensibilidade moral”¹⁰⁷. Para exemplificar esse projeto, Rawls elabora uma polêmica analogia com a teoria linguística de Chomsky, na qual escreve que descrever um senso de justiça é como a tarefa de descrever um senso de gramaticalidade: “o objetivo é caracterizar a habilidade de reconhecer sentenças bem formadas formulando com clareza os princípios expressos que façam as mesmas discriminações que o falante nativo”¹⁰⁸.

¹⁰⁶ *TJ*, p.46n24/rev, p.40n24.

¹⁰⁷ *TJ*, p.46/rev, p.41.

¹⁰⁸ “A useful comparison here is with the problem of describing the sense of grammaticality that we have for the sentences of our native language. In this case the aim is to characterize the ability to recognize well-formed sentences by formulating clearly expressed principles which make the same discrimination as the native speaker” (*TJ*, p.47/rev, p.41).

Embora o projeto exemplificado nesse tipo de comparação não seja apresentado mais como uma “investigação empírica” (como foi no *Outline*), trata-se claramente de um programa descritivo, isto é, um “tipo de psicologia”¹⁰⁹ em que o objetivo da teoria moral é trazer à luz certos princípios mais básicos que regem a capacidade moral de deliberação. Segundo Rawls, a melhor teoria da justiça é aquela que oferece a melhor descrição dessa capacidade. Recentemente Mikhail dedicou um livro inteiro para discutir essa analogia, colocando-a no centro da proposta de Rawls em *TJ*. Na seção 2.3 discutirei a sua interpretação.

Assim como no *Outline*, a teoria moral descreveria essa capacidade moral tendo como seu “material” juízos morais ponderados. Em *TJ* não há uma lista de critérios para determinar o que conta como um juízo ponderado. Rawls faz alguns comentários gerais afirmando que devemos descartar juízos feitos com hesitação e com pouca confiança, assim como juízos realizados em estados emocionais adversos como raiva ou juízos “influenciados por uma excessiva atenção ao nosso autointeresse”¹¹⁰. Embora em *TJ* não se faça referência à categoria de agentes morais competentes, Rawls afirma que a pessoa que profere um juízo ponderado deve ter a habilidade, a oportunidade e o desejo para alcançar a decisão correta¹¹¹. De uma forma geral é mantido o mesmo posicionamento do *Outline* analisado no capítulo anterior. É importante destacar que em *TJ* Rawls mantém as características morais das restrições do ponderado (elas continuam excluindo o egoísta moral por assumirem o valor da imparcialidade como correto). No capítulo anterior busquei mostrar as sérias implicações que advém dessas restrições. Na subseção 2.4.1 voltarei a elas.

Para Rawls esse programa descritivo é classificado de “teoria moral”, uma expressão que possui um sentido técnico e é utilizada para se contrapor à “filosofia moral”. Em *The Independence of Moral Theory* Rawls conceitua a teoria moral como o estudo de noções morais básicas tais como elas são utilizadas por agentes morais nas suas vidas cotidianas. O seu objetivo é analisar como essas noções morais podem ser arranjadas para formar diferentes concepções de justiça (estruturas morais). A partir do seu estudo a teoria moral pretende arbitrar entre essas estruturas morais especificando as condições que uma concepção de justiça que se pretenda válida deve satisfazer¹¹². Diferentemente da teoria moral, a filosofia moral não investiga a moralidade substantiva para proceder nesse exame comparativo dos diversos arranjos possíveis dessa moralidade. O seu objetivo é lidar com questões de epistemologia

¹⁰⁹ *IMT*, p.290.

¹¹⁰ *TJ*, p.47/rev, p.42.

¹¹¹ *TJ*, p.47-48/ rev, p.42.

¹¹² *IMT*, p.286

moral, semântica moral e filosofia da mente. Rawls entende que a teoria moral é independente da filosofia moral e deve precedê-la metodologicamente. Para exemplificar, ele afirma que a verdade moral pode em princípio ser entendida como o resultado da convergência em ERA, mas que para investigar essa questão (questão de filosofia moral) é preciso, antes, saber se há uma concepção moral substantiva que cada agente subentende nos seus juízos morais ponderados (questão de teoria moral). Se a teoria moral realmente identificar um nível relevante de convergência, então, depois, a filosofia moral poderá analisar especulativamente à luz dos seus propósitos as implicações dessa descoberta¹¹³. Segundo Rawls, *TJ* é limitada ao âmbito da teoria moral. Voltarei a este ponto ao caracterizar a concepção epistemológica suposta pelo ERA.

Em que lugar dessa discussão metateórica entra o equilíbrio reflexivo? Na seção 9 de *TJ* ele é utilizado para explicar por que a finalidade descritiva da teoria moral é apenas “provisória”. Em primeiro lugar, é provisória porque os juízos morais ponderados, embora realizados sob circunstâncias especiais, estão sujeitos a certas “irregularidades e distorções”. Como uma questão de fato, eles são incoerentes entre si e mutuamente exclusivos. O corolário dessa incoerência é que não é realístico supor que seja possível oferecer uma caracterização (Rawls não se refere mais ao conceito de explicação do *Outline*) dos juízos morais ponderados que seja abrangente o bastante para ser compatível com todos esses juízos. Qualquer explanação oferecida precisará descartar ou revisar certos juízos morais ponderados, do que se segue a necessidade de contar com o equilíbrio reflexivo. Conforme Rawls:

Quando é apresentada a uma pessoa uma explicação intuitivamente apelante do seu senso de justiça (uma que, digamos, incorpora várias pressuposições razoáveis e naturais), ela pode muito bem revisar os seus juízos para que se conformem aos princípios dessa explicação mesmo se a teoria não se ajusta exatamente aos seus juízos atuais. É especialmente provável que ela faça isso se ela puder encontrar uma explicação para os desvios que enfraquecem a sua confiança nos seus juízos originais e se a concepção apresentada produzir juízos que ela pode agora aceitar. Do ponto de vista da filosofia moral, a melhor explicação para o senso de justiça de uma pessoa *não é* aquela que se ajusta aos seus juízos anteriores ao seu exame de qualquer concepção de justiça, mas aquela que se ajusta com os seus juízos em equilíbrio reflexivo... Esse é um estado alcançado após a pessoa ter pesado as várias concepções propostas e ter ou revisado os seus juízos para que se acordem com uma delas ou mantido as suas convicções iniciais¹¹⁴.

¹¹³ *IMT*, p.290.

¹¹⁴ “When a person is presented with an intuitively appealing account of its sense of justice (one, say, which embodies various reasonable and natural presumptions), he may well revise his judgments to conform to its principles even though the theory does not fit his existing judgments exactly. He is especially likely to do this if he can find an explanation for the deviations which undermines his confidence in his original judgments and if the conception presented yields a judgment which he finds he can now accept. From the standpoint of moral

Como a passagem acima elucida, a descrição que a teoria moral oferece não é uma investigação empírica porque ela descreve o senso de justiça que a pessoa *teria* caso ela seguisse o método do ERA. Se engajar no método significa assumir uma postura crítico-reflexiva diante de seus juízos morais ponderados à luz da descrição desse senso de justiça que o teórico moral oferece. Em larga medida a descrição do teórico moral será contraditória com muitos dos juízos morais ponderados pré-ERA, mas se a descrição for adequada, e o agente moral for razoável, o agente poderá ver que a descrição oferecida é a descrição daquilo que ele realmente sustenta em questões de justiça. Notadamente, como Goodman tornou clássico, tudo é revisável. O agente pode chegar à conclusão de que além de alguns juízos morais ponderados, a própria descrição oferecida precisa ser revista. Portanto, não se trata de uma descrição de um estado de coisas real tão-somente, mas de um estado de coisas ao mesmo tempo ideal. Rawls desenvolve uma comparação com a descrição do mundo oferecida por uma teoria física para ilustrar este ponto. Se nós não considerarmos a descrição de certos fatos naturais atraente, nós simplesmente não podemos alterar esses fatos para acomodá-los a outra teoria que nós consideramos mais sedutora¹¹⁵. A ontologia das ciências naturais é rígida, por assim dizer. Já no caso da teoria moral, a relação entre teoria-mundo é dinâmica: a teoria moral e os juízos morais ponderados podem ser revisados reciprocamente¹¹⁶. A consequência é que os nossos juízos morais ponderados constituem aquilo que contará como fato moral para a teoria. E se nós considerarmos que a relação entre a teoria e os juízos é de reciprocidade, então a teoria pode alterar os nossos juízos e, com isso, também os fatos morais. A “descrição” é ideal porque o senso de justiça sistematizado pelos princípios de justiça é alcançado após reflexão. É isso o que está em jogo quando Rawls afirma que o objetivo é descrever juízos morais ponderados *em* equilíbrio reflexivo. É uma descrição ao mesmo tempo prescritiva ou uma *descrição-ideal*.

philosophy, the best account of a person's sense of justice *is not* the one which fits his judgments prior to his examining any conception of justice, but rather the one which matches his judgments in reflective equilibrium...this state is one reached after a person has weighed various proposed conceptions and he has either revised his judgments to accord with one of them or held fast to his initial convictions” (*TJ*, p.48/rev, p.42-43. Os itálicos são meus.)

¹¹⁵ *TJ*, p.49 (excluída da edição revisada).

¹¹⁶ Realistas morais ortodoxos, como Marcus Singer, certamente rejeitarão esse modelo com o argumento de que nenhuma teoria moral pode alterar o fato de que “torturar por torturar é sempre errado” (SINGER, 1987, p.518). Para Singer quando nós alteramos os nossos juízos nós simplesmente alteramos a descrição dos fatos morais, não os próprios fatos. Contrariamente a essa visão, para Rawls não há fatos morais independentes das nossas descrições. Uma série de referências autoriza essa afirmação. A mais explícita delas é em *KC*, em que Rawls escreve que “não há fatos morais” e o que conta como fato moral é selecionado pelas partes na posição original (*KC*, p.350). Mas já em *TJ* nós podemos encontrar diversas referências apontando para a mesma conclusão, como quando é afirmando que os princípios primeiros dependem de fatos a respeito da sociedade (*TJ*, p.158-159/rev, p.137) que são (evidentemente) contingentes relativamente às escolhas morais e não-morais realizadas pelos membros dessa sociedade.

E em segundo lugar, esse programa descritivo é provisório porque o objetivo último da teoria moral é normativo. Ela não busca por princípios que refletem fatos sociais brutos, mas fatos morais especificados pelos próprios indivíduos que utilizam o ERA. A referência descritiva deve, contudo, ser mantida. É da essência do equilíbrio reflexivo partir de certos “fatos” empiricamente analisáveis do mundo. Contudo, esse descritivismo não pode ser exagerado na medida em que Rawls não supõe que haja qualquer contrapartida factual para a justiça como equidade inteira. Há, por assim dizer, uma contrapartida factual para “porções” suas. A justiça como equidade é amplamente contraditória com muitos juízos morais ponderados, mas ela consegue mostrar que aqueles juízos morais ponderados com os quais ela é compatível são mais relevantes e devem ser mantidos à custa dos juízos desviantes. A dinâmica desse processo será analisada ao longo desta dissertação.

Mas qual é a alegada diferença entre o ERE e o ERA? Eis como Rawls estabelece, em *TJ*, o que há de novo no ERA:

A noção varia dependendo se é apresentado à pessoa apenas aquelas descrições que mais ou menos se ajustam aos seus juízos correntes exceto por discrepâncias menores, ou se é apresentado à pessoa todas as descrições possíveis para as quais alguém pode plausivelmente conformar os seus juízos juntamente com todos os argumentos filosóficos em favor deles. No primeiro caso nós descreveríamos o senso de justiça de uma pessoa mais ou menos como ele é, apesar de admitir a correção de certas irregularidades; no segundo caso o senso de justiça de uma pessoa pode ou não pode passar por uma mudança radical. É claramente com o segundo tipo de equilíbrio reflexivo que alguém está interessado em filosofia moral¹¹⁷.

Apesar de não utilizar a terminologia, o segundo tipo de equilíbrio reflexivo acima é o amplo e o primeiro é o estreito¹¹⁸. Em *RH* e *JF*, nos é oferecida uma formulação um tanto mais clara dessa distinção (agora já recorrendo à terminologia):

Focando agora em uma pessoa qualquer, suponha que nós (como observadores) encontramos a concepção de justiça política que faz menos revisões nos juízos iniciais dessa pessoa e se prova ser aceitável quando a concepção é apresentada e explicada. Quando a pessoa em questão adota essa concepção e a ela alinha os seus outros juízos nós dizemos que essa pessoa

¹¹⁷ “The notion varies depending upon whether one is to be presented with the only those descriptions which more or less match one’s existing judgments except for minor discrepancies, or whether one is to be presented with all possible descriptions to which one might plausibly conform one’s judgments together with all relevant philosophical arguments for them. In the first case we would be describing a person’s sense of justice more or less as it is although allowing for the smoothing out of certain irregularities; in the second case a person’s sense of justice may or may not undergo a radical shift. Clearly it is the second kind of reflective equilibrium that one is concerned with in moral philosophy” (*TJ*, p.49/rev, p.43).

¹¹⁸ Em *IMT* (p.289) Rawls apresenta uma definição bastante parecida do ERA, dessa vez utilizando pela primeira vez a terminologia estreito e amplo e se referindo ao amplo como um método que satisfaz “certas condições de racionalidade”.

está em equilíbrio reflexivo estreito. O equilíbrio é estreito porque, enquanto convicções gerais, primeiros princípios e juízos particulares estão alinhados, nós procuramos a concepção de justiça que exigisse menos revisões para alcançar consistência, e nem concepções alternativas de justiça nem a força dos argumentos em favor dessas concepções foram considerados pela pessoa em questão. Isso sugere que nós vemos o equilíbrio reflexivo amplo (ainda no caso de uma pessoa) como o equilíbrio reflexivo alcançado quando alguém tenha cuidadosamente considerado concepções alternativas de justiça e a força dos vários argumentos em favor delas. Mais exatamente, essa pessoa considerou as principais concepções políticas de justiça encontradas na nossa tradição filosófica e pesou a força das razões filosóficas e outras razões em favor delas. Neste caso, nós supomos que convicções gerais, primeiros princípios e juízos particulares estão alinhados; mas agora o equilíbrio reflexivo é amplo, dado a amplitude de reflexão e possivelmente as muitas mudanças de visão que o precederam¹¹⁹.

Como Rawls coloca, a diferença entre o ERE e o ERA é de amplitude da reflexão. Enquanto o ERE consiste em um estado de coerência entre juízos morais ponderados de todos os níveis de generalidade, de um lado, e princípios de justiça, de outro, o ERA é o estado em que há um equilíbrio entre (i) juízos, (ii) princípios e (iii) outras descrições possíveis desse senso de justiça. Essas outras descrições são oferecidas pelas tradicionais teorias morais encontradas na história da filosofia ocidental. O argumento de Rawls é que os seus princípios de justiça são justificados não porque eles descrevem adequadamente o senso de justiça de uma pessoa, mas porque a pessoa pode ver esses princípios como a melhor descrição possível (em ERA) dentre todas as descrições alternativas possíveis. Aparentemente, o ERA vai além do ERE por incluir elementos externos aos juízos morais ponderados (a tradição filosófica).

Segundo Rawls, é por isso que o ERA é “filosoficamente motivado”, isto é, não é apenas descritivo, e é devido a esse terceiro componente que o ERA pode almejar produzir “muitas mudanças” nos juízos morais ponderados. Esse componente “externo” aos juízos morais ponderados é a primeira razão que Rawls mobiliza para mostrar que o ERA não é “conservador” e não se “limita à investigação do que as pessoas (incluindo eu mesmo) agora sustentam”¹²⁰. A

¹¹⁹ “Focusing now on any one person, suppose we (as observers) find the conception of political justice that makes the fewest revisions in that person’s initial judgments and proves to be acceptable when the conception is presented and explained. When the person in question adopts this conception and brings other judgments in line with it we say this person is in narrow reflective equilibrium. The equilibrium is narrow because, while general convictions, first principles, and particular judgments are in line, we looked for the conception of justice that called for the fewest revisions to achieve consistency, and neither alternative conceptions of justice nor the force of the various arguments for those conceptions have been taken into account by the person in question. This suggests that we regard as wide reflective equilibrium (still in the case of one person) that reflective equilibrium reached when someone has carefully considered alternative conceptions of justice and the force of various arguments for them. More exactly, this person has considered the leading conceptions of political justice found in our philosophical tradition and has weighed the force of the different philosophical and other reasons for them. In this case, we suppose this person’s general convictions, first principles and particular judgments are in line; but now the reflective equilibrium is wide, given the wide-ranging reflection and possibly many changes of view that have preceded it” (*JF*, p.30-31). Uma versão reduzida dessa mesma passagem está presente em *RH*, p.384.

¹²⁰ *IMT*, p.288.

segunda razão invocada é o fato de que a descrição pertinente para a teoria moral é do tipo descritivo-ideal: os princípios caracterizam o senso de justiça que a pessoa *afirmaria em ERA*, não o que ela afirma antes de reflexão¹²¹. No restante da discussão acerca do ERA desacreditarei essas duas razões invocadas por Rawls. Penso que elas são exageradas e que as diferenças entre o ERE e o ERA situam-se em outro nível.

Para Rawls, então, o ERA é um estado em que a pessoa chega após adotar uma atitude reflexiva diante dos seus juízos morais ponderados. Nessa atitude reflexiva, a pessoa avalia diversas concepções de justiça (dentre elas a justiça como equidade) para decidir qual delas representa com mais acuidade o seu senso de justiça. Como o seu senso de justiça pré-ERA não é provavelmente coerente, ela terá de descartar alguns juízos e revisar outros, ou mesmo modificar as teorias propostas para que se ajustem aos seus juízos, até que haja um equilíbrio entre os seus juízos morais ponderados e alguma teoria moral. Sobre esse processo, Rawls enfatiza, é “duvidoso se alguém pode alcançar esse estado”¹²². Mesmo se fosse possível definir adequadamente o que está incluindo sob o título de “todas as descrições possíveis”, não é razoável supor que alguém seja capaz de examiná-las reflexivamente. Há, portanto, uma natural indeterminação.

Como superar essa dificuldade? Embora o ERA seja em certa medida um ideal utópico inalcançável, Rawls sustenta que é possível “se aproximar dele”. O seu argumento parece ser o de que, via teoria moral, é possível interpretar o seu resultado e assim descobrir qual teoria seria aceita em ERA. Para interpretar o resultado dessa reflexão filosófica, Rawls limita as descrições possíveis às principais teorias morais conhecidas (utilitarismo, intuicionismo, perfeccionismo e a própria justiça como equidade) e provê um procedimento capaz de orientar a escolha por uma dessas concepções¹²³. Com esse apoio da teoria moral, seria possível antever o resultado da reflexão moral individual. Conforme o texto:

O máximo que nós podemos fazer é estudar as concepções de justiça conhecidas através da tradição da filosofia moral e quaisquer outras que nos ocorram... É quase isso o que eu farei, já que ao apresentar a justiça como equidade eu compararei seus princípios e argumentos com algumas outras visões familiares. À luz desses comentários, a justiça como equidade pode ser entendida como dizendo que os dois princípios previamente mencionados seriam escolhidos na posição original preferencialmente a outras concepções tradicionais de justiça... E esses princípios têm um melhor ajuste com os

¹²¹ *IMT*, p.288.

¹²² *TJ*, p.49/rev, p.43.

¹²³ *TJ*, p.50/rev, p.44.

nossos juízos ponderados em reflexão do que essas alternativas reconhecidas¹²⁴.

O que deve ser destacado da citação acima é a introdução da posição original como um procedimento auxiliar para incluir o terceiro elemento característico do ERA. O que essa citação mostra, eu devo argumentar, é que não há diferença entre afirmar que os princípios seriam escolhidos na posição original em detrimento de princípios rivais, ou afirmar que eles seriam vistos, por quem emprega o ERA, como a melhor descrição possível do seu senso de justiça. Assim a posição original funciona como um complemento ao ERA e não como uma estratégia justificacional conflitante ou paralela que o substitui.

É extremamente dúbio exatamente o que está sendo alegado na passagem acima e em que sentido a posição original é um complemento. Creio que haja duas formas não excludentes de interpretar a natureza da posição original. Por um lado, Rawls parece desenhá-la como um dispositivo teórico cuja finalidade é orientar ou ajudar aqueles que se empenham na reflexão filosófica tal como o ERA a caracteriza. Nesse sentido, qualquer pessoa que desejar refletir sobre moralidade, uma tarefa indiscutivelmente árdua para pessoas comuns (e mesmo para filósofos profissionais), poderá valer-se da posição original como um recurso assistencial. Essa interpretação é favorecida por diversas passagens. Rawls algumas vezes se preocupou em oferecer uma concepção de justiça que poderia ser efetivamente absorvida pela cultura geral de uma sociedade, e também disse que a posição original é designada para resolver dilemas morais internos a uma pessoa e para fornecer orientação (*guidance*) quando a orientação é necessária¹²⁵. Ela forneceria orientação ao ajudar “a colocar em ordem nossas convicções ponderadas sobre justiça em todos os níveis de generalidade, da mais geral a mais particular”¹²⁶, assim nos ajudando a “desenvolver o que nós agora pensamos”¹²⁷.

Por outro lado, Rawls parece estar utilizando a posição original para ele mesmo, na posição de teórico, interpretar e antever qual seria o resultado de um hipotético curso de reflexão cuja praticabilidade efetiva é duvidosa. A posição original cumpre ambos os papéis e pode ser lida de ambas as formas. Quando lida como um recurso assistencial para pessoas comuns,

¹²⁴ “The most we can do is to study the conceptions of justice known to us through the tradition of moral philosophy and any further ones that occur to us ... this is pretty much what I shall do, since in presenting justice as fairness I shall compare its principles and arguments with a few other familiar views. In light of these remarks, justice as fairness can be understood as saying that the two principles previously mentioned would be chosen in the original position in preference to other traditional conceptions of justice ... and these principles give a better match with our considered judgments on reflection than these recognized alternatives” (*TJ*, p.49-50/rev, p.43).

¹²⁵ *TJ*, p.19-20/rev, p.17.

¹²⁶ “No political conception of justice could have weight with us unless it helped to put in order our considered convictions of justice at all levels of generality, from the most general to the most particular” (*PL*, p.45).

¹²⁷ “As a device of representation the idea of the original position serves a means of public reflection and self-clarification. We can use it to help us work out what we now think” (*JPFNM*, p.402).

valoriza-se mais a especificidade metodológica de primeira pessoa do ERA, isto é, a ideia de que se apresenta à pessoa diversas concepções alternativas e então pede-se que *ela* siga os passos do ERA para averiguar qual é a concepção mais aceitável.

A inserção da posição original, se nós a compreendemos a partir da segunda leitura, se explica provavelmente por duas razões principais. Em primeiro lugar, pelo fato de que o ERA é utilizado com o intuito de oferecer como resultado uma *teoria da justiça*. Rawls não está primariamente preocupado em descrever como nós devemos deliberar moralmente e no que consiste a ideia de reflexão moral apropriada. Antes, o seu objetivo é valer-se dessas ideias apenas instrumentalmente para desenvolver uma teoria moral que especifique o que é o correto em questões de justiça a partir dos juízos morais ponderados de pessoas reais. Como espero mostrar, parte da sua tese básica em *TJ* é a seguinte: a justiça como equidade é superior às suas teorias rivais porque ela pode ser aceita em ERA. Qualquer problema de indeterminação sobre o verdadeiro significado do nosso senso de justiça, ou sobre a que resultado alguém engajado no ERA alcançaria, deve ser resolvido para que o próprio argumento seja convincente. A posição original, quando vista da segunda maneira, é um procedimento para determinar qual seria o resultado do ERA e precisamente por isso que Rawls a detalha em minúcias.

Uma segunda razão para essa segunda leitura é dada pela tentativa de Rawls de se afastar do que ele mesmo denomina de intuicionismo racional. Rawls rejeita o intuicionismo com o argumento de que ele contaria com um conjunto de primeiros princípios alegadamente básicos e não disporia de nenhuma regra ou princípio de ordenamento para ajuizar nas situações em que esses princípios conflitam entre si. Segundo Rawls, o intuicionista diz que quando há conflito entre os princípios primeiros o indivíduo deve se guiar pela sua intuição. Para Rawls quando há conflito a teoria moral deve resolvê-lo oferecendo regras de prioridade ou “critérios construtivos” (ver *TJ* seção 7). Se a hipótese de Rawls é que ele pode extrair as consequências exatas dos nossos juízos morais ponderados e do nosso engajamento reflexivo no ERA (as consequências relevantes para uma teoria preocupada com a justiça), então ele precisa de um procedimento que desvele a ordem ou a hierarquia das crenças que compõem o nosso senso de justiça. Esse seria o objetivo principal da posição original, se vista pela segunda interpretação. Quando vista pelo primeiro viés, a posição original é para Rawls também um procedimento que nos ajuda, pessoas que empregamos o ERA, a perceber qual é a ordem de prioridade e importância que concebemos a nossos juízos e, assim, como devemos resolver quando houver conflito no interior do nosso senso de justiça (a prescrição fundamental será a seguinte: devemos descartar ou revisar aquelas crenças que sustentamos com menos comprometimento

tomando como critério epistêmico aquelas que sustentamos com mais comprometimento). De posse dessa ordem, a nossa reflexão será mais bem “orientada”¹²⁸.

Ao longo desta dissertação o leitor perceberá ora referências para a primeira leitura, ora referências para a segunda. Em vários momentos será difícil precisar qual das duas posições está mais em foco. Apesar disso, não creio que essa ambiguidade represente sérias dificuldades. Ambas as interpretações são harmonizáveis e se confundem ao longo das obras de Rawls.

2.2 A relação com a posição original

Para Rawls “uma concepção de justiça é mais razoável do que outra, ou justificável com respeito a ela, se pessoas racionais em uma situação inicial escolheriam seus princípios em detrimento de outros”, de modo que as “concepções de justiça são classificadas pela sua aceitabilidade para pessoas nessas circunstâncias”¹²⁹. Rawls defende ao longo de *TJ* que a sua teoria da justiça possui a caracterização mais “filosoficamente favorecida” dessa situação inicial, por ele denominada de posição original, e que os seus princípios de justiça seriam escolhidos nessa situação. Assim o problema da justificação parece ser definido, ao menos em um primeiro momento, como um problema de escolha ou de deliberação a partir de uma situação idealizada¹³⁰. O ERA, aparentemente, não desempenharia nenhuma função.

Esse tipo de caracterização pode dar a entender, e parece que muitos interpretaram dessa forma, que a posição original é um tipo de ponto arquimediano moralmente neutro a partir do qual a razoabilidade de conflitantes teorias da justiça poderia ser acessada. O problema justificacional seria reduzido assim a um problema de escolha racional. Essa interpretação deve ser afastada. Como o próprio Rawls afirma, cada teoria da justiça tem a sua própria situação inicial e a posição original é apenas uma dentre muitas caracterizações possíveis dessa situação¹³¹ (o espectador ideal ou judicioso invocado por teorias utilitaristas é uma interpretação rival da situação inicial da justiça como equidade). Isso significa, para adotar as palavras de Dworkin, que a posição original é um dos “resultados substantivos da teoria como um todo”¹³². Por conseguinte, a justificação dos princípios depende da adequação da própria posição original, adequação essa que não pode ser assumida como dada. O fundamental não é saber

¹²⁸ *TJ*, p.19-20/rev, p.17

¹²⁹ “One conception of justice is more reasonable than another, or justifiable with respect to it, if rational persons in the initial solution would choose its principles over those of the other for the role of justice. Conceptions of justice are to be ranked by their acceptability to persons so circumstanced” (*TJ*, p.17/rev, p.16).

¹³⁰ *TJ*, p.17/rev, p.16.

¹³¹ *TJ*, p.18/rev, p.16; *TJ*, p.121/rev, p.105.

¹³² DWORKIN, 1989, p.125.

quais princípios seriam escolhidos, mas qual a interpretação da situação inicial devemos adotar e por quê.

Obviamente Rawls estava ciente dessa necessidade de justificar a posição original. Quando o ERA é pela primeira vez mencionado em *TJ*, a sua função é justamente a de justificá-la (note que na seção 4 o ERA é introduzido para complementar a descrição da situação inicial, enquanto que na seção 9 abordada na subseção anterior a posição original é introduzida para complementar o ERA). Cito a já clássica e longa passagem em que o ERA é definido:

Ao buscar a descrição mais favorecida dessa situação... nós começamos descrevendo-a de tal modo que ela representa condições preferivelmente fracas e geralmente compartilhadas. Nós então vemos se essas condições são fortes o suficiente para produzir um conjunto significativo de princípios. Se não o forem, nós buscamos por premissas adicionais igualmente razoáveis. Mas se é assim, e esses princípios se ajustam com as nossas convicções ponderadas de justiça, então até aqui tudo bem. Porém, presumivelmente haverá discrepâncias. Neste caso nós temos uma escolha. Podemos ou modificar a explicação da situação inicial ou revisar nossos juízos correntes, pois mesmo os juízos que tomamos provisoriamente como pontos fixos são suscetíveis de revisão. Vindo e indo, em algumas vezes alterando as condições das circunstâncias contratuais, em outras descartando nossos juízos e os conformando aos princípios, eu assumo que eventualmente encontraremos uma descrição da situação inicial que tanto expressa condições razoáveis quanto produz princípios que se ajustam aos nossos juízos ponderados devidamente apurados e ajustados. A esse estado de coisas eu me refiro como equilíbrio reflexivo. É um equilíbrio porque ao menos alguns dos nossos princípios e juízos coincidem, e ele é reflexivo porque nós sabemos os princípios aos quais os nossos juízos se conformam e as premissas da sua derivação¹³³.

A passagem acima esclarece que uma interpretação particular da situação inicial é aceitável quando ela está em ERA com os juízos morais ponderados. Essa sustentação da posição original com o uso do ERA é que serve como base de Rawls para argumentar que os princípios de justiça mobilizam um interesse genuinamente moral e obrigante em nós, a despeito de serem escolhidos em uma situação hipotética por agentes artificiais. A ideia de que

¹³³ “In searching for the most favored description of this situation...we begin by describing it so that it represents generally shared and preferably weak conditions. We then see if these conditions are strong enough to yield a significant set of principles. If not, we look for further premises equally reasonable. But if so, and these principles match our considered convictions of justice, then so far well and good. But presumably there will be discrepancies. In this case we have a choice. We can either modify the account of the initial situation or we can revise our existing judgments, for even the judgments we take provisionally as fixed points are liable to revision. By going back and forth, sometimes altering the conditions of the contractual circumstances, at others withdrawing our judgments and conforming them to principle, I assume that eventually we shall find a description of the initial situation that both express reasonable conditions and yields principles which match our considered judgments duly pruned and adjusted. This state of affairs I refer to as reflective equilibrium. It is an equilibrium because at last our principles and judgments coincide; and it is reflective since we know to that principles our judgments conform and the premises of their derivation” (*TJ*, p.20/rev, p.18).

a justiça como equidade mobiliza algo que já é de fato aceito é decisiva. Rawls argumenta que por se tratar de uma expressão de juízos morais ponderados devidamente ajustados, a posição original representa ou condições que nós, pessoas reais, já aceitamos ou que “aceitaríamos após reflexão filosófica”. A sua função é a de um “*dispositivo expositório* que sintetiza o significado dessas condições e nos ajuda a extrair as suas consequências”¹³⁴; ela meramente “explica os nossos juízos morais”¹³⁵ ou, como Laden se referiu, ela é meramente um “*handy way*” para articular as pressuposições que estão na base dos nossos juízos¹³⁶. Dado esse quadro, os princípios de justiça podem ser vistos como consequências daquilo que em reflexão acreditamos representar o razoável e a justiça como equidade é de uma maneira geral uma interpretação dos nossos juízos morais ponderados em ERA.

Conforme visto, a posição original é constituída de “condições” que julgamos razoáveis e que são implicitamente aceitas nos nossos juízos morais ponderados, e a justiça como equidade é apresentada como justificada em razão de ser uma descrição desses elementos. É possível refinar um pouco mais essa ideia de teoria moral. Aqueles familiarizados com o pensamento de Rawls devem se lembrar que ele frequentemente afirma que nós temos juízos em todos os níveis de generalidade¹³⁷. Por exemplo, proferimos juízos bastante concretos como “atear fogo em gatos por diversão é errado”, assim como juízos bastante complexos e sofisticados, como o juízo de que “nenhuma sociedade deve privilegiar juridicamente uma pessoa meramente pela posição social que ela ocupa”. A pressuposição de Rawls é que é possível encontrar uma sistematicidade ou um esquema coerente nessa variedade de juízos¹³⁸. Mas é possível para quem encontrar? Indiscutivelmente para o teórico moral: a justiça como equidade é a “tentativa de ver como as pessoas ajustariam seus vários juízos em um esquema coerente”, isto é, é uma tentativa de construir o resultado do uso do ERA. Porém, essa frase também assume que as próprias pessoas podem encontrar essa sistematicidade. Se elas “descartarem alguns juízos e revisarem outros”, e se elas se valerem da “persuasão intuitiva de um para avaliar a plausibilidade de outro”, ou seja, se elas seguirem o ERA, então *elas* estarão em equilíbrio reflexivo¹³⁹. Note a ambiguidade: por um lado o ERA é um método que demanda certo comportamento deliberativo e revisionista das pessoas, assim funcionando na primeira pessoa, mas, por outro, é um estado de coerência que o teórico moral pode identificar para

¹³⁴ TJ, p.21/rev, p.19.

¹³⁵ “The conception of original position is not intended to explain human conduct except insofar as it tries to account for our moral judgments and helps to explain our having a sense of justice” (TJ, p.120/rev, p.104).

¹³⁶ LADEN, 2003, p.379.

¹³⁷ IMT, p.289.

¹³⁸ FG, p.270.

¹³⁹ IMT, p.289.

formular a sua teoria, assim operando na terceira pessoa. É necessário algum cuidado aqui. Se reduzirmos o ERA ao segundo sentido, como muitos fazem, acabamos incapazes de explicar como as pessoas alcançaram aquele estado, além de precisarmos simplesmente equacionar “ERA” com “senso de justiça coerente”. Não é isso o que Rawls faz, pois ele afirma, por exemplo, que “o equilíbrio reflexivo requer somente que o agente faça as revisões com convicção e confiança, e que continue a afirmar esses princípios quando tiver de aceitar as suas consequências na prática”¹⁴⁰.

Em *TJ* Rawls não é explícito a respeito dos pressupostos da posição original, o que gerou muita discussão e desentendimento na recepção imediata da sua obra. Uma breve incursão pela coleção de ensaios que marca essa recepção, o *Reading Rawls* (1989), é suficiente para perceber esse ponto¹⁴¹. Dworkin afirmou que o argumento contratual (a posição original) deveria ser visto como uma “conclusão intermediária” cuja utilidade é apontar para uma razão de justiça que credencie os princípios resultantes independente do fato da escolha ou do acordo. A razão de justiça que Dworkin identifica como subjacente à posição original é um direito natural à

¹⁴⁰ “Reflective equilibrium requires only that the agent makes these revisions with convictions and confidence, and continues to affirm these principles when it comes to accepting their consequences in practice” (*IMT*, p.289).

¹⁴¹ Como está documentado no *Reading Rawls*, originalmente publicado em 1975, a posição original foi criticada de todas as frentes. Foi afirmado que ela não seria obrigante sobre indivíduos reais ao menos que servisse meramente para indicar que os princípios de justiça estão de acordo com um direito natural (Dworkin), que ela não se diferenciaria significativamente da situação inicial das teorias utilitaristas e é menos filosoficamente favorecida do que ela (Hare). Foi defendido que ela conta com um apelo não justificado aos valores da imparcialidade e da equidade por estar subordinada a um argumento coerentista ele mesmo infundado (Lyons). Também foi sustentado que os princípios utilitaristas é que seriam o seu real resultado (Lyons). De qualquer forma, a principal fonte de controvérsias em torno do argumento a partir da posição original foi a reconhecidamente infeliz expressão de Rawls, que se referiu à posição original como um caso de teoria da escolha racional. Muitos leitores se entusiasmarão com essa expressão, criticando e defendendo Rawls como se Rawls estivesse empenhado em alguma revitalização de Hobbes (o projeto que certos leitores julgaram ser o de Rawls foi depois levado a cabo por David Gauthier em *Morals by Agreement*). Por exemplo, Buchanan escreveu em 1975 um ensaio devotado à defesa da “teoria da escolha racional” de Rawls (BUCHANAN, 1975), enquanto Williams escreveu um ensaio para refutar a “teoria da decisão racional sob condições de incerteza” de Rawls (WILLIAMS, 1981, p.94-100). Dois anos depois da publicação de *TJ*, a revista *Ethics* publicou um artigo em que a autora, Schwartz, objeta Rawls por, dentre outras coisas, não ter provido informações suficientes para guiar a escolha racional das partes e por incluir premissas morais na posição original (Schwartz também diz que o equilíbrio reflexivo é naturalmente equivocado, já que os juízos morais ponderados de um nazista estariam em equilíbrio reflexivo com princípios de justiça nazistas. Ver SCHWARTZ, 1973, p.298. Quero destacar dois pontos equivocados dessa primeira recepção de *TJ*: em primeiro lugar, uma excessiva concentração na posição original, isto é, a redução de todo o argumento justificacional em favor dos dois princípios à posição original, e, em segundo lugar, uma equivocada compreensão da própria posição original, vista equivocadamente como um caso de escolha racional contratualista. Ao longo deste capítulo buscarei mostrar como essa leitura do projeto de Rawls (uma leitura que ainda hoje ecoa em artigos especializados) torna a justiça como equidade muito menos interessante do que ela efetivamente é. Esse foi um ponto enfatizado, sobretudo, por Rorty. Ele admite que foi um dos maus leitores iniciais de *TJ* e que, assim como muitos outros, compreendeu erradamente as assunções de racionalidade envolvidas no interior da posição original. Rorty inclusive afirma que é “tentador sugerir que a posição original poderia ser completamente excluída de *TJ* sem perdas” (RORTY, 1991, p.184). Creio que essa sugestão (de fato tentadora) é um tanto radical: a posição original não causa dano algum ao argumento justificatório de Rawls, contanto que seja bem interpretada. Uma boa interpretação da posição original deve começar reconhecendo que ela é simplesmente um movimento interno à teoria, por isso mesmo sem qualquer função de justificação. A justificação depende inteiramente do equilíbrio reflexivo.

igual consideração e respeito, direito esse que seria válido independentemente de qualquer “convenção”, “legislação” ou “contrato hipotético”¹⁴². O argumento principal de Dworkin não é simplesmente o de que a posição original realmente supõe tais restrições morais, mas que ela tem de pressupor caso aspire por força justificatória. Dworkin está correto quando defende que a posição original tem de pressupor certos valores morais (ou ideais morais) e que deve ser vista como uma conclusão intermediária para apontar para esses valores, mas nós não precisamos (nem devemos) segui-lo no tipo de valor que identifica.

Para caracterizar os pressupostos morais da posição original eu me afastarei de *TJ*, onde eles não estão estabelecidos com clareza, e recorrerei à posterior formulação construtivista que Rawls fornece à sua interpretação da situação inicial. Nos textos publicados como resposta aos críticos nos anos imediatamente seguintes à publicação de *TJ*, como *Reply to Alexander and Musgrave* (1974), *A Kantian Conception of Equality* (1975) e *Fairness to Goodness* (1975), Rawls já se preocupa em esclarecer os pressupostos da posição original, mas é somente em *Kantian Constructivism in Moral Theory* (1980) que eles são completamente elucidados¹⁴³.

Nas três conferências de *KC* a intenção é esclarecer “as raízes kantianas da justiça como equidade”, um tema que Rawls reconhece não ter sido suficientemente abordado em *TJ*. Pelo adjetivo “kantiana” ele adverte que está exprimindo uma relação de semelhança e aproximação, e não de identidade, pois o seu objetivo é “superar os dualismos da doutrina de Kant”¹⁴⁴ e

¹⁴² DWORKIN, 1989, p.26; 46; 50.

¹⁴³ Cumpre aqui uma nota: a seção 23 de *TJ* é reservada à apresentação das cinco “restrições formais do conceito de correto” (*formal constraints of the concept of right*) que moldam parte da posição original. Todavia, essas restrições não contam a história toda. Rawls entende que todas as concepções de justiça tradicionais as satisfazem (o espectador ideal também as incluiria), mas apenas a justiça como equidade seria escolhida na posição original. Assim, tem de haver outras restrições que determinam a escolha dos dois princípios propostos por Rawls. As cinco restrições formais são as seguintes: (i) generalidade, (ii) universalidade, (iii) publicidade, (iv) ordenamento, (v) finalidade (*TJ*, p.130-136/rev, p.112-118). Essas restrições, apesar de formais, não são *a priori* ou analíticas, no sentido de constituir o significado do conceito de correto ou de moralidade. Para Rawls o mérito dessas restrições formais (que reconhecidamente são morais) depende da adequabilidade (*soundness*) da teoria que resulta delas. No caso da justiça como equidade, essas restrições são autorizadas porque elas podem ser endossadas em ERA juntamente com a teoria como um todo. É absolutamente imperioso e significativo que haja essa justificação holista para as restrições. Embora as restrições sejam bastante fracas, e aceitas por todas as tradicionais teorias morais (por isso que elas são denominadas de “formais”), elas excluem o egoísmo moral. E o egoísmo moral, seja visto como uma teoria moral ou como uma forma de ceticismo moral, deve ser respondido. Portanto quando Rawls exclui o egoísmo da lista de teorias morais com as quais ele comparará a justiça como equidade, ele não está, em certa medida, simplesmente rejeitando discutir com o egoísta, mas argumentando que a sua teoria contém pressupostos que não seriam afirmados em ERA.

¹⁴⁴ Em *KCE* Rawls esclarece os dualismos que ele tem em mente: entre o necessário e o contingente, a forma e o conteúdo, a razão e o desejo e o noumeno e o fenômeno. É importante destacar que o equilíbrio reflexivo é, de uma forma geral, comprometido com a superação de dicotomias. A mais significativa delas é a dicotomia descrição vs normatividade que Goodman supera ao apontar para a dimensão normativa do tratamento descritivo que Hume deu à questão da indução. Rawls em *TJ* também busca superar essas dicotomias porque ao mesmo tempo em que ele propõe uma teoria normativa, ele define a teoria moral como um tipo de psicologia que busca descrever o nosso senso de justiça. O papel do equilíbrio reflexivo é justamente o de compatibilizar esferas aparentemente inconciliáveis na busca por padrões epistêmicos efetivos que não tenham comprometimentos metafísicos

reinterpretá-la no “interior do escopo de uma teoria empírica”¹⁴⁵. Para Rawls, o construtivismo é caracterizado por especificar duas “concepções-modelo” (*model-conceptions*), uma concepção de pessoa enquanto livre e igual e uma concepção de sociedade bem-ordenada, e estipular um procedimento de construção (posição original) que conecta essas concepções-modelo com um conjunto definido de princípios de justiça¹⁴⁶. Ao oportunizar esse tipo de combinação, a teoria “articula e torna explícitas aquelas noções compartilhadas” e “latentes no senso comum” ou se o senso comum é “hesitante e não sabe o que pensar”, como é frequentemente o caso, propõe certos “princípios afins com as suas convicções mais essenciais e tradições históricas” que constituem “uma profunda base de acordo”¹⁴⁷. Assim, o propósito do construtivismo é “invocar uma concepção de pessoa implicitamente afirmada na cultura” ou que poderia ser aceitável, quando devidamente explicada e apresentada, para cada partícipe dessa cultura. Como resultado, o construtivismo oferece uma que cada um pode ver (em ERA) como congruente com o seu profundo entendimento de si mesmo¹⁴⁸.

Como essa última frase sugere, as concepções de pessoa e sociedade nada mais são do que compreensões morais que cada indivíduo possui acerca de si mesmo. Rawls esclarece que essas concepções são extraídas daqueles juízos morais ponderados ou daquelas convicções que são mais “essenciais” e “profundas”. É plausível conjecturar que as crenças mais profundas moldam a imagem que nós temos de nós mesmos e dos outros, e assim orientam e restringem o nosso raciocínio prático. A compreensão de Rawls parece ser a de que essas concepções são firmes convicções que julgamos prioritárias sobre as nossas outras convicções, convicções que não estamos dispostos a abandonar, a revisar ou a admitir qualquer tipo de maleabilidade. Por de certa forma constituírem a nossa identidade moral, descartá-las implicaria também o abandono do *nosso* próprio modo de raciocinar praticamente. Em *TJ* Rawls afirma:

Há questões que nós sentimos com certeza que têm de ser respondidas de uma certa forma. Por exemplo, nós estamos confiantes de que a intolerância religiosa e a discriminação raciais são injustas... essas convicções são pontos fixos provisórios que nós presumimos que qualquer concepção de justiça tem

indesejáveis, como tem a filosofia kantiana segundo Rawls. Silveira recentemente se preocupou em enfatizar essa aversão do ERA às dicotomias (SILVEIRA, 2011).

¹⁴⁵ *KC*, p.303-304 e *KCE*, p.264. A expressão “uma teoria empírica” não estava presente na edição original de *TJ*, mas foi cuidadosamente inserida por Rawls na edição revisada (*TJ*, rev, p.226).

¹⁴⁶ *KC* p.304-308.

¹⁴⁷ “The aim of political philosophy, when it presents itself in the public culture of a democratic society, is to articulate and to make explicit those shared notions and principles thought to be already latent in common sense; or, as is often the case, if common sense is hesitant and uncertain, and doesn’t know what to think, to propose to it certain conceptions and principles congenial to its most essential convictions and historical traditions” (*KC*, p.306).

¹⁴⁸ *KC*, p.307.

de se ajustar. Mas nós temos muito menos segurança sobre qual é a distribuição correta da renda e da autoridade. Aqui nós podemos buscar por um modo de remover nossas dúvidas. Nós podemos checar uma interpretação da situação inicial, então, pela capacidade dos seus princípios acomodarem nossas convicções mais firmes, e prover orientação quando a orientação é necessária¹⁴⁹.

Como a passagem indica, a teoria é construída a partir de certos juízos morais tomados provisoriamente como fixos. Rawls parece assumir que há um subconjunto de juízos (os mais fixos) que restringe outro subconjunto (os menos fixos). O primeiro subconjunto possuiria assim uma certa prioridade epistêmica sobre o segundo. Nos termos da citação acima, os primeiros seriam “feel sure” e os segundos com “much less assurance”. Como nós não sabemos o que pensar sobre diversas questões morais (sobre quais devem ser os princípios para regular a nossa sociedade, por exemplo), o objetivo da teoria moral, cumprido por meio do procedimento de construção, é ordenar todas as nossas convicções ponderadas em ordem de importância para nós e consequentemente extrair um conjunto de princípios que revela aquilo que acreditamos representar o que é justo em questões de justiça¹⁵⁰. Este ponto revela como Rawls possui uma noção de teoria moral: ela estabelece um conjunto de deveres e obrigações que devem ser seguidos por todos, mas esses deveres e obrigações, antes que imposições externas, são rearticulações daquilo que cotidianamente já aceitamos como constituindo deveres e obrigações.

Cumpriria agora especificar qual o conteúdo dessas concepções de pessoa e sociedade e como elas são incorporadas e funcionam no interior da posição original. Contudo, a abordagem desse ponto fugiria ao escopo do presente trabalho¹⁵¹. Para os presentes propósitos, é suficiente considerar que a posição original adquire a sua força justificatória em função de espelhar ou modelar adequadamente os ideais especificados nas concepções-modelo de sociedade bem-ordenada e concepção de pessoa como livre e igual, concepções essas que são

¹⁴⁹ “There are question which we feel sure must be answered in a certain way. For example, we are confident that religious intolerance and racial discrimination are unjust... these convictions are provisional fixed points which we presume any conception of justice must fit. But we have much less assurance as to what is the correct distributions of wealth and authority. Here we may be looking for a way to remove our doubts. We can check an interpretation of the initial situation, then, by the capacity of its principles to accommodate our firmest convictions and to provide guidance where guidance is needed” (*TJ*, p.19-20/rev, p.17).

¹⁵⁰ Esta passagem do *PL* é especialmente esclarecedora e não deixa dúvidas do que Rawls tem em mente com o construtivismo: “We collect such settled convictions as the belief in the religious toleration and the rejection of slavery and try to organize the basic ideas and principles implicit in these conviction into a coherent political conception of justice. These convictions are provisional fixed points” (*PL*, p.8).

¹⁵¹ A minha caracterização não se distingue daquela oferecida por Baynes na primeira parte do seu ensaio (e apenas na primeira parte) “Construtivism and Practical Reason in Rawls”, em que é desenvolvida a ideia de que a posição original adquire a sua força justificatória em função de espelhar ou modelar adequadamente os ideais especificados nas concepções-modelo (BAYNES, 1992, p.20-25).

uma extensão dos nossos juízos morais ponderados¹⁵². O quadro acima é pouco profundo, mas o meu objetivo aqui é apenas destacar que as restrições da posição original “não são externas”¹⁵³ aos juízos morais ponderados idealmente configurados em ERA e que a sua validade depende *inteiramente* das restrições que ela expressa. Isso vem a significar que examinar se um princípio seria acordado pelas partes é o *mesmo* que examinar se ele seria coerente com as concepções-modelo de sociedade e pessoa, ou, é o *mesmo* que afirmar que ele seria aceita em ERA¹⁵⁴.

2.3 Primeira pessoa vs terceira pessoa

A exposição do ERE no capítulo passado assume que ele é apresentado como no *Outline* como uma investigação empírica cujo objetivo é descrever o nosso senso de justiça. Rawls supõe que há uma competência cognitiva confiável e estipula que o *desideratum* do ERE é oferecer uma “explicação” dessa competência a partir da consideração dos seus resultados empíricos, que no caso são os juízos morais ponderados. Segundo o quadro esboçado até aqui, o ERA não nutre essa aspiração descritivista. Como Mikhail recentemente publicou um interessante livro no qual defende uma interpretação em terceira pessoa do ERA, uma interpretação que não reconhece qualquer diferença significativa entre a posição de Rawls no

¹⁵² KCE, p.256. RAM, p.233; KC, p.308-309

¹⁵³ KC, p.319.

¹⁵⁴ Dado esse quadro, a posição original não deve ser vista como uma estratégia contratualista de justificação. Hampton escreveu um relativamente extenso ensaio para demonstrar que não há espaço para argumentos contratualistas em TJ. O cerne da sua tese é apontar para a inexistência de um critério de individuação das partes situadas na posição original. Todas as características das partes (personalidade, modo de raciocinar, motivações) seriam completamente idênticas, e, postula Hampton, objetos indiscerníveis uns dos outros no interior dos termos de um dado discurso devem ser construídos como idênticos para esse discurso (HAMPTON, 1980, p.334). Antes de Hampton, Alexander havia sustentado que a justiça como equidade não é contratualista porque seria possível alcançar, desde que se mantivesse “as normas substantivas básicas” inalteradas, o mesmo resultado construindo a posição original em termos de uma teoria da escolha individual ou do observador ideal (ALEXANDER, 1974, p.597). Em mais de um momento depois de TJ nós podemos ler Rawls afirmando que para cada um de nós (individualmente) entrar na posição original, basta “raciocinar em conformidade com as restrições que ela modela e as razões que essas restrições permitem” (JF, p.86)¹⁵⁴. Rawls escreveu um ensaio para responder a Alexander, argumentando que há uma restrição operante no momento da escolha que é essencial e unicamente contratualista: as *imposições do comprometimento* (*strains of commitment*). A ideia é a seguinte: uma vez que o acordo é um empreendimento conjunto, e não simplesmente a mesma escolha realizada individualmente, ninguém pode acordar um princípio se houver dúvidas de que esse princípio e as suas consequências não poderão ser honradas (RAM, p.250). Hampton respondeu essa réplica de Rawls invocando as restrições formais do conceito de correto, restrições essas que, segundo Rawls, estão incorporadas na posição original. Uma dessas restrições é a de finalidade: ela determina que as partes devem avaliar o sistema de princípios como a corte última de apelo em questões de raciocínio prático, de modo que a concepção de justiça “acordada” deve ser conclusiva. De acordo com essa restrição “nós não podemos contar com uma segunda vez se não gostamos do resultado” (TJ, p.135/rev, p.116). O problema é que Rawls também afirma que essas restrições formais compõem o conceito de correto e que elas são gerais o bastante ao ponto de que todas as teorias morais tradicionais as satisfazem (TJ, p.130). Quer dizer, Rawls afirma que o utilitarismo, que utiliza a ideia de uma escolha individual, também satisfaz a restrição de finalidade ou irrevogabilidade, justamente o que ele afirma em *Reply to Alexander and Musgrave* que o utilitarismo é incapaz de fazer. Como essa restrição é a única razão alegada por Rawls para defender a ideia de acordo a partir das imposições do comprometimento, temos de concluir que há uma considerável contradição aqui.

Outline e em *TJ*, pretendo nesta subseção argumentar diretamente contra Mikhail para defender que o ERA, diferentemente do ERE, funciona na primeira pessoa.

O objetivo de Mikhail é defender a existência de um naturalismo em Rawls. A analogia gramatical da seção 9 de *TJ* comporia o núcleo dessa interpretação. Ao propor essa analogia, Rawls assumiria que o objetivo principal de *TJ* é descrever um objeto factual que Mikhail caracteriza como “o sistema moral da mente/cérebro humana” (*I-Morality*). Esse sistema moral seria um composto de certas regras ou princípios operativos implicados no uso de conceitos morais pré-teóricos em nossos juízos morais ponderados¹⁵⁵. Basicamente, Rawls teria o mesmo programa empírico dos linguistas que seguem a tradição de Chomsky, exceto por buscar descrever uma capacidade moral em vez de uma capacidade linguística.

Esse programa descritivista é indicado por Rawls quando ele afirma, dentre outras coisas, que a sua teoria é um “tipo de psicologia”. São essas referências que Mikhail coleta para defender uma leitura naturalizada do ERA. O seu ponto não é defender que Rawls executou esse projeto plenamente, mas que ele esboçou as bases de tal programa. Mas Rawls é naturalista? Não é exatamente claro qual o ponto a que Mikhail quer chegar ao propor essa maneira de ler Rawls. Ele se foca apenas na interpretação das seções 4 e 9 da *TJ* e no *IMT*, e também no *Outline*, ignorando todos os outros textos publicados por Rawls e até mesmo as outras seções de *TJ*. Embora essa explicação seja um tanto obscura, é preciso reconhecer, na própria *TJ* Rawls rejeita estar oferecendo uma justificação naturalista dos seus princípios porque não é possível definir os termos morais em termos não-morais¹⁵⁶. Em *KC* é expressamente afirmado que os princípios “não são, como em algumas teorias naturalistas, derivados de verdades da ciência e ajustados de acordo com os avanços na psicologia moral e na teoria social”¹⁵⁷. Mikhail reconhece em mais de um momento que Rawls depois de *TJ* e do *IMT* abandonou o seu projeto naturalista em nome do construtivismo¹⁵⁸. Ele também reconhece que o construtivismo, visto então como oposto ao naturalismo, é mais ajustável às pretensões normativas mais gerais de Rawls, o que demonstraria, segundo ele, haver uma certa “tensão” entre as aspirações descritivas e normativas de *TJ*¹⁵⁹. A despeito de enfatizar a analogia gramatical, Mikhail por duas vezes admite que há semelhanças fundamentais entre a ética e a linguística, principalmente no que concerne à questão normativa da racionalidade, mas ele

¹⁵⁵ MIKHAIL, 2011, p.63.

¹⁵⁶ *TJ*, p.578/rev, p.506.

¹⁵⁷ *KC*, p.353.

¹⁵⁸ MIKHAIL, 2011, p.267; p.274; p.294-295.

¹⁵⁹ MIKHAIL, 2011, p.294-295.

não explora (e admitidamente) todas as implicações dessa diferença¹⁶⁰. Por essas razões eu confesso não compreender exatamente as pretensões de Mikhail. De qualquer forma, o que eu quero mesmo extrair da sua leitura é a caracterização do equilíbrio reflexivo como um “estado de coisas alcançado quando o *teórico moral* sabe os princípios com os quais o conjunto de juízos morais ponderados se conformam, e as premissas daqueles princípios” (o itálico é do autor)¹⁶¹.

O ERA é definido por ele como um método em terceira pessoa:

No modelo efetivo de Rawls, o teórico moral adota um ponto de vista na *terceira*-pessoa ao abordar as suas intuições morais e a dos outros (isto é, àquelas que se qualificam como juízos morais ponderados), mas ele faz isso da perspectiva de um cientista cognitivo que toma os juízos como evidência para as propriedades de um sistema cognitivo particular¹⁶².

Mas por que é necessário essa perspectiva de terceira pessoa? Rawls afirma, na seção 9 de *TJ*, “que uma explicação adequada das nossas capacidades morais certamente envolverá princípios e construções teóricas que vão muito além das normas e padrões citados na vida diária”, e que por isso não deveríamos assumir que elas podem ser caracterizadas por “preceitos do senso comum”¹⁶³. Amparado exatamente por essa passagem, Mikhail sustenta que o linguista não assume que o usuário comum da linguagem seja consciente das regras que constituem o seu conhecimento da linguagem. Até aqui, nada que não devidamente autorizado pelas palavras de Rawls. O problema é o que Mikhail extrai a seguir como consequência dessa passagem: o linguista também não assume que a pessoa “pode tornar-se consciente dessas regras por meio da introspecção”. “Como resultado de uma investigação empírica”, essas regras estão “além de qualquer conscienciosidade real ou potencial”¹⁶⁴. Por conseguinte, há uma distinção entre dois tipos de regras ou princípios que regem uma capacidade moral. Há os princípios expressos, “enunciados que uma pessoa verbaliza na tentativa de descrever, explicar ou justificar os seus juízos”, que não são relevantes para o teórico moral, e há os princípios denominados de operativos, aqueles que funcionam como os dados para teoria científica. Esses são princípios “realmente operativos no exercício do senso de justiça”, pois é o cientista, e não o próprio dono dos princípios, quem os identifica¹⁶⁵.

¹⁶⁰ MIKHAIL, 2011, p.32-33; p.209.

¹⁶¹ MIKHAIL, 2001, p.205.

¹⁶² MIKHAIL, 2011, p.283.

¹⁶³ “A correct account of moral capacities will certainly involve principles and theoretical constructions which go much beyond the norms and standards cited in everyday life” (*TJ*, p.47/rev, p.41-42).

¹⁶⁴ MIKHAIL, 2011, p.19; p.50-51.

¹⁶⁵ MIKHAIL, 2011, p.19-21.

Essa distinção é crucial para os propósitos cientificistas de Mikhail, pois ele entende que princípios expressos são “intuições morais” ou meras descrições distorcidas e não revelam as regras que *de fato* regulam a deliberação moral. Ela também oportuniza uma outra ideia central do projeto de Mikhail que é a separação entre a aceitabilidade de uma sentença e a sua gramaticalidade. Uma sentença é gramatical se está de acordo com regras válidas, e ela é aceitável se ela parece, para o falante nativo, natural ou conforme às suas intuições linguísticas. O segundo tipo é irrelevante para o projeto de descrever a capacidade cognitiva (ele é útil para aqueles que analisam o desempenho cognitivo)¹⁶⁶. O que é indispensável é assinalar a assimetria que há entre os dois: uma sentença pode ser aceitável, mas não ser gramatical, como também pode ser gramatical, mas não ser aceitável.

Se essa distinção entre gramaticalidade e aceitabilidade (ou entre princípios expressos e operativos) é necessária ou não para o programa desenvolvido por Mikhail, eu não estou interessado. O que eu gostaria de criticar é a apresentação dessa distinção como uma interpretação do ERA em Rawls. A passagem de Rawls citada algumas linhas atrás não autoriza em nenhum momento essa assimetria forte que determina que a gramaticalidade não é (ou não pode ser) acessível via introspecção¹⁶⁷. Ao contrário do que Mikhail afirma¹⁶⁸, para Rawls são os falantes nativos que têm a palavra final sobre a precisão de uma determinada descrição do seu senso de justiça, e não o linguista.

Mikhail compreende erroneamente o que o ERA assume (i) por descrição do senso de justiça e por consequência não dispõe de elementos para vê-lo como um (ii) método na primeira pessoa. Considere (i) primeiro. A primeira coisa a se notar é a inferência *non sequitor* de Mikhail: uma coisa é afirmar, como Rawls, que uma descrição do senso de justiça envolve teorização e necessita de recursos que vão além do que é citado na vida diária, outra coisa, bem diferente, é defender que esses recursos não são acessíveis à consciência via introspecção e reflexão. A segunda afirmação não é autorizada pela primeira, portanto ela não pode ser vista, sem qualificação, como uma interpretação da primeira proposição. Um segundo ponto a ser destacado é a ideia enfatizada por Daniels de que a descrição oferecida pela justiça como equidade é ideal ou hipotética. Embora essa alegação seja exagerada, como será discutido na seção seguinte, Daniels chega ao ponto de defender que Rawls não busca por nenhuma

¹⁶⁶ MIKHAIL, 2011, p.233-234.

¹⁶⁷ Mesmo no *Outline* Rawls não defende que os princípios são inacessíveis ao agente. Pelo contrário, para que possam ser vistos como justificados, eles têm de mostrar “uma capacidade para serem aceitos por juízes morais competentes após eles livremente pesarem os seus méritos por meio de criticismo e discussão aberta” (*Outline*, p.11). Mikhail em nenhum momento analisa essa informação, ignorando-a completamente.

¹⁶⁸ MIKHAIL, 2011, p.235.

“descrição”¹⁶⁹. De qualquer modo, não é possível perder de vista que a justiça como equidade tomada como um todo descreve um senso de justiça *em* ERA, o que mostra uma segunda razão pela qual a interpretação de Mikhail é inadequada.

Para Rawls, o ERA é o estado alcançado “após *a* pessoa ter pesado as várias concepções propostas e ter ou revisado os seus juízos para conformá-los a uma dessas concepções, ou mantido as suas convicções iniciais” (o *itálico* é meu)¹⁷⁰. Esse tipo de comentário mostra que o utilizador do ERA não é o teórico, como Mikhail afirma. Mas há mais. Observe como Rawls apresenta a característica da revisabilidade:

Quando é apresentada a uma pessoa uma *explicação intuitivamente apelante* do seu senso de justiça, ele pode muito bem revisar... É especialmente provável que ele fará isso *se ela puder* encontrar uma explanação para os desvios que enfraquecem *a sua confiança* nos seus juízos originais e se a concepção apresentada produz um juízo que *ela pode agora aceitar*¹⁷¹.

Note que os meus grifos acima mostram que Rawls tanto não endossa a distinção entre gramaticalidade e aceitabilidade, como explicitamente se posiciona contrariamente à ela. A revisão dos juízos em ERA é guiada pelo nível de confiança com que o próprio indivíduo vê a situação e o objetivo final da teoria é apresentar uma descrição que ele *pode agora aceitar*. Longe de buscar um distanciamento das intuições morais ordinárias (no sentido especificado por Mikhail), Rawls sustenta que um princípio válido é um princípio que se aplicado “nos levaria a fazer os mesmos juízos sobre a estrutura básica da sociedade que nós agora fazemos intuitivamente e nos quais nós temos a maior confiança”. São princípios que em momentos de dúvida e hesitação “oferecem uma resolução que nós podemos endossar em reflexão”¹⁷². A linguagem do ERA é a da aceitabilidade (“intuitivamente apelante”, “sua confiança”, “ela pode aceitar”), não da gramaticalidade.

Mikhail solenemente ignora esse elemento da confiança ou comprometimento *em reflexão*. Para ele, o ERA não é um método ou uma técnica, mas um “estado de coisas no qual princípios morais e juízos morais ponderados coincidem, e o pesquisador entende os princípios

¹⁶⁹ DANIELS, 1996, p.71.

¹⁷⁰ “This state is one reached after a person has weighed various proposed conceptions and he has either revised his judgments to accord with one of them or held fast to his initial convictions” (TJ, p.48/rev, p.42).

¹⁷¹ “When a person is presented with a intuitively appealing account of his sense of justice, he may well revise ... He is especially likely to do this if he can find an explanation for the deviations which undermines his confidence in his original judgments and if the conception presented yields a judgment which he finds he can now accept” (TJ, p.48/rev, p.42-43). Os *itálicos* são de minha autoria.

¹⁷² “In cases where our present judgments are in doubt and given with hesitations, these principles offer a resolution which we can affirm on reflection” (TJ, p.19/rev, p.17).

para os quais os juízos se conformam, junto com as premissas daqueles princípios”¹⁷³. O curioso é que Mikhail atribui essa formulação do ERA a Rawls, citando a página 20 da edição original de *TJ* como referência. De fato nessa página Rawls se refere ao ERA como um “estado de coisas” (*state of affairs*), mas essa é apenas uma opção estilística. Logo após apresentar o ERA Rawls explica que ele é reflexivo porque “nós sabemos os princípios aos quais os nossos juízos se conformam e as premissas da sua derivação”¹⁷⁴. Mikhail interpreta essa frase como significando que o pronome na primeira pessoa do plural se refere a Rawls e seus colegas teóricos morais¹⁷⁵, e por isso ele escreve “o pesquisador entende”. Esse movimento é implausível. Trinta páginas depois de utilizar esse pronome Rawls explica que “para os propósitos deste livro, as visões do leitor e do autor são as únicas que contam”¹⁷⁶. O “nós” é nós que utilizamos o ERA, em que o leitor e o autor são ambos vistos como agentes morais, não “nós” teóricos que descrevemos o senso de justiça dos outros. Rawls faz essa qualificação para afastar a necessidade de lidar com a questão de saber se cada pessoa que utilizasse o ERA chegaria à mesma conclusão. Ele afirma que é suficiente que o leitor e ele possam convergir em ERA.

Creio que o entendimento de teoria moral de Scanlon é mais ajustável a esse quadro do que aquele oferecido (e sugerido pela analogia gramatical) por Mikhail. Para Scanlon, o objetivo da teoria moral é obter uma clara ideia de que coisas são corretas e erradas por meio de um claro entendimento das razões pelas quais nós professamos certos juízos morais e, de posse dessas razões, resolver conflitos. Um “claro entendimento” mostra as nossas várias razões, como as nossas conclusões dependem delas, quais as suas limitações e como os vários conflitos entre elas devem ser entendidos¹⁷⁷. Trata-se de uma compreensão de teoria moral enquanto interpretação, e do teórico moral enquanto intérprete. Scanlon enfatiza que o componente de primeira pessoa exerce um papel essencial nesse projeto, pois ele assegura o sentido da ideia de revisabilidade como o ERA entende essa ideia¹⁷⁸:

Uma vez que a força especial dos juízos ponderados consiste simplesmente no fato de que nós estamos fortemente inclinados a pensá-los como corretos, eles perdem sua força completamente quando nós decidimos, em reflexão, que nós

¹⁷³ MIKHAIL, 2011, p.289. Além de Mikhail, Habermas também interpreta o equilíbrio reflexivo como um método na terceira pessoa: “O equilíbrio reflexivo é alcançado no momento em que o filósofo esteja certo de que aqueles envolvidos não podem rejeitar com boas razões as intuições reconstruídas e clarificadas” (HABERMAS, 1995, p.120).

¹⁷⁴ *TJ*, p.20/rev, p.18.

¹⁷⁵ MIKHAIL, 2011, p.205.

¹⁷⁶ “So for the purposes of this book, the views of the reader and the author are the only ones that count” (*TJ*, p.50/rev, p.44).

¹⁷⁷ SCANLON, 1992, p.8.

¹⁷⁸ Outro que faz esse ponto é RAZ, 1982, p.316.

estamos errados sobre eles. A revisibilidade da classe de juízos ponderados assim ilustra o fato de que a busca por Equilíbrio Reflexivo é essencialmente um empreendimento em primeira pessoa; se os juízos em questão são aqueles de outras pessoas, tratados como um tipo de fato sociológico, então eles não seriam suscetíveis a esse tipo particular de revisão¹⁷⁹.

Mikhail entende que métodos na primeira pessoa limitam a construção da teoria justamente porque nesse ponto de vista nós estamos propensos a diversas distorções¹⁸⁰. Porém, considere as consequências epistemológica da posição de Mikhail. De acordo com essa leitura naturalista, os justificadores (princípios) dos juízos morais ponderados são, em princípio, inacessíveis à reflexão. Pode-se dizer que eles são externos à capacidade de reflexão. A posição de Mikhail configura-se como uma forma radical do que é conhecido em epistemologia como externalismo epistêmico. Essa é a posição segundo a qual, conforme definido por Bonjour, a pessoa que possui uma crença justificada não precisa ter (e em geral não terá) qualquer apreensão cognitiva daquilo que justifica as suas crenças¹⁸¹. Por outro lado, a interpretação que será apresentada no capítulo seguinte, e que é compartilhada por uma variedade de autores como Brink, Scanlon, Daniels, DePaul, Sosa e Scheffler, realça que o ERA incorpora um internalismo epistêmico de acordo com o qual as razões justificadoras têm de ser cognitivamente acessíveis, via reflexão, ao agente¹⁸². O problema do externalismo epistêmico é que ele colapsa o conceito de racionalidade e de justificação epistêmica, essencialmente conceitos de primeira pessoa¹⁸³. Se aquilo que determina a correteza das *minhas* crenças (segundo o modelo de Mikhail seriam as regras oferecidas pelo cientista) é inacessível via reflexão para mim, digamos porque todas as minhas crenças após reflexão me apontam para um princípio contrário, então segue-se que será irracional, do meu ponto de vista, sustentar uma crença que é, apesar de irracional para mim, em algum sentido justificada¹⁸⁴.

2.4 ERA no *Political Liberalism* e *Reply to Habermas*

Antes de encerrar a apresentação do ERA, é preciso considerá-lo no ambiente conceitual do *PL*. Quero compreender o que Rawls tem em mente quando ele afirma que cada cidadão alcança um ERA distinto¹⁸⁵ e por que essa multiplicidade de “ERAs” é vista como o problema fundamental para a filosofia política atualmente. Até esta seção, busquei direcionar o leitor para

¹⁷⁹ SCANLON, 1992, p.10.

¹⁸⁰ MIKHAIL, 2011, p.286.

¹⁸¹ BONJOUR, 2003, p.13.

¹⁸² BONJOUR, 2003, p.25.

¹⁸³ BONJOUR, 2003, p.37.

¹⁸⁴ Aqui apenas sigo o raciocínio da bem conhecida objeção de Bonjour contra o confiabilismo (BONJOUR, 2003, p.17-20; 2003, p.26-30).

¹⁸⁵ *JF*, p.31.

uma compreensão do ERA enquanto um método para ser utilizado na primeira pessoa. No capítulo seguinte eu explorarei a importância de reconhecer esse ponto, mas já nesta subseção essa ideia se revestirá de toda a importância, pois, como será abordado, permite associar o ERA ao internalismo defendido por Bernard Williams.

No capítulo seguinte eu também defenderei que a tese fundamental de Rawls com o ERA é a de que *se um* indivíduo seguir o ERA adequadamente, ele endossará a justiça como equidade tal como é apresentada em *TJ*. Isso pode ser expresso dizendo que o ERA descreve uma ideia de justificação pessoal ou racional. Creio que nesta subseção já é possível introduzir preliminarmente esse conceito. Apesar de não o encontrarmos claramente definido em *TJ* ou no *PL*, penso que em *RH* Rawls claramente desenvolve um conceito de justificação pessoal. O conceito é oficialmente intitulado *full justification*, conforme cito:

Justificação plena (*full*) é alcançada por um cidadão individual como um membro da sociedade civil. (Nós assumimos que cada cidadão afirma tanto uma concepção política quanto uma doutrina abrangente). Nesse caso, o cidadão aceita uma concepção política e *completa a sua justificação* incorporando-a de algum modo nas doutrinas abrangentes afirmadas como verdadeiras ou razoáveis, dependendo do que a doutrina permite. Alguns podem considerar a concepção política plenamente justificada mesmo se ela não é aceita por outras pessoas. Se nossa concepção é endossada por ela (isto não implica dar) não é dado peso suficiente para suspender a justificação plena em nossos próprios olhos¹⁸⁶.

Nessa passagem não é feita uma conexão explícita entre a justificação *full* e o ERA. O que é efetivamente pressuposto é que a justificação pública pressupõe o ERA, mas não se identifica inteiramente com ele porque ela necessita de um equilíbrio reflexivo geral (*general*). Mas se nós consideramos que é dito que “a justificação *full* é realizada por um cidadão individual” e que para Rawls o ERA é o ponto de vista de um cidadão tomado individualmente¹⁸⁷, então nós podemos concluir que a justificação *full* se aplica ao ERA. Conforme pode ser extraído da passagem acima, a ideia fundamental é a de que a justiça como equidade é (pessoalmente) justificada a partir de crenças abrangentes em ERA. A meu ver, não compreendemos o equilíbrio reflexivo sem examinar em detalhe o que isso significa.

¹⁸⁶ “Full justification is carried out by an individual citizen as a member of civil society. (We assume that each citizen affirms both a political conception and a comprehensive doctrine.) In this case, the citizen accepts a political conception and fills out its justification by embedding it in some way into the citizen’s comprehensive doctrines as either true or reasonable, depending on what that doctrine allows. Some may consider the political conception fully justified even though it is not accepted by other people. Whether our view is endorsed by them is not given sufficient weight to suspend its full justification in our own eyes” (*RH*, p.384n16).

¹⁸⁷ *JF*, p.30-31.

Antes de efetivamente abordar essa questão, é preciso registrar dois pontos. O primeiro é a explícita enumeração, por parte de Rawls, de um sentido de primeira pessoa em que podemos falar da justificação da sua teoria. Ele enuncia outros, como as intimamente conectadas ideias de justificação *pro tanto* e pública, e, pela primeira vez em um texto seu (e após dois livros inteiros publicados), explicitamente confirma algo que eu insistirei daqui em diante, a saber, que para falar da “justificação” da justiça como equidade é preciso se ater a diferentes modos de entender esse conceito¹⁸⁸. Um segundo ponto é terminológico, já que não opto por usar a nomenclatura oficial de Rawls. Faço isso por duas razões. Em primeiro lugar, falar de justificação *full* (i) sugere que se trata do tipo de justificação próprio do equilíbrio reflexivo *full*, o que não é definitivamente o caso. Essa versão do equilíbrio reflexivo lida com a ideia de justificação pública, que não deve ser confundida com a ideia de justificação pessoal do ERA. No próprio *PL* Rawls utiliza a expressão *full justification* em outro sentido, nessa ocasião em mais íntima conexão com a ideia de justificação pública¹⁸⁹. E em segundo lugar, o adjetivo *full* (ii) pode indicar ao leitor que a justificação *full* é efetivamente tudo o que nós precisamos ou podemos alcançar, o que também não é verdadeiro. Nem tudo o que é pessoalmente justificável é publicamente justificável.

Em virtude do que já foi discutido, sabemos que quem se engaja no ERA empreende um processo de avanços e recuos entre as suas convicções em vários níveis de generalidade, com a estrita finalidade de encontrar um conjunto de princípios que explique o seu senso de justiça como um todo. Para obter tal conjunto, ela precisará partir das ideias subjacentes nos seus pontos fixos provisórios para, a partir delas, revisar, ajustar ou destacar outros juízos, até o ponto em que os seus juízos morais ponderados estejam todos em equilíbrio com essas ideias subjacentes. O que essa breve descrição torna particularmente claro é que quando uma pessoa utiliza o ERA ela não reflete apenas a partir dos seus juízos morais ponderados sobre justiça, mas a partir de todas as suas convicções. Isso pode parecer estranho haja vista que a tese central do *PL* sustenta que a justiça como equidade é construída unicamente a partir de valores políticos e descreve, portanto, apenas esse tipo de juízo¹⁹⁰. Mas se a justiça como equidade é justificada por poder ser afirmada em ERA, ela inevitavelmente dependerá de outras convicções que não são políticas. E essas outras convicções, por sua vez, serão igualmente justificadas em ERA. Rawls nunca negou essas implicações da sua posição. O que ele afirmou foi que, se alguém seguir o ERA, então a justiça como equidade explicará as suas crenças sobre justiça. Está

¹⁸⁸ *RH*, p.386-387.

¹⁸⁹ *PL*, p.67.

¹⁹⁰ *PL*, p.10.

completamente aberta a possibilidade de que outras teorias morais expliquem as suas outras crenças morais justificadas em ERA¹⁹¹. O *ponto* da ideia de político é justamente lidar com a consequência de que as pessoas possuem crenças morais (abrangentes) pessoalmente justificadas.

Mas como compreender essa relação aparentemente contraditória entre o ERA e uma concepção política como o seu resultado? Creio que a ameaça de contradição é desfeita se analisarmos a justiça como equidade de duas perspectivas distintas. Uma é a perspectiva do próprio teórico moral que a constrói como uma descrição de valores políticos presentes na cultura política pública ou de nós, na posição de estudiosos dos seus livros. É nessa perspectiva que a justiça como equidade é uma concepção política de justiça. Uma concepção é dita política se ela satisfizer o requisito do *freestanding*, segundo o qual uma teoria que se pretenda legítima deve poder ser apresentada independentemente de qualquer doutrina abrangente. Alternativamente, o *freestanding* demanda que o conteúdo de uma teoria da justiça seja expresso exclusivamente em termos de certos ideais fundamentais vistas como implícitas na cultura política pública de uma sociedade democrática¹⁹². É isso que Rawls buscou fazer ao expressar que na formulação da teoria se “olha para a nossa cultura política pública” entendida como um “fundo compartilhado de ideias básicas e princípios implicitamente reconhecidos” na esperança de que eles “possam ser combinados em uma concepção de justiça política” que é aceitável “após a devida reflexão” (no contexto do *PL*, *due reflection* sempre funciona como sinônimo de equilíbrio reflexivo)¹⁹³.

Que uma concepção política de justiça seja construída unicamente a partir de valores políticos não implica que quem se engaja no ERA para descobrir que a justiça como equidade é a concepção de justiça mais razoável para si mesmo ignore as suas outras convicções não-políticas, ou busque justificar apenas as suas convicções políticas. E isso não poderia ser diferente uma vez que o ERA é um método na primeira pessoa. Quem reflete sobre o seu senso de justiça, necessariamente o faz com todas as suas convicções, sejam elas religiosas, morais, filosóficas ou políticas. Isso nos conduz à segunda perspectiva: a perspectiva interna e de primeira pessoa peculiar ao ERA. É, por assim dizer, o ponto de vista do cidadão comum

¹⁹¹ É perfeitamente viável isolar o método do ERA e dissociá-lo de qualquer teoria moral, pois para alguém estar justificado em ERA não é necessário que alguma teoria moral particular explique os seus juízos.

¹⁹² *PL*, p.12-14.

¹⁹³ “We look, then, to our public political culture itself, including its main institution and the historical traditions of their interpretation, as the shared fund of implicitly recognized basic ideas and principles. The hope is that these ideas and principles can be formulated clearly enough to be combined into a conception of political justice congenial to our most firmly held convictions. We express this by saying that a political conception of justice, to be acceptable, must be in accordance with our considered conviction, at all levels of generality, on due reflection (or in what I have called ‘reflective equilibrium’)” (*PL*, p.8; *JPFNM*, p.393).

quando ele reflete sobre o seu senso de justiça e descobre que a justiça como equidade é a explicação mais racional dos seus juízos morais ponderados sobre justiça, ou o nosso próprio, quando não estamos pensando filosoficamente sobre a justiça como equidade e estamos efetivamente decidindo o que fazer em questões de justiça. Nessa perspectiva cada um de nós afirma a concepção proposta por Rawls pesando o seu mérito a partir das nossas crenças morais, religiosas e filosóficas (sempre em ERA). Aqui a justiça como equidade deixa de ser política e passa a ser abrangente. Para um cristão bem-informado (e bem situado historicamente), o repúdio à escravidão nada mais é do que uma forma de respeitar o fato de que Deus criou todos os homens à Sua imagem e semelhança, e que por isso seria um insulto à Sua imagem desrespeitar a liberdade dos negros. Por outro lado, um liberal ateu discordaria do cristão e entenderia que repudiar a escravidão é respeitar o valor moral da liberdade e da autonomia, intrínsecos a todos os seres humanos. Tanto o ateu quanto o cristão afirmam o mesmo juízo ponderado sobre justiça, embora um veja as razões do outro para afirmá-lo como falsas ou incompletas (no caso do meu exemplo acima, o cristão diria para o liberal ateu que ele ignorou que a liberdade é um valor sagrado, ao passo que o liberal ateu diria para o cristão que a sua visão é simplesmente falsa).

O quadro acima mostra que é um grave erro considerar que uma concepção política, no sentido definido por Rawls, não depende (em certo sentido) de valores abrangentes. Em seu último escrito sobre o assunto, *The Idea of Public Reason Revisited*, Rawls expressa com clareza que as raízes da fidelidade pelas ideias de uma concepção política reside nas respectivas doutrinas abrangentes, religiosas e não-religiosas, que os cidadãos democráticos sustentam¹⁹⁴. Por que alguém acredita, para citar os dois exemplos de pontos fixos provisórios sempre mencionados, que a tolerância religiosa deve ser respeitada e que o repúdio à escravidão deve ser assumido como correto? *Do ponto de vista de quem adota o ERA*, o que fundamenta essas crenças? E o que motiva o indivíduo a agir com base em uma concepção de justiça que as descreve? Como uma concepção meramente política gera algum tipo de motivação? Como os exemplos do cristão e do ateu liberal antecipam, uma resposta de Rawls é que, do ponto de vista do ERA, o que fundamenta a concepção política são as doutrinas abrangentes. Tomando esse exemplo, pode-se dizer que os valores políticos articulados pela justiça como equidade são uma extensão dos valores não-políticos, e que os indivíduos que afirmam a justiça como equidade a veem como uma aplicação da sua doutrina abrangente a questões de justiça política. Como Rawls diz, algumas doutrinas abrangentes simplesmente veem os valores políticos como

¹⁹⁴ PRR, p.463.

deduções a partir das suas crenças morais (o seu exemplo preferido para ilustrar esse ponto é o de alguém que professa a filosofia moral de Kant)¹⁹⁵. Naturalmente, um cristão só segue os valores políticos porque ele os relaciona com os seus valores religiosos não-políticos. Esse é um caso exemplar para mostrar em que sentido uma concepção política de justiça depende de doutrinas abrangentes e porque, em si mesma, ela é incapaz de conduzir alguém à ação política de acordo com as suas recomendações. Ela não gera motivações, ela apenas articula motivações preexistentes.

Rawls também considera o caso de cidadãos que sustentam doutrinas parcialmente abrangentes. Diferentemente das doutrinas abrangentes, que associam o caráter meritório dos valores políticos exclusivamente ao mérito dos seus valores abrangentes dos quais eles podem ser vistos como consequência, as doutrinas parcialmente abrangentes não se pronunciam sobre todas as questões, especialmente questões básicas de justiça¹⁹⁶. As pessoas que professam essas doutrinas afirmam os valores políticos por um “comprometimento” independente e em virtude do mérito próprio desses valores. Rawls alega que isso é possível porque os valores políticos “são grandes valores” que constituem o nosso próprio mundo social e asseguram as condições que permitem a cooperação social e o reconhecimento da reciprocidade; são valores que se desenvolveram lenta e, poderíamos acrescentar, dolorosamente ao longo do tempo e que por isso representam verdadeiras conquistas sociais que a maior parte das pessoas reconhece com base na sua experiência e conhecimento do passado¹⁹⁷.

É claro que a relação entre valores abrangentes e políticos é uma questão complexa. Como os críticos não deixaram de notar, Rawls deveria ter dedicado mais tempo para explicar como ela funcionaria (o próprio Rawls reconhece que o seu tratamento da questão não é exaustivo¹⁹⁸). Não é evidente que algumas doutrinas abrangentes razoáveis realmente poderiam endossar os valores políticos da justiça como equidade¹⁹⁹. Quero, porém, proceder alheamente a essa questão para poder focalizar na ideia central de que “*todos* aqueles que afirmam a concepção política partem do interior da sua própria doutrina abrangente e a baseiam nos fundamentos religiosos, filosóficos e morais que ela fornece”²⁰⁰. Em outras palavras, a

¹⁹⁵ *PL*, p.169.

¹⁹⁶ *PL*, p.159.

¹⁹⁷ *PL*, p.157.

¹⁹⁸ *PL*, p.170.

¹⁹⁹ Uma infinidade de intérpretes levanta essa objeção. Wenar dedicou um extenso ensaio a ela (WENAR, 1995, p.32-62).

²⁰⁰ “All those who affirm the political conception start from within their own comprehensive view and draw on the religious, philosophical, and moral grounds it provides” (*PL*, p.147). O “*todos*” é significativo. Mesmo em casos de doutrinas parcialmente abrangentes em que os valores políticos são sustentados por razões independentes, é possível dizer que, após eles serem efetivamente endossados, eles passam a funcionar como valores abrangentes. O indivíduo, embora tenha sido movido em direção aos valores políticos independentemente de outras

concepção política é aceita *porque* ela pode ser baseada em tais fundamentos, e as obrigações que ela impõe são legítimas porque os cidadãos podem ver essas obrigações como extensões das obrigações estipuladas pela suas doutrinas abrangentes. O ponto crucial é que, do ponto de vista do ERA, os valores políticos sempre são fundamentados a partir de outras crenças, crenças que são abrangentes e, portanto, relativas a cada ERA particular. O imperativo da justiça como equidade, que comanda que o indivíduo deve seguir os dois princípios de justiça, é condicionado à existência de uma motivação, um desejo ou uma crença particular, elementos estes que não são introduzidos pela própria justiça como equidade. O fato de que esses elementos não devem ser introduzidos pela própria teoria explica em que sentido ela é política e *freestanding*.

É útil investigar por que Rawls adota essa compreensão de normatividade. Tradicionalmente, talvez muitos estejam inclinados a pensar que o filósofo deveria utilizar a sua *expertise* argumentativa para produzir uma teoria que convença as pessoas do seu mérito independentemente das crenças que as pessoas sustentam. Afinal, o filósofo deveria justamente dizer que crenças as pessoas deveriam sustentar, e o bom filósofo seria aquele cuja argumentação é capaz de conduzir todos para o seu lado. Rawls evidentemente não nutre essa pretensão quando afirma que “a concepção política é um módulo, uma parte constituinte essencial, que se ajusta em diversas doutrinas abrangentes razoáveis e pode ser suportadas por ela”²⁰¹ e que, como visto, deve o seu mérito ou a sua justificação pessoal a essa possibilidade de ajuste.

Claro que Rawls se empenha em convencer as pessoas de que elas poderiam ver a justiça como equidade como pessoalmente justificável (como a melhor descrição do seu senso de justiça em ERA), mas ele não diz por que a sua teoria “vale a pena ser seguida”. Ele não diz que eu deveria seguir a justiça como equidade porque ela pode funcionar como módulo, porque ela é uma concepção política de justiça ou porque ela pode fundar uma sociedade estável. Da perspectiva de primeira pessoa, razões como essas parecem ser para Rawls “frias” (nada impede que um cidadão particular utilize essas razões como suas razões últimas; talvez o próprio Rawls, enquanto indivíduo que vive segundo a sua própria filosofia, seja um exemplo. O que Rawls não pode é, na posição de teórico, afirmar que essas são as razões últimas para a sua teoria).

considerações, tende a, posteriormente, oferecer uma fundamentação abrangente para esses valores. Ele pode dizer que eles são verdadeiros por serem valores que representam o desenvolvimento da própria capacidade humana de utilizar a razão livremente, ou algo assim. O que Rawls parece querer defender é que um valor político, no âmbito de primeira pessoa, não se sustenta por si mesmo: ele sempre precisa estar apoiado em um valor não-político mais fundamental.

²⁰¹ “The political conception is a module, an essential constituent part, that fits into and can be supported by various reasonable comprehensive doctrine” (PL, p.12).

Sou eu quem determina a fundamentação da justiça como equidade e, conseqüentemente, especifico as razões últimas que determinam por que ela vale a pena ser seguida. Daniels discutiu esse assunto e chamou a tentativa de inculcar razões últimas como o “sonho filosófico” que Rawls estava destruindo (ou tentando) no *PL*. A característica básica dos “sonhadores”, como destacado por Daniels, é que eles imaginam que o ERA alcançado pelo filósofo pode ser um modelo para todos²⁰², de forma que as razões que o filósofo oferece como razões últimas em favor da sua teoria poderiam ser também aceitas pelos outros como suas razões. Cada pessoa deveria ajustar as suas crenças às aquelas alcançadas pelo ERA ideal obtido pelo filósofo, visto então como uma espécie de *phronimus*. Contrariamente a esse modelo, no modelo efetivo de Rawls tudo o que o filósofo pode fazer é mostrar para a pessoa que, dadas algumas de suas crenças, é pessoalmente racional que ela aceite outras (a justiça como equidade) e que, consideradas as suas doutrinas abrangentes, ela pode ver que essas crenças são dignas do seu assentimento. Essa é, em certo sentido, uma “perda filosófica”. Representa também um enorme deflacionamento da noção de obrigação moral: em questões de justiça política, ninguém está incondicionalmente “obrigado” a praticar qualquer ação.

2.4.1 O internalismo de razões

Aceitando o que foi exposto, alguém poderia sentir-se inclinado a questionar se Rawls não estaria desistindo tão facilmente da possibilidade de inculcar essa razão última por meio da filosofia, assim salvaguardando uma noção mais robusta de obrigação. Essa é uma questão de extrema relevância. Creio que Rawls possui duas razões fortes como suporte para a sua abordagem. A primeira é moral e se explica pelo seu liberalismo político: tentar oferecer razões últimas é uma forma de desrespeitar a concepção de pessoa enquanto livre e igual, racional e razoável. Precisamente, essa tentativa restringiria inaceitavelmente a liberdade que cada um tem para perseguir a concepção de bem razoável que lhe aprouver, com isso desrespeitando a sua autonomia. O domínio das razões últimas é o domínio das doutrinas abrangentes, e aqui o papel da filosofia política deve ser o de preservar a autonomia individual e o livre exercício da razão²⁰³.

A segunda razão não é invocada por Rawls, mas, tentarei mostrar, está devidamente implicada pela sua posição. A segunda razão ampliaria a primeira para defender que, além de política e eticamente indesejável, é uma tarefa que, conforme especificado por uma determinada concepção de razão prática, é inapropriada. O objetivo desta subseção é analisar essa concepção

²⁰² DANIELS, 1996, p.161.

²⁰³ *PL*, xvi. Referências para esse argumento de Rawls encontram-se por todo o *PL*.

de razão prática subjacente ao ERA. Sugerirei que é plausível relacionar o ERA com uma interpretação possível do *internalismo de razões normativas* defendido por Bernard Williams ao longo da sua carreira. Creio que esse internalismo ajuda a compreender a natureza indestrutível do pluralismo razoável e, conseqüentemente, por que o que é publicamente justificável tem de necessariamente ser pessoalmente justificável (para todas as pessoas relevantes concernidas). Em acréscimo, ele permite explicar e destacar a relevância da ideia de que o ERA descreve um senso de justiça. No que se segue o meu objetivo será apenas expositório: não pretendo pormenorizar os detalhes desse internalismo, tampouco entrar no debate internalismo vs externalismo. Mais modestamente, tenciono apenas sugerir uma direção de interpretação, isto é, um esboço do que julgo ser o ponto de partida para um futuro projeto de oferecer uma formulação completa do tipo de internalismo pressuposto pelo ERA.

Em “Internal and External Reasons”, dentre outros ensaios, o objetivo de Williams é determinar as condições de verdade de enunciados da forma “S tem razão para ϕ ”, em que S é uma pessoa qualquer e ϕ está para algum verbo de ação (note que aqui Williams não está restringindo-se a um agente idealizado, como é o agente razoável de Rawls, e nem está exclusivamente preocupado com razões éticas, mas com razões práticas em geral). Consideradas as intenções deste trabalho, com o símbolo ϕ podemos nos referir simplesmente a uma ação recomendada pela justiça como equidade. Como Williams apresenta, há duas posições rivais sobre como entender enunciados sobre razões: uma posição internalista e outra externalista. A distinção fundamental entre as duas posições é que a primeira afirma que a condição seguinte é necessária para a verdade de enunciados sobre razões para a ação, enquanto que a segunda nega que essa *restrição internalista* precise ser satisfeita:

S tem uma razão para ϕ se e somente se ele puder alcançar a conclusão para ϕ por meio de um processo deliberativo adequado a partir do seu *c* atual²⁰⁴

O *c* de S é o que Williams chama de seu “*set* motivacional subjetivo” (*subjective motivational set*) que inclui, além de desejos, coisas como “disposições de avaliação, padrão de reação emocional, fidelidades pessoais e vários projetos” ou, mais genericamente, incorpora os “comprometimentos do agente”²⁰⁵. Eu proponho que tratemos de comprometimentos e que encaremos parte desses comprometimentos como algo que é dado pela doutrina abrangente do

²⁰⁴ Essa formulação é, em essência, a mesma oferecida por Williams: “A tem uma razão para ϕ se ele puder alcançar a conclusão para ϕ por uma rota deliberativa adequada a partir das motivações que ele já tem” (WILLIAMS, 1995, p.35).

²⁰⁵ WILLIAMS, 1981, p.105.

indivíduo. Por processo deliberativo “adequado” (*sound*) referido por Williams, proponho que entendamos o processo descrito pelo ERA, embora, como se verá, Williams muito provavelmente protestaria contra essa identificação. Sendo assim, a posição internalista afirma que S tem uma razão para ϕ se e somente se S puder alcançar a conclusão para ϕ em ERA a partir dos seus comprometimentos atuais (dentre os quais as suas crenças abrangentes). Contrariamente ao internalista, para o externalista se ϕ for uma ação correta que constitui uma genuína obrigação moral, então S tem uma razão para ϕ mesmo se ele não for capaz, via ERA a partir do seu *c*, de alcançar uma conclusão para ϕ .

Em geral, uma posição externalista rejeita que a restrição internalista deva ser satisfeita porque defende que as razões normativas para o agir não são dadas por elementos subjetivos do agente, como suas crenças, desejos e planos de vida, mas são dados por fatos, valores ou verdades normativas cuja validade é independente desses elementos subjetivos. O externalista evidentemente não nega que existam razões subjetivas e particulares a cada agente, o que ele afirma é que essas razões subjetivas e particulares derivam o seu peso normativo dessas razões externas²⁰⁶.

Em contraposição, para um internalista como Williams (e Rawls), as razões são dadas por fatos sobre o que cumpriria ou realizaria os nossos desejos, crenças e comprometimentos. Para o externalista, se fumar causa câncer, e conseqüentemente morte, eu tenho uma razão para parar de fumar independentemente dos meus desejos, pois a morte é em si mesma um fato indesejável. Para o internalista, eu apenas devo parar de fumar se eu, de alguma forma, escolher agir desse modo após reflexão. A normatividade de razões subjetivas para ações não depende de fatos, mas de uma escolha deliberada e racional. É claro que eu escolheria parar de fumar porque isso prolongaria a minha vida, mas esse fato constitui uma razão, segundo o internalista, apenas se o seu reconhecimento como um fato que impõe razões for resultado de uma escolha racional do agente ao qual a razão em questão se aplica²⁰⁷. O que nós temos razão para fazer, de acordo com uma posição internalista, é o mesmo o que nós, se fôssemos racionais, escolheríamos. Adaptando para o ERA, pode-se definir assim: S tem uma razão para ϕ se e somente se S puder alcançar uma conclusão para ϕ em ERA. Como razões especificam o que alguém deve fazer, a ação que S deve realizar é idêntica àquela que ele realizaria se se engajasse no ERA.

Para ilustrar a divergência das duas posições, Williams cita o personagem Owen Wingrave, de Henry James. A família de Owen exige dele que ele se aliste ao exército, com o

²⁰⁶ Essa apresentação de uma posição externalista segue Parfit (2011, p.45-61).

²⁰⁷ PARFIT, 2011, p.63.

argumento de que é uma tradição antiga da sua família que os seus homens sirvam ao exército. Porém, Owen não tem nenhuma motivação em servir como militar e o seu c contém diversos desejos que apontam para a direção exatamente oposta: ele nutre um profundo repulso por tudo o que é relacionado à vida militar. Indiferentemente aos desejos e comprometimentos de Owen, e mesmo contrariamente a eles, a família persiste em afirmar que ele tem uma razão para se alistar ao exército. Talvez ela possa acrescentar ao seu argumento o raciocínio de que é correto seguir as tradições familiares e que essa é a razão pela qual Owen deve se alistar. Nesse caso, a família faz um enunciado sobre razões externas que viola a restrição internalista acima. O externalista está disposto a assumir que, se se alistar no exército for a coisa certa a se fazer, então Owen realmente possui uma razão para respeitar a prescrição da sua família mesmo se não houver uma rota deliberativa adequada que o leve do seu c para a conclusão requerida²⁰⁸.

Como Williams apresenta, o externalista defenderia a tese de que enunciados de razão externa podem ser verdadeiros independentemente da motivação do agente: se Owen deliberar racionalmente, ele, independentemente da sua motivação inicial, poderia aceitar que possui uma razão para se alistar no exército e assim adquirir a motivação correspondente para agir de acordo com ela. O internalista típico, por outro lado, afirma que o agente apenas tem uma determinada razão para ϕ se essa razão estiver conectada com uma motivação prévia que contribua para ϕ ²⁰⁹.

No caso de Owen, há inúmeras situações em que ele pode vir a se alistar no exército. Por exemplo, a sua família poderia ameaçar retirar-lhe a herança e essa ameaça pode provocar o surgimento de uma motivação para servir como soldado. Porém, o externalista supõe que a aquisição de uma nova motivação pode ocorrer *porque* o agente veio a perceber, via reflexão, a verdade de um enunciado sobre razões externas. Williams não acredita que a reflexão tenha essa capacidade, e ele argumenta que o externalista não consegue explicar convincentemente como o reconhecimento da verdade de um enunciado de razões externas pode gerar a motivação necessária. Uma razão para a ação, segundo a posição de Williams, é uma razão que o indivíduo teria ou reconheceria se fosse racional, isto é, se refletisse adequadamente a partir do seu *set* motivacional. Em outros termos, é uma razão que, sendo o agente racional, figuraria na

²⁰⁸ WILLIAMS, 1981, p.107.

²⁰⁹ Gaus propõe uma formulação alternativa para demarcar essa divergência. Segundo ele, para o externalista razões são fatos do mundo ao passo que para o internalista razões são comprometimentos do agente. Por exemplo, um externalista diria que a razão para eu acreditar que Pelotas é uma cidade do estado do Rio Grande do Sul é o fato de que Pelotas é uma cidade do estado do Rio Grande do Sul. Contrariamente, um internalista diria que um fato do mundo é uma genuína razão apenas se a pessoa possuir outras razões para aceitar esse fato como uma razão (GAUS, 1996, p.35-38). Assim, a razão para eu acreditar que Pelotas é uma cidade do estado do Rio Grande do Sul é o fato de que *eu acredito* que Pelotas é uma cidade do estado do Rio Grande do Sul. Embora Gaus esteja aí se referindo a razões para crer e não em razões para agir, o exemplo é útil para ilustrar o internalismo de razões para a ação.

explicação de sua ação. Essa *dimensão explanatória* do conceito de razão para a ação não é, segundo Williams, preservada por um conceito de razão externa²¹⁰. Como resultado, o próprio conceito de uma razão externa deve ser rejeitado por ser incoerente: não faz sentido dizer que alguém tem uma razão para executar uma ação se, mesmo sendo racional, essa razão não pode figurar em uma explicação de sua ação. Quando alguém profere um enunciado de razão externa, portanto, tudo o que ele faz é uma “alegação interna otimista” de que, por exemplo, seria melhor se Owen se alistasse no exército²¹¹.

Porém, conforme Scanlon apresenta, Williams parece ter também um segundo argumento em mente. De acordo com esse outro argumento, a posição externalista viola uma dimensão de racionalidade²¹²: o externalista não pode sugerir que o agente que após a reflexão adequada não reconhecer que possui uma determinada razão para ϕ está sendo irracional em deixar de ϕ . O exemplo abordado tanto por Williams quanto por Scanlon é o de um marido que trata mal a sua esposa. Mesmo que ele deva (em algum sentido possível da ideia de dever) tratá-la bem, não se segue necessariamente que ele estará sendo irracional se tratá-la mal²¹³. No que concerne ao ERA, a restrição internalista parece funcionar como uma restrição de racionalidade que pode ser expressa da seguinte forma: ϕ é pessoalmente racional para S se e somente se S puder alcançar uma conclusão para ϕ em, ou após, ERA.

Com esse tipo de internalismo em mente, podemos entender em que sentido o projeto descritivo do ERA é descritivo-ideal. Daniels o denomina de descritivo-hipotético, mas a ideia básica é que se trata de uma descrição de juízos morais ponderados *em* ERA, não de juízos morais ponderados pré-reflexão. Mas se esse é o caso, poderia se objetar contra Rawls, qual a significância de se falar em “descrição”? Exagerando o lado ideal ou hipotético, prossegue a objeção, significa tornar a ideia de descrição vazia: qualquer teoria moral “descreve” um estado de coisas ideal. Eu penso que essa objeção é completamente acertada. O que distingue o ERA de (algumas) outras abordagens é que ele busca um *input* para a teoria no modo como as pessoas efetivamente deliberam. O *input* é descritivo e o *output* é (pretensamente) normativo. Esse é o espírito do método logo quando Goodman o formulou pela primeira vez.

Eis como eu vejo que é possível manter a significância do conceito de descritividade. A justiça como equidade tem de descrever um número mínimo de juízos morais ponderados prévios ao ERA. Suponha que antes de se engajar no ERA, S possui o conjunto de crenças c_1 ,

²¹⁰ WILLIAMS, 1981, p.106.

²¹¹ WILLIAMS, 1981, p.111.

²¹² WILLIAMS, 1995, p.39.

²¹³ SCANLON, 1998, p.365.

c2, c3, c4, c5, e que ao final do ERA, S possui o conjunto de juízos morais ponderados v1, v2, v3, v4, v5. É possível que o conjunto de crenças pré-ERA, c, não interfira no conjunto v, afirmado pós-ERA? Se sim, como S poderia reconhecer que uma determinada descrição corresponderia ao seu senso de justiça idealmente configurado, se antes do processo ele não possuía nenhuma crença moral que o indique isso? Nesse caso, como S poderia ter sequer um interesse em seguir a teoria moral e utilizar o ERA? Se a justiça como equidade pretende descrever o que um cidadão democrático poderia aceitar, ela precisa descrever algo que esse cidadão, em um sentido relevante, já aceita²¹⁴. E parece óbvio que ela faz isso justamente porque ela se limita a apenas descrever o que um cidadão especificamente *democrático* poderia reflexiva e livremente aceitar.

A interpretação que Daniels faz dessa ideia de descritividade corre seriamente o risco de não fazer jus ao espírito internalista do método, pois ele reconstrói *toda* a justiça como equidade a partir de uma linguagem contrafactual (como uma teoria que descreve o que aceitaríamos, julgaríamos, endossaríamos, se nos engajássemos no ERA). Daniels defende que o ERA não busca caracterizar uma competência moral real e que seria perverso imaginar que todas as revisões que o ERA pode promover é uma questão de explicar crenças implícitas já aceitas²¹⁵. Daniels aparentemente entendeu que esta é a única saída argumentativa para mostrar que o ERA não é descritivista e conservantista. O problema é que esse argumento vem a um custo muito alto, um custo que é razoável supor que Rawls não estaria disposto a pagar: implica anular ou tornar irrelevante o *input* do modo como as pessoas *realmente* realizam inferências. Como visto no capítulo passado, essa é a característica essencial do equilíbrio reflexivo tal como Goodman o formulou. Se esse *input* é anulado, a revisibilidade promovida pela teoria é utópica ou arbitrária. No cenário sugerido por Daniels, a própria teoria é externa às razões que o indivíduo tem para ser motivado em iniciar a deliberar moralmente.

Rawls utiliza frases condicionais contrafactuais para se referir à justiça como equidade, mas nem sempre. Por exemplo, em um dado momento ele defende que a justiça como equidade

²¹⁴ Esse me parece um ponto que Rawls reconhece e enfatiza em alguns momentos, como quando ele afirma que se o senso comum é “hesitante e não sabe o que pensar” a teoria moral deve propor “princípios afins com as suas convicções mais essenciais e tradições históricas” que constituem “uma profunda base de acordo” (KC, p.306). Por outro lado, ele presumivelmente se contradiz em outros contextos, como no *IMT* quando ele constrói o projeto descritivo utilizando apenas frases contrafactuais: “nós estamos interessados em que concepções as pessoas *afirmariam*”, nós investigamos que princípios as pessoas *reconheceriam*”, “alguém busca a concepção, ou a pluralidade de concepções, que *sobreviveriam*” (*IMT*, p.289). Rawls sublinhou esse aspecto porque estava pressionado pela objeção de que a sua teoria seria conservadora “por se limitar àquilo que as pessoas agora sustentam”. Rawls pode argumentar que não se limita àquilo que as pessoas sustentam, mas ele não pode alegar (como essa linguagem contrafactual sugere) que ignora totalmente todas as crenças que alguém sustenta. Nós aceitaríamos a justiça como equidade porque nós *efetivamente* aceitamos alguns valores morais que ela descreve.

²¹⁵ DANIELS, 1996, p.71.

pode ser aceita “porque as condições incorporadas na posição original são tais que *de fato aceitamos*”, embora complementemente com o seguinte: “ou se não aceitamos, então talvez *poderíamos* ser persuadidos a aceitá-la pela reflexão filosófica”²¹⁶. Como a reflexão filosófica pode nos persuadir sem mobilizar *nenhuma* crença que de fato já aceitamos? Penso que se não houver essa contrapartida factual, mínima que seja, a própria noção de justificação é comprometida. Rawls na maior parte das vezes confirma sem espaço para dúvidas esse ponto, como quando defende que “aqueles que não sentem afinidade pela noção de uma sociedade bem-ordenada”, “não serão movidos pela justiça como equidade (mesmo admitindo a validade do argumento)”²¹⁷. A consequência desse excerto é que se nós já não aceitamos alguma crença mobilizada pela justiça como equidade, então não poderíamos ser persuadidos pela reflexão filosófica a endossá-la.

Isso não significa que as crenças alcançadas pós-ERA são completamente determinadas pelas crenças pré-ERA (o que não é o mesmo que dizer que o ERA é capaz de impedir esse resultado, como eu discutirei na seção dedicada ao problema do descritivismo no capítulo seguinte). A corrigibilidade das crenças básicas, mais o fato de que o ERA é um processo contínuo sem fim, bloqueia essa consequência. Suponha uma série de revisões: S pode fazer uma primeira revisão no seu conjunto de crenças pré-ERA, c1, c2, c3, c4, c5, e manter apenas uma crença original, sustentando então o conjunto c1, v2, v3, v4, v5. Como o ERA é um

²¹⁶ “The conditions embodied in the description of the original position are ones that we do in fact accept. Or if we do not, then perhaps we can be persuaded to do so by philosophical reflection” (*TJ*, p.21/rev, p.19).

²¹⁷ *RAM*, p.236. De uma maneira geral, o tópico que eu aqui abordo foi declaradamente um problema que Rawls não estava disposto a resolver. Em *TJ* ele escreve o seguinte: “Uma vez que nós entendemos a teoria, nós podemos reconhecer essas interpretações como adequados desdobramentos do que em reflexão nós desejamos sustentar. Embora nós não usaríamos normalmente esses substituidores, talvez porque eles são muito complexos, ou seriam mal entendidos, ou qualquer outra coisa, nós estamos preparados para admitir que eles cobrem em essência tudo aquilo que se quer que seja dito. Certamente esses substitutos podem não significar o mesmo que os juízos ordinários com que eles são emparelhados. Até que ponto esse é o caso é algo que eu não examinarei. Além do mais, essas substituições podem indicar uma mudança mais ou menos drástica nos nossos juízos morais iniciais tal como eles existem antes da reflexão filosófica”. (“Once we understand the theory, we can acknowledge these interpretations as suitable renderings of what on reflection we now wish to maintain. Even though we would not normally use these replacements, perhaps because they are too cumbersome, or would be misunderstood, or whatever, we are prepared to grant that they cover in substance all that wants to be said. Certainly these substitutes may not mean the same as the ordinary judgments with which they are paired. How far this is the case is a question that I shall not examine. Moreover, the replacements may indicate a shift more or less drastic from our initial moral judgments as they existed prior to philosophical reflection”) (*TJ*, p.451; rev, p.396). Essa passagem é, no limite, curiosa. Há claramente no centro do projeto justificacional de Rawls a ideia de uma aceitação em reflexão filosófica, isto é, aceitação em ERA. A despeito da centralidade dessa ideia, que é uma ideia que supõe uma metodologia moral sobre como deliberar moralmente, Rawls resiste em trabalhá-la em detalhe (seja em *TJ*, seja nos escritos posteriores). Para Rawls, os caminhos que levam à aceitação em ERA são pormenores. O bastante, para o seu argumento, é analisar o resultado da reflexão: se a justiça como equidade seria ou não aceita nessas condições. O problema, contudo, é que se trata sempre de um curso hipotético de reflexão de uma pessoa real. Se esse curso não for detalhado, como ele não é, sempre haverá dúvidas a respeito de qual seria exatamente o seu resultado. Não é incorreto objetar que Rawls não oferece razões suficientes para mostrar que a sua interpretação dos juízos morais ponderados é a melhor interpretação possível desses juízos ou, equivalentemente, que a justiça como equidade efetivamente seria aceita em ERA.

processo contínuo indefinido, S pode fazer uma segunda revisão e recusar c1 em nome de outra crença, b1, extraída das crenças v2, v3, v4 e v5, crenças que não estavam presentes antes da primeira revisão. Nessa fase S terá alcançado o sistema de crenças b1, v2, v3, v4, v5, completamente distinto do seu sistema c1, c2, c3, c4, c5 original. Esse é um exemplo bastante esquemático, poderíamos imaginar sucessivas e incontáveis “revisões”, ocorrendo em processos dinâmicos e não divisíveis. Basta destacar que, para os propósitos que Rawls nutre com o ERA, S precisa ter alguns juízos morais ponderados específicos antes de iniciar o processo e, por isso, ele não pode descrever apenas juízos morais ponderados *em ERA*²¹⁸.

A minha insatisfação com essa construção inteiramente contrafactual da justiça como equidade é, ao menos eu penso, uma que Williams poderia endossar: se não houver um *link* entre as motivações que alguém já sustenta antes de começar a refletir sobre moralidade e as motivações que ele adquire após a reflexão, nós não conseguimos explicar como que alguém veio a adquirir a nova motivação em questão. Talvez Daniels presuma que se S seguir adequadamente o ERA, ele poderia, *independentemente das crenças com que iniciou o processo*, “ver” a correção da justiça como equidade e assim adquirir a motivação necessária²¹⁹. Como esclarecido por Thomas, é isso o que externalista faz: ele parte da verdade ou correção de uma razão externa e, via deliberação racional, para a sua aceitação e geração de um motivo. O internalista adota o caminho oposto: ele parte do *set* motivacional inicial e, via deliberação racional, para os enunciados de razão interna verdadeiros para aquele agente²²⁰. Alguns internalistas de razões, como Rawls e Williams, estão felizes em admitir que o *output* da

²¹⁸ Alguém poderia questionar se a legitimidade epistêmica dos juízos pré-ERA determina a legitimidade das crenças pós-ERA. Essa é uma questão diferente e complexa, sobre a qual apenas posso conjecturar. No meu exemplo acima, S aceita o conjunto v em virtude de c1, e depois rejeita c1 a partir de uma crença b1 originária das crenças v. Se a crença c1 não for ela mesma justificada, então o conjunto v será ele mesmo injustificado, e por consequência b1 também terá o mesmo destino, pois ela está indiretamente relacionada com c1. Ao mesmo tempo, creio que esse tipo de raciocínio pode levar a resultados absurdos. Se as minhas crenças atuais, não importa qual seja o seu conteúdo, dependem da legitimidade epistêmica das crenças prévias que, embora agora inexistentes, me permitiram formular meu novo conjunto de crenças, então eu tenho de, seguindo a lógica desse raciocínio, supor que a legitimidade das crenças prévias às minhas crenças prévias determina diretamente a legitimidade destas últimas e, por implicação, indiretamente a legitimidade das minhas crenças correntes. Mas se esse é o caso, então nós temos de chegar à conclusão de que a legitimidade de qualquer crença moral sustentada hoje depende indiretamente da legitimidade das crenças morais que eu mesmo possuía aos dez anos de idade. Apesar do aparente absurdo envolvido, creio temos de reconhecer que o valor epistêmico das crenças pós-ERA é ao menos fortemente dependente do valor das crenças pré-ERA. O meu modelo simplista e reduzido de um número baixo de revisões parece fazer mais sentido do que um modelo que imagina revisões potencialmente infinitas.

²¹⁹ Penso que também não faz sentido afirmar, como Brink, que “é possível, ao menos em princípio, para aqueles que entram no processo com crenças morais completamente diferentes convergir com as mesmas crenças morais” (BRINK, 1989, p.143). Rawls estava bastante ciente dessa impossibilidade quando afirmou que quem não aceitar a ideia de sociedade bem-ordenada não será movido pela justiça como equidade (RAM, p.236). É também por isso que ele não destina a justiça como equidade para todos os seres capazes de razão prática, mas somente para aqueles que compartilham certas crenças básicas típicas de um regime democrático.

²²⁰ THOMAS, 2006, p.70.

reflexão é condicionado pelos seus pontos de partida²²¹. Presumivelmente porque as razões para a ação têm sempre de dizer algo de distintivo sobre o agente particular aos quais elas se aplicam²²².

Não estou afirmando que o elemento contrafactual não está presente. De fato, a justiça como equidade descreve o que as pessoas *afirmariam* após o ERA, não o que elas afirmam antes. Porém, na medida em que Rawls em *TJ* constrói a aceitação da justiça como equidade como orientada pelos nossos pontos fixos provisórios a partir dos quais as concepções-modelo são inferidas, ele parece reconhecer que a justiça como equidade não poderia ser aceita por alguém que não compartilha desses pontos fixos. O que há de contrafactual é o seguinte: é possível afirmar que alguém tem uma razão (interna) para aceitar a justiça como equidade mesmo se essa pessoa antes do ERA não endossa a teoria proposta por Rawls, ou o seu *set* motivacional antes do ERA a faz agir conforme uma prescrição contrária ou até mesmo se ela acreditar que possui uma razão para não agir conforme a justiça como equidade. Inúmeros fatores podem contribuir para esse resultado, sendo o maior deles a incoerência interna de um senso de justiça. Se um senso de justiça estiver em ERA, então a justiça como equidade descreve as razões que o indivíduo *de fato* aceita; se esse senso de justiça ainda não estiver em ERA, o que é mais provável, então ela descreverá o que ele *deve* (no sentido de é racional que) aceitar ou o que ele *aceitaria*. Porém, apesar de haver esse nível relevante de contrafactualidade, a restrição internalista é respeitada porque o que alguém *deve* acreditar é aquilo que alguém pode afirmar após tomar uma rota deliberativa adequada que leva do seu *set* motivacional para a aceitação em questão. No caso de Rawls, que leva dos seus pontos fixos provisórios para a aceitação em questão. Se aplicado ao caso de Owen Wingrave, o ERA implicaria que ele não tem uma razão para se alistar no exército²²³. Se aplicado a pessoas que não têm os pontos fixos

²²¹ Neal propõe uma distinção entre dois níveis de comprometimento que julgo ser útil para esclarecer este ponto. Por um lado, a teoria depende de comprometimentos em um primeiro nível que podem ser vistos como comprometimentos de fato empiricamente enraizados. Esses são os comprometimentos efetivamente descritos pela justiça como equidade. Por outro lado, segue Neal, a teoria também articula comprometimentos de segundo nível, que são resultado de uma reflexão individual a partir dos comprometimentos de primeiro nível. Esses comprometimentos de segundo nível são condicionais contrafactuais por depender da reflexão. Pode-se considerar os princípios de justiça como sendo desse segundo tipo. Neal chama atenção que para Rawls o consenso no primeiro nível é assumido como ponto de partida para a tarefa realmente necessária: buscar o consenso no segundo nível. Sem aquele primeiro consenso, a justiça como equidade da forma que a conhecemos não existiria. Esse tipo de constatação é interessantemente sublinhada pelo autor. Ele a expressa na ocasião em que está criticando aqueles que tendem a ler como se Rawls estivesse supondo que as pessoas se veriam como livres e iguais *se* seguissem a teoria. Para Neal, tudo o que a teoria faz é permitir às pessoas que elas percebam que *em virtude do que elas já acreditam*, elas devem sustentar os comprometimentos de segundo nível que a teoria recomenda (NEAL, 1994, p.92).

²²² Conforme esclarece THOMAS, 2006, p.83.

²²³ Como Scanlon identificou, é em virtude dessa característica que o internalismo proposto por Williams sofre muitas críticas: ele supõe que “as nossas próprias razões são todas contingentes relativamente à presença de elementos apropriados em nosso conjunto motivacional subjetivo” (SCANLON, 1998, p.367). Mas essa é uma

provisórios de cidadãos democráticos contemporâneos, o ERA implicaria que essas pessoas não têm razão para afirmar a justiça como equidade. E Rawls estava ciente disso ao não aplicar o conteúdo da justiça como equidade no âmbito internacional.

Williams também admite espaço para níveis de revisibilidade e contrafactualidade. Para ele, enunciar que “S tem uma razão para ϕ ” não é simplesmente dizer que “S está presentemente disposto a ϕ ”, do que se segue que uma posição internalista é “normativa” por buscar descrever o que as pessoas deveriam estar dispostas a fazer²²⁴ e por estar preocupada com a “racionalidade do agente”²²⁵. Porém, Williams nos alerta para o nível de contrafactualidade que devemos permitir e como devemos ver a conexão entre o que há antes da reflexão e o que pode haver depois. De uma forma evidentemente clara, a reflexão deve corrigir o que havia antes. Mas essa correção é limitada porque ela deve, para respeitar a restrição internalista, se limitar a erros de fato e de raciocínio cometidos pelo agente, e não erros ou deficiências morais²²⁶.

Como eu o interpreto, o ERA também está comprometido apenas com a correção de erros de raciocínio e de fato²²⁷. Como Rawls diz, “nós estamos procurando por um modo de remover nossas dúvidas” e por uma concepção que “acomode nossas convicções mais firmes e forneça orientação quando a orientação é necessária”²²⁸. Metodologicamente, o método parece funcionar especificando condições para a deliberação que não pressupõem quaisquer considerações morais. Ele prescreveria que um determinado indivíduo tem de ranquear os seus comprometimentos segundo o nível de confiança, identificar aqueles que sustenta com mais certeza e considerar uma diversidade de concepções de justiça para avaliar qual oferece a articulação mais satisfatória desses juízos. De posse desses critérios, ele deve revisar os seus outros juízos menos fixos para ajustá-los aos critérios oriundos dos mais fixos. Se ele não conseguir completar esse último passo, então ele deverá ou ponderar a revisão dos seus pontos fixos, ou reconsiderar a descrição das ideias básicas subjacentes a eles ou mesmo ambas as coisas ao mesmo tempo. O ponto final do processo é um estado de equilíbrio que o indivíduo

constatação em terceira pessoa. Quando Owen conclui que não tem uma razão para se alistar ao exército, ele pode crer (e o internalismo ou o ERA não se opõem a esse resultado) que ele não tem *porque* a vida militar é imoral. Ele não precisa compreender que a sua conclusão (vida militar é imoral) é condicionada pelo seu *set* motivacional. Nada impede que Owen veja essa conclusão como uma verdade moral *a priori*, dessa forma compreendendo (erradamente) a sua razão para não ceder às pressões da sua família como uma razão externa.

²²⁴ WILLIAMS, 1995, p.36.

²²⁵ WILLIAMS, 1981, p.103.

²²⁶ WILLIAMS, 1981, p.101-104.

²²⁷ Com isso não quero me comprometer com a tese de que o ERA faz isso competentemente e que, como está agora apresentado por Rawls, seja suficiente para corrigir todos esses erros.

²²⁸ “Here we may be looking for a way to remove our doubts. We can check an interpretation of the initial situation, then, by the capacity of its principles to accommodate our firmest convictions and to provide guidance where guidance is needed” (TJ, p.20/rev, p.17).

em questão pode ver como aceitável. Note que nenhuma restrição moral é operante nesse retrato (bastante grosseiro) do método.

Para explicar essa característica, é usual definir um internalismo como o de Williams como supondo uma concepção procedimental de razões para ação: o indivíduo tem uma razão para ϕ apenas se ao seguir um determinado procedimento (ERA) ele alcançar uma conclusão para essa razão. Em oposição, um externalista teria uma concepção substantiva de razões para ação (e, conseqüentemente, de racionalidade): o indivíduo tem uma razão para ϕ apenas se ao seguir um determinado procedimento ele alcançar uma conclusão para essa razão e ϕ for uma ação correta, verdadeira ou em si mesma valorosa²²⁹. Uma concepção procedimental não afirma que há um conteúdo moral específico que deve ser reconhecido por uma pessoa racional. Ser racional é apenas uma questão de deliberar corretamente, enquanto que para uma concepção substantiva de racionalidade, ser racional é, além de deliberar corretamente, reconhecer um determinado conteúdo como racional – em que a racionalidade do conteúdo é independente do procedimento²³⁰.

Ao abandonar o sonho filosófico caricaturado por Daniels, e utilizar uma concepção procedimental de racionalidade, Rawls abdica da pretensão de afirmar que todas as pessoas que não sustentam a justiça como equidade são pessoalmente injustificadas ou irracionais, e, mais importante, não atrela o mérito da sua concepção de justiça à possibilidade de ela ser aceita por qualquer indivíduo capaz de racionalidade prática²³¹. Note que isso se explica em função da contrafactualidade permitida estar limitada apenas à correção de erros de fato e de raciocínio, de modo que considerações éticas não compõem o próprio conceito de ERA e de deliberação prática adequada. Para alguém estar justificado em ERA, basta fazer as conexões corretas entre as suas crenças. O conteúdo da justiça como equidade não é um critério de racionalidade, a despeito de que todo o cidadão democrático contemporâneo o afirmaria em ERA.

Como objeção à caracterização do ERA que proponho, alguém poderia lembrar das restrições do ponderado, especialmente da restrição segundo a qual se deve excluir juízos realizados com uma excessiva atenção ao autointeresse, e então argumentar que o ERA embute no próprio conceito de racionalidade prática considerações morais normativas em favor da imparcialidade. Não me oponho a essa linha de raciocínio e concordo com ela: se as restrições do ponderado fizerem parte da definição de ERA, então o ERA não consegue cumprir o seu

²²⁹ PARFIT, 2011, p.61-63.

²³⁰ AUDI, 2001, p.79-80.

²³¹ É por isso que em *The Law of Peoples* (1999) Rawls não estende o conteúdo da justiça como equidade para a aplicação no âmbito internacional. Diferentemente de *TJ* ou *PL*, a aplicação *The Law of Peoples* não é limitada a povos cujos pontos fixos provisórios dos seus habitantes podem ser representados pelas concepções-modelo.

objetivo de assegurar a racionalidade pessoal porque pessoas que empregam o ERA e não são imparciais sempre terão como *output* um resultado que lhes é irracional. Nesse caso o egoísta moral é considerado por Rawls alguém que é irracional e seria verdadeiro dizer que para Rawls o egoísta tem uma razão para ser imparcial mesmo se ele não puder alcançar essa conclusão via uma rota deliberativa adequada a partir do seu *set* motivacional. Eu penso que esse resultado é desastroso para o ERA porque ele viola a restrição internalista. Suponho que é por esse motivo que o ERA seria incapaz de cumprir o papel de processo deliberativo adequado na tese do Williams.

No primeiro capítulo desta dissertação esse problema já havia sido identificado e uma solução apresentada. A solução, proposta por DePaul, afirma que as restrições do ponderado deveriam ser justificadas em ERA com referência a princípios morais e epistêmicos do seu utilizador. Contanto que esses princípios morais e epistêmicos sejam relativizados ao *set* motivacional de cada agente, esse reparo *ad hoc* de DePaul é satisfatório. Ele permite que qualquer agente moral, independentemente do seu *set* motivacional, possa se engajar no ERA. Alternativamente, podemos extrair outra solução igualmente satisfatória, ainda que mais simples, de um ensaio em que Scanlon comenta brevemente o equilíbrio reflexivo. Nesse ensaio, Scanlon escreve que o que conta como um juízo ponderado deve ser selecionado meramente pelo nível de confiança com que o agente sustenta o juízo. Os não-ponderados seriam aqueles sustentados, do ponto de vista de quem emprega o ERA, com um nível insatisfatório de confiança²³². As restrições do ponderado seriam apenas a estipulação do que as pessoas em geral tomam como circunstâncias inconfiáveis que geram juízos inconfiáveis, mas não necessariamente contariam como uma exigência para o emprego do ERA: cada ERA pode definir diferentemente o que é e o que não é ponderado.

Partindo desse comentário de Scanlon, penso que nós podemos descartar completamente a ideia de restrições do ponderado como algo que é fixado externamente ao agente. Rawls já admite que os juízos ponderados devem ser revisados ao longo do ERA mediante o nível de confiança com que são sustentados. Essa consideração me parece suficiente para excluir os juízos não-ponderados também. Pense nos juízos realizados com excessiva atenção ao autointeresse: a aposta do proponente do ERA deve ser a de que o agente descartaria juízos demasiadamente parciais porque reconheceria que esses juízos são distorcidos por circunstâncias que o impedem de realizar um juízo moral que dá voz às suas convicções morais realmente mais fundamentais. Porém, se o agente não reconhecer a inconfiabilidade desses

²³² SCANLON, 1992, p.9

juízos, ou, mais dramaticamente, se ele possuir convicções morais fundamentais para não reconhecê-la, então não é claro como esse juízo, apesar de violar a restrição de imparcialidade, pode ser considerado não-ponderado.

Então, contrariamente ao que quer Rawls, os juízos morais não-ponderados não precisam nem devem ser excluídos antes do ERA sequer iniciar. O leitor pode agora dizer que é assim que as coisas devem ser, mas não é assim que Rawls as concebe. De fato, mas essa qualificação é necessária e embora não seja de Rawls, segue dos seus pressupostos e, aceitando que o internalismo de razões é preferível ao externalismo, torna a sua teoria melhor. O próprio Rawls em um dado momento escreve que “sabendo que pessoas são racionais nós não sabemos os fins que elas perseguirão, somente que elas os perseguirão inteligentemente”²³³. Essa própria frase estabelece a natureza procedimental do conceito de racionalidade²³⁴.

Mesmo considerando esses fatos, creio que não precisamos ser tão severos com Rawls. Como desenvolverei no capítulo seguinte, o seu objetivo é utilizar o ERA para mostrar que a justiça como equidade pode ser pessoalmente justificada para pessoas como nós, que são pessoas que aceitam o valor da imparcialidade. Para pessoas como nós a arbitrariedade das restrições do ponderado não são um impeditivo para obter um *output* racional com o uso do método. Esse tipo de comentário não o inocenta, claro, mas diminui os danos²³⁵.

2.5 Amplo versus estreito

Creio que uma maneira útil de sintetizar os resultados alcançados até este ponto é fazer uma comparação entre o ERE e o ERA.

Não raro é sustentado que o ERA representa um grande distanciamento em relação ao ERE porque o segundo consistiria na busca de coerência entre juízos morais ponderados e princípios de justiça, enquanto que o primeiro, mais plausivelmente, acrescentaria um terceiro elemento que atenuaria a dependência da teoria em relação aos juízos morais ponderados. Essa demasiada dependência do ERE foi vista por Daniels como problemática, e nós não temos razões para pensar que Rawls não subscreve essa maneira de ver as coisas. Assim como Daniels, ele também enfatiza que o ERA é “filosoficamente motivado” justamente porque ele busca por

²³³ “Knowing that people are rational we don’t know the ends they will pursue, only that they will pursue them intelligently” (PL, p.49n1).

²³⁴ Parfit utiliza exatamente esse excerto em suporte à alegação de que Rawls é um subjetivista sobre razões, isto é, de que Rawls defenderia que as razões de um agente são dadas pelos elementos subjetivos desse agente (como seus desejos, objetivos, ou comprometimentos) (PARFIT, 2011, p.78).

²³⁵ Thomas é um autor que utilizou o ERA e esteve sensível a essa dificuldade de relacionar racionalidade prática e imparcialidade. Como destacado por Thomas, o internalismo é violado quando se afirma que a imparcialidade é constitutiva do que significa ser racional praticamente. Ele tentou acomodar a imparcialidade com as restrições internalistas de Williams no capítulo quarto do seu livro *Value and Context* (2006).

descrições alternativas ao senso de justiça, descrições que supostamente envolveriam o apelo a algo além do mero equilíbrio entre juízos e princípios característico do ERE²³⁶. Essa também é uma das alegações de Rawls (assim como de Daniels) contra a crítica de que o ERA seria conservador no sentido de não dispor de meios para superar um dado *status quo* moral estabelecido por nossos juízos morais ponderados. O acréscimo de teorias morais alternativas, que o indivíduo deveria considerar quando estivesse buscando pela melhor descrição do seu senso de justiça, conduziria então a uma revisibilidade mais ampla do seu senso de justiça²³⁷. Assim, o ERA possuiria duas vantagens em relação ao ERE: (i) o ERA não dependeria tão fortemente da confiabilidade dos juízos morais ponderados com que se inicia o processo, assim permitindo que nós tenhamos *outputs* apropriados com base em *inputs* inapropriados; e (ii) os princípios resultantes do método são testados não apenas contra os juízos ponderados, mas também contra concepções-modelo organizadas pela posição original, dessa forma havendo um duplo suporte (juízos mais teorias) para o *output* do método.

Pace Daniels e Rawls, essas diferenças não são suficientes para colocar o ERA em uma posição mais favorável que a do ERE²³⁸. Considere (i) primeiro. Na seção anterior foi defendido que o ERA necessariamente precisa descrever algo que as pessoas sustentam pré-ERA, e que a justiça como equidade não pode ser construída com uma linguagem inteiramente contrafactual, pois ela apenas poderá ser aceita por aqueles que já aceitam determinadas crenças morais. Isso mostra que tanto o ERE quanto o ERA dependem dos *inputs* no sentido de que, se em determinadas circunstâncias um dado indivíduo não possuir um *input* adequado, então o emprego do ERE ou do ERA não fará com que esse indivíduo tenha um *output* drasticamente superior. Especificamente, se eu não aceitar as concepções-modelo que compõem a posição original, então eu não terei razões para tomá-las como uma autoridade para mim e, por conseguinte, não adquirirei a motivação necessária para agir de acordo com elas. No seu último ensaio sobre o equilíbrio reflexivo, Daniels explicitamente defende que Rawls abandonou o sonho filosófico de supor que argumentos filosóficos convenceriam alguém a aceitar a posição original e a justiça como equidade²³⁹. Sendo assim, o filósofo está refém dos *inputs* da reflexão

²³⁶ *IMT*, p.289.

²³⁷ *IMT*, p.289 e DANIELS, 1996, p.27.

²³⁸ Que não há diferença entre o ERE e o ERA foi também defendido por Holmgren (1989). O ponto básico para Holmgren é que, uma vez que se aceita que (i) os nossos juízos morais ponderados têm de possuir uma credibilidade *prima facie* (isto é, o *output* do método apenas será adequado se o seu *input* também o for), e que (ii) as teorias morais derivam a sua credibilidade do fato de que elas sistematizam adequadamente esses juízos, então não sobra espaço para demarcar uma diferença substancial entre o ERE e o ERA. Estou inteiramente de acordo com o ensaio de Holmgren. No que respeita à alegação (i), a defenderei no capítulo seguinte.

²³⁹ Como o próprio Daniels admite no seu ensaio mais recente, o seu primeiro sentimento ao entrar em contato com os argumentos de Rawls após *TJ* foram de decepção em virtude da desprestigiabilidade filosófica implicada

e o ERA não gera, para usar a expressão de DePaul, “ouro a partir do chumbo”. Para DePaul, “quais revisões são necessárias é ditado pelas relações lógicas e evidenciais entre as crenças de *input*”, e que revisões serão efetivamente feitas “é virtualmente determinado pelo nível inicial de comprometimento da pessoa com as suas crenças de *input*”²⁴⁰. Como o próprio Daniels esclarece mais recentemente, fundamentalistas não afirmarão a justiça como equidade em ERA e não serão irresponsáveis epistemicamente em rejeitá-la²⁴¹.

O argumento originalmente apresentado por Daniels em favor de (i) também depende da sua leitura coerentista do método, mas isso eu tratarei apenas no capítulo seguinte. Agora considere (ii). A despeito do empenho de Daniels, não vejo em que sentido pode efetivamente haver um duplo suporte²⁴². Para Rawls está claro que o novo componente do ERA é a posição original e eu me esforcei para mostrar que a sua função justificatória consiste em elencar certas restrições morais em uma fórmula intuitivamente apelante. Considerando que essas restrições morais nada mais são do que sistematizações de nossos pontos fixos provisórios, aqueles juízos que sustentamos com mais comprometimento e estamos mais indispostos a descartar, o ERA não oportuniza um mediador externo à relação de reciprocidade entre juízos e princípios ou um critério independente dessa relação.

A motivação para as alegações (i) e (ii) é responder diretamente às objeções abordadas no capítulo anterior, assim fazendo com o que o ERA não seja vitimado pelas mesmas fraquezas

no desencorajamento do sonho filosófico. Contudo, Daniels admite que essa foi apenas a sua reação inicial, pois depois ele percebeu que Rawls nunca nutriu aquele sonho (DANIELS, 1996, p.151).

²⁴⁰ DEPAUL, 1993, p.49.

²⁴¹ DANIELS, 1996, p.167. Eu creio que há uma tensão entre os ensaios clássicos de Daniels e o seu ensaio mais recente de 1996 voltado à caracterização do equilíbrio reflexivo no *PL*.

²⁴² Como se sabe, Daniels define o ERA como um método para produzir coerência entre (a) juízos morais ponderados, (b) princípios e (c) teorias de fundo (*background theories*) ou o que Rawls denomina de concepções-modelo. O elemento-chave da sua definição é o postulado de que (c) deve possuir um suporte epistêmico diferente de (b). Para a diferença entre o ERE e o ERA ser relevante, defende Daniels, (c) tem de satisfazer a seguinte “restrição de independência” (*independence constraint*): o conjunto de juízos morais ponderados que justifica (c) tem de ser distinto daquele caracterizado por (b) (DANIELS, 1996, p.22-26). Assim, os princípios de justiça seriam legitimados por se tratarem de descrições de um conjunto (a) de juízos morais ponderados, e a posição original seria justificada por depender de um conjunto diferente, chamado por Daniels de conjunto (a’). Daniels tentou mostrar que a concepção de pessoa, uma das *background theories*, apenas satisfaz a restrição de independência se for possível mostrar que os juízos morais que a legitimam contém uma linguagem distinta daqueles dos quais os princípios são derivados. Assim, a concepção de pessoa não poderia invocar juízos morais ponderados que empregam um vocabulário de direitos, pois esse é o vocabulário característico dos princípios (DANIELS, 1996, p.56-57). Mas o que acontece é que ela emprega esse vocabulário. A concepção de pessoa reflete, a um só tempo, o tipo de pessoas que nós queremos ser e que julgamos que temos o direito de ser. Daniels não pode operar aqui com uma dicotomia entre direitos e deveres, porque me parece claro que para Rawls os deveres informam quais são os nossos direitos ao mesmo tempo em que os direitos moldam os nossos deveres. A distinção entre dois tipos de juízos morais ponderados nesses termos é ilusória (no capítulo seguinte eu proporei uma distinção entre dois tipos de juízos morais ponderados, mas será uma distinção completamente diferente da de Daniels e motivada por razões igualmente discrepantes das suas). Porém, mesmo se considerássemos que tal dicotomia é viável, nada altera o fato de que os juízos morais ponderados, seja (a) ou (a’), ainda são *nossos* juízos. Considero a restrição de independência uma proposta ininteligível, principalmente quando nos transportamos para o ponto de vista de quem emprega o ERA. Nesse ponto de vista, que sentido faz essa distinção?

do ERE. Se esses são os argumentos, então o ERA definitivamente falha também. Mas essa não é a minha conclusão, contudo. Neste capítulo foi defendido que o ERA funciona na primeira pessoa, e que essa característica não é compartilhada pelo ERE. Partindo dessa ideia, no capítulo seguinte será defendido que o ERA representa um genuíno avanço em relação ao ERE.

3. Justificação em Equilíbrio Reflexivo Amplo

Na parte final do capítulo primeiro foram apresentadas duas objeções ao ERE cujo mérito dependia exclusivamente das pretensões do *Outline*. É porque Rawls pretende traçar (1) uma distinção objetiva entre o certo e o errado que especifique as condições sob as quais há e não há *conhecimento moral* que ele tem de responder a objeção do conservantismo. É porque ele defende (2) a normatividade de princípios justificados *de acordo com a sua concepção de justificação* que ele precisa afastar a objeção do descritivismo. Para efetivamente analisar como essas objeções se aplicam ao ERA, é preciso, antes, identificar se ele nutre as mesmas pretensões que o ERE. O objetivo central deste capítulo é fazer isso. Para tanto, ele é dividido em duas seções. Na primeira parte é defendido que o ERA deve ser visto como uma forma de fundacionalismo moderado que supõe exclusivamente o que pode ser denominado de uma concepção de justificação pessoal. Na segunda parte é argumentado que assim entendido o ERA não sucumbe à objeção (1) e que, com respeito à (2), o ERA não tem como superá-la de um modo que satisfaça os seus críticos, mas é deixado em aberto se o descritivismo realmente precisa ser visto como uma falha.

3.1 A posição epistemológica

3.1.1 O conceito geral de justificação

Antes de investigar que tipo de posição epistemológica subjaz ao ERA é necessário legitimar a própria pergunta. Rawls escreveu um ensaio inteiro (*IMT*) para mostrar que a teoria moral é independente da filosofia moral, isto é, metodologicamente indiferente a questões epistêmicas, semânticas e de filosofia da mente (é um mistério por que Rawls não incluiu também a ontologia aqui). Muitos podem pensar que o *IMT* deslegitima uma investigação metaética sobre a obra de Rawls. Como esta dissertação é exatamente uma investigação desse tipo, tenho de, ao menos preliminarmente, defendê-la. A defesa completa é a própria dissertação em si.

A tese do *IMT* é que o ERA é independente da epistemologia porque ele “não admite que há uma concepção correta”. Supor que há uma concepção correta, Rawls esclarece, equivale a aceitar a existência de “verdades morais objetivas”²⁴³. Ao longo do *IMT*, Rawls implicitamente mostra entender epistemologia como sinônimo de teoria do conhecimento. O que nós temos de perguntar agora é se essa desconexão entre teoria moral e verdade moral é suficiente para a alegação de que a justiça como equidade é independente da epistemologia. Não acredito que seja. A teoria moral busca mostrar que certas concepções de justiça são *justificadas*, e o próprio Rawls se empenhou em mostrar que quando o nosso senso de justiça é coerente (justificado) ele suporta a justiça como equidade. O conceito de justificação é essencialmente epistemológico e admite variações. Proponho ver o que Rawls chama de teoria moral como uma dessas variações, como a tese de que os problemas filosóficos de justificação

²⁴³ *IMT*, p.289-290.

moral podem ser reinterpretados e resolvidos sem que seja necessário lidar com a ideia de verdade moral ou sem que seja necessário discorrer sobre o que é o conhecimento moral. Assim, a expressão “teoria moral” meramente designa um conceito de justificação epistêmica em que a justificação não é entendida como conducente à verdade moral (= conhecimento moral). Em 1975 Rawls opera com um conceito bastante restrito e limitado de “epistemologia”, um conceito que hoje já não faz o mesmo sentido.

A melhor maneira de analisar a posição epistemológica de Rawls é começar chamando a atenção para os vários momentos em que ele assere com relativa nitidez o seu holismo. Na seção 87 de *TJ* há a talvez mais clara enunciação: “a justificação repousa na *concepção inteira* e como ela se ajusta com e organiza nossos juízos ponderados em equilíbrio reflexivo”²⁴⁴. Remetendo ao capítulo passado, nós podemos ver que citações como essa autorizam a afirmação de que o ERA não se afasta do holismo característico do ERE em Goodman, Scheffler e no próprio *Outline*.

O holismo é significativo porque é a partir dele que Rawls descarta, na seção 87 de *TJ*, duas posições diferentes sobre a justificação moral. A primeira é a que ele denomina de cartesiana: a posição segundo a qual há primeiros princípios que podem ser vistos como necessariamente verdadeiros. O conhecimento moral seria construído dedutivamente, transferindo esse status epistêmico das premissas para as conclusões, ou das crenças básicas infalíveis para as crenças não-básicas²⁴⁵. Aqui Rawls rejeita sem dúvidas o fundacionalismo clássico e o intuicionismo de autores como Moore e Ross (segundo alguma interpretação mais ou menos padrão do que seja o intuicionismo de Moore e de Ross). É digno de nota que na seção 87 Rawls rejeita o intuicionismo não pelo seu pluralismo de primeiros princípios, como é feito na seção 7, mas pela própria existência desses primeiros princípios tal como o intuicionismo os entende. A outra posição recusada é a naturalista, que define os valores morais em termos de conceitos não-morais. Essa posição, Rawls entende, necessita de uma teoria do significado que não está disponível (ou não estava). Pode-se ver que essas duas posições são recusadas por conflitarem com o holismo. A primeira afirma que o status epistêmico dos princípios primeiros é completamente independentemente da sua relação com o restante da teoria, o que implica que eles são incorrigíveis a partir de outras crenças. A segunda pretende uma redução dos valores morais a certos fatos naturais, então ela precisaria assumir uma teoria

²⁴⁴ “Justification rests upon the entire conception and how it fits in with and organizes our considered judgments in reflective equilibrium” (*TJ*, p.579/rev, p.507. Também em *TJ*, p.21/rev, p.19; *TJ*, p.53/rev, p.46; *KC*, p. 357; *IMT*, p.295).

²⁴⁵ *TJ*, p.578/rev, p.506.

do significado atomista. Não só isso, um naturalismo clássico (que é o tipo de naturalismo do qual Rawls estava se afastando) também pressupõe a distinção entre juízos analíticos e sintéticos. Tanto o holismo quanto a rejeição da distinção entre juízos analíticos e sintéticos são teses desenvolvidas por Quine profundamente influentes na época em que Rawls escreveu e publicou *TJ*²⁴⁶. Não há razão para conjecturar que Rawls não estivesse refletindo essas teses ao escrever *TJ*.

Voltarei ao holismo nas subseções posteriores. Agora quero analisar brevemente o que alcinho de *noção geral de justificação*. Cito-a:

Justificação é um argumento endereçado àqueles que discordam de nós, ou a nós mesmos quando nós estamos indecisos. Ela presume um conflito de visões entre pessoas ou no interior de uma pessoa, e busca convencer os outros, ou nós mesmos, da razoabilidade dos princípios nos quais nossas alegações e juízos são fundados. Sendo designada para reconciliar pela razão, a justificação procede daquilo que todas as partes da discussão sustentam em comum. Idealmente, justificar uma concepção de justiça para alguém é dar uma prova dos seus princípios a partir de premissas que ambos aceitamos, tendo esses princípios, por sua vez, consequências que se ajustam aos nossos juízos ponderados²⁴⁷.

A justificação moral não busca conhecimento moral, mas razoabilidade (*reasonableness*). Razoabilidade é livre aceitabilidade em reflexão. Para obter aceitabilidade, a justificação precisa partir de algum consenso, quando o âmbito é intersubjetivo, ou de algumas crenças morais assumidas como dadas, quando o conflito é no interior de uma pessoa. É com esse tipo de justificação em mente que “as várias condições de escolha dos primeiros princípios são incorporadas conjuntamente na noção de posição original”²⁴⁸ como um expediente para estender um consenso mínimo já existente, quando o âmbito é intersubjetivo, ou “para fixar ideias”²⁴⁹, quando o âmbito é subjetivo e pessoal.

²⁴⁶ No *IMT*, quando discutindo questões de significado, Rawls também afirma que as restrições do conceito de correto não operam isoladamente das concepções de justiça a que pertencem, e que “devemos tentar ver como elas se unem com as estruturas morais como um todo” (*IMT*, p.295). Nesse viés, as definições necessitam de um corpo de teoria no interior do qual elas fazem sentido, e isso implica que juízos analíticos não podem ser independentes dos sintéticos. Em *IMT* Rawls claramente se aproxima do holismo semântico de Quine. Em mais de um momento em *TJ* Rawls apoia considerações semânticas na posição desenvolvida por Quine em *Word and Object* e outros livros (*TJ*, p.579n/rev, p.507n).

²⁴⁷ “Justification is argument addressed to those who disagree with us, or to ourselves when we are of two minds. It presumes a clash of views between persons or within one person, and seeks to convince others, our ourselves, of the reasonableness of the principles upon which our claims and judgments are founded. Being designed to reconcile by reason, justification proceeds from what all parties to the discussion hold in common. Ideally, to justify a conception of justice to someone is to give him a proof of its principles from premises that we both accept, these principles having in turn consequences that match our considered judgment” (*TJ*, p.580-581/rev, p.508-509).

²⁴⁸ *TJ*, p.582/rev, p.509.

²⁴⁹ *KC*, p.357.

É por adotar essa preocupação com a aceitabilidade que Rawls defende que o real problema de justificação em filosofia política é um problema prático, e não um problema teórico²⁵⁰. A justificação teórica diria que uma teoria deve ser conducente à verdade e ao conhecimento proposicional, e que quando há justificação, o senso de justiça de alguém está mais próximo da verdade. Em contraste, a justificação prática sustentaria que a justificabilidade e a validade da justiça como equidade dependem apenas do aceite daqueles a quem ela é destinada (ou de um indivíduo apenas, quando a perspectiva é subjetiva)²⁵¹. Com relação à ideia de justificação pública, abordar o problema da justificação como uma questão prática é buscar por uma concepção de justiça que possa ser acordada livremente por cidadãos considerados livres e iguais²⁵².

É por insistir nesse sentido prático do conceito de justificação que Rawls defende que a filosofia política deve evitar as controversas disputas filosóficas entre cétricos e não-cétricos, realistas e antirrealistas²⁵³. Como visto no capítulo passado, doutrinas abrangentes oferecem razões últimas em favor das suas alegações, e essas razões são incompatíveis entre si. Uma concepção de justiça, para Rawls, deve ser construída independentemente das razões abrangentes dessas doutrinas, pois essa é a única forma de ela conseguir o apoio de todas as doutrinas abrangentes e cumprir o seu objetivo prático de ser foco de um acordo livre. Ou, colocando diferentemente, essa é a única forma de defender uma concepção de justiça que cada cidadão razoável, deliberando em ERA a partir do seu *set* motivacional subjetivo, pode ver como aceitável. Se a justificação fosse uma questão teórica, Rawls teria de defender que se um cidadão não puder, a partir da sua doutrina abrangente, reconhecer a correção da teoria, então o problema está no cidadão que “não vê as coisas corretamente”, mas não na teoria. Seria possível haver uma concepção de justiça teoricamente justificada que poucas pessoas podem

²⁵⁰ KC, p.306; RH, p.394-395.

²⁵¹ Freeman escreveu respeito deste ponto enfatizando que a tese básica de Rawls pode ser lida a partir da alegação de Kant (segundo a maneira como Freeman lê Kant) a respeito da “prioridade da razão prática”. Cito Freeman sobre Kant: “Ele quis dizer que nós estamos autorizados a assumir que juízos feitos de um ponto de vista prático podem ser objetivos, que princípios morais e racionais podem ser universalmente válidos, e que quando aplicados as conclusões desses princípios são adequadas (razoáveis ou verdadeiras), embora elas não satisfazem os mesmos padrões que se aplicam à razão teórica (verificabilidade empírica, correspondência com fatos antecedentes, etc). A razão prática tem o seu âmbito distinto e funciona de acordo com os seus próprios princípios; assim ela tem os seus próprios padrões de objetividade e ‘validade’ que se aplicam unicamente a ela” (FREEMAN, 2007, p.311). Desde que nós diferenciemos adequadamente o que Rawls e Kant entendem por razão prática, e o que cada um entende como suficiente para falar em crença moral justificada, a explicação de Freeman é esclarecedora.

²⁵² JPFNM, p.394.

²⁵³ “Filosofia como a busca pela verdade sobre uma ordem metafísica e moral independente não pode, eu acredito, fornecer uma base compartilhada e praticável para uma concepção política de justiça em uma sociedade democrática” (“Philosophy as the search for truth about an independent metaphysical and moral order cannot, I believe, provide a workable and shared basis for political conception of justice in a democratic society”) (JPFNM, p.395)

aceitar livremente. Como consequência, a filosofia se tornaria uma disciplina intolerante e a sociedade filosoficamente ideal seria uma sociedade autoritária. Nós devemos, como Rawls afirmou, aplicar “o princípio de tolerância à filosofia ela mesma”²⁵⁴ mesmo que isso signifique permanecer “na superfície, filosoficamente falando”²⁵⁵.

Note que esse conceito de justificação é genérico e especifica muito pouco. Para tornar as coisas mais específicas, podemos dividi-lo em dois, completamente distintos. Um deles é subjetivo ou pessoal e pode ser extraído da ideia de que a justificação é um argumento que um indivíduo endereça a si mesmo quando está indeciso. Remetendo ao capítulo passado, esse é o sentido do que chamei de justificação pessoal e que Rawls denomina em *RH* de justificação *full*. O outro é interpessoal ou público, e diz respeito à ideia de dirigir argumentos com vistas a convencer o outro. Esse é o conceito do que Rawls alinha de “justificação pública” nos seus escritos pós-*TJ*. O equilíbrio reflexivo foi tão erradamente interpretado ao longo dos últimos anos porque poucos se ativeram a essa simples, mas crucial, distinção. Como defenderei, o ERA concerne à ideia de justificação pessoal e não se pronuncia a respeito da justificação pública, que é objeto do que Rawls chama de equilíbrio reflexivo geral (*general reflective equilibrium*). Nas duas subseções seguintes, focar-me-ei exclusivamente no tipo de justificação pessoal.

3.1.2 Fundacionalismo moderado ou coerentismo holista?

Agora eu devo enfrentar uma questão que tem sido adiada desde o capítulo primeiro e que é fonte de considerável controvérsia no que diz respeito ao equilíbrio reflexivo: o ERA é um caso de coerentismo ou é mais bem visto como uma forma de fundacionalismo moderado (o fundacionalismo clássico é obviamente excluído)? Não podemos abordar essa questão ignorando a distinção acima realizada, entre um sentido pessoal e outro público de justificação. A pergunta que será abordada nesta seção é a seguinte: *quando alguém sustenta crenças justificadas em ERA, qual a estrutura do seu corpo de crenças?* Para respondê-la, assumirei as seguintes definições de fundacionalismo e coerentismo, propostas por Audi:

Fundacionalismo moderado: para qualquer *S* e qualquer *t*, (1) a estrutura do corpo de crenças justificadas de *S* é, em *t*, fundacional no sentido de que a justificação de quaisquer crenças inferenciais justificadas (crenças não-fundacionais) de *S* depende de uma ou mais crenças não-inferenciais justificadas (crenças fundacionais ou básicas) de *S*. A (2) justificação das crenças fundacionais de *S* é anulável (*defeasible*) em *t*, se, em *t*, elas forem incoerentes com uma ou mais crenças de *S* e (3) as crenças não-fundacionais não derivam toda a sua justificação das crenças fundacionais, mas derivam

²⁵⁴ “To formulate such a conception, we apply the principle of toleration to philosophy itself” (*JFPNM*, p.388).

²⁵⁵ “Justice as fairness deliberately stays on the surface, philosophically speaking” (*JFPNM*, p.395).

apenas o suficiente para que permaneçam justificadas se a justificação que elas obtêm de outras fontes for eliminada²⁵⁶.

Esse fundacionalismo admite o falibilismo ao menos em dois sentidos: (i) as crenças da superestrutura (crenças não-fundacionais) podem ser falsas mesmo quando as crenças fundacionais são verdadeiras (o *link* entre elas não é necessariamente dedutivo) e (ii) há espaço para que se descubra que, em um momento *t* futuro, as crenças fundacionais conflitam ou com outras crenças fundacionais ou com crenças não-fundacionais adequadamente suportadas. A aquisição de novas crenças pode anular a justificação das crenças fundacionais porque as crenças não-fundacionais não derivam toda a sua justificação das crenças fundacionais.

Já o coerentismo holista pode ser definido como se segue:

Coerentismo holista: para qualquer *S* e qualquer *t*, se *S* possui crenças justificadas em *t*, então, em *t*, (1) elas são justificadas pela sua coerência com um sistema coerente de crenças de *S*; e (2) elas permanecem justificadas mesmo se (outras coisas sendo igual) a justificação que elas obtêm de outras fontes que não a coerência for eliminada²⁵⁷.

Conforme BonJour, o coerentismo holista pode ser caracterizado a partir de três características centrais²⁵⁸. A primeira é o caráter (i) holista ou não-linear da justificação: todas as crenças em um sistema de crenças relevante estão em uma relação de *suporte mútuo*, nenhuma crença sendo considerada epistemicamente anterior ou superior a outras. Sendo assim, as crenças no interior do sistema são todas justificadas apenas derivativamente, em virtude de sua participação em um sistema coerente de crenças. A segunda característica (ii) é a própria noção de coerência, que deve ser definida não apenas em termos de consistência lógica entre

²⁵⁶ Essa definição extrai elementos de duas definições oferecidas por Audi (1993, p.134; 2011, p.233).

²⁵⁷ AUDI, 1993, p.138. BonJour oferece uma definição de coerentismo holista que é, no essencial, a mesma de Audi. Cito: “uma visão de acordo com a qual (1) não há crenças fundacionais ou básicas e (2) a base primária para a justificação empírica é o fato de que tais crenças se ajustam conjuntamente e apoiam-se umas nas outras, assim formando um sistema coerente de crenças” (BONJOUR, 2003, p.42).

²⁵⁸ De fato, BonJour apresenta quatro características centrais. A quarta característica é denominada de (iv) “presunção doxástica” (*doxastic presumption*). Para uma posição coerentista questões de justificação devem ser remetidas ao sistema de crenças, do que se segue que o indivíduo deve ser capaz de ter uma crença adequada e justificada sobre o seu próprio sistema de crenças. Essa crença é uma meta-crença que especifica o conteúdo do seu sistema de crenças. O problema com esse quadro é que para um coerentista qualquer crença deve ser justificada pela sua relação com um sistema coerente de crenças. Por conseguinte, a meta-crença de que eu tenho crenças ocorrentes também deve ser justificada pela sua coerência com essas crenças que ela está representando. Mas se o que eu quero saber é justamente quais são as crenças que compõem o meu sistema de crenças, contar com essas mesmas crenças para justificar a minha meta-crença é tomar por corretas justamente as crenças cuja justificação está em questão. O raciocínio é completamente circular. Para lidar com esse problema, BonJour insere a presunção doxástica: a meta-crença é uma hipótese injustificada a partir da qual questões de justificação são condicionalmente avaliadas, produzindo resultados da forma “se a minha representação do meu sistema de crenças é correta, então tal-e-tal crenças são justificadas” (BONJOUR, 2003, p.52). BonJour considera que esse é um problema fatal para uma posição coerentista preocupada em assegurar condições para o conhecimento empírico.

crenças, mas também em termos de relações probabilísticas e explanatórias. Uma terceira característica, que é mais importante para um coerentismo aplicado para lidar com crenças empíricas, é uma (iii) explicação do papel da noção de observação. Grosso modo, o coerentista defende que embora certas crenças sejam formadas não-inferencialmente ou “espontaneamente” em função de algum processo causal, a sua justificação depende da sua relação com o conjunto de crenças do indivíduo²⁵⁹.

Que tipo de evidência textual há para uma posição coerentista tal como é caracterizada por Audi e BonJour? Basicamente aquelas que apontam para o holismo, como esta: “é melhor ver essas condições [que constituem a posição original] simplesmente como estipulações razoáveis que devem ser eventualmente avaliadas por toda a teoria à qual elas pertencem”²⁶⁰. Essa afirmação sugere fortemente a característica (1) do coerentismo holístico, e por implicação a rejeição da característica (1) do fundacionalismo moderado. Também há aquelas em que Rawls recusa operar com “verdades básicas”: “eu não afirmo que os princípios de justiça propostos são verdades necessárias ou deriváveis de tais verdades. Uma concepção de justiça não pode ser deduzida de premissas autoevidentes ou condições impostas a princípios”²⁶¹. Passagens como essas parecem não deixar margem para dúvidas: para avaliar se uma crença é justificada é preciso considerar como ela se relaciona com outras crenças, sendo que essas outras crenças são a “teoria como um todo”. Não haveria espaço para a característica (1) do fundacionalismo, que afirma a necessidade de haver crenças não-inferencialmente justificadas, uma vez que a justificação seria vista “como uma questão de suporte mútuo de muitas considerações, de *tudo* se ajustar em uma visão coerente”²⁶². Afirmar que a justificação é simplesmente uma questão de apoio mútuo parece representar a defesa da característica (2) do coerentismo holista: se a justificação que uma crença obtém de outras fontes for eliminada, mas ela ainda for coerente, ela ainda será justificada. A consequência fundamental dessa característica (2) é a tese de que a coerência é suficiente para a justificação. Outras referências em favor do holismo são aquelas em que Rawls textualmente afirma não supor “crenças fundacionais”. Essas eu avaliarei ao final desta subseção.

²⁵⁹ BONJOUR, 2003, p.45-53.

²⁶⁰ “Seems best to regards these conditions simply as reasonable stipulations to be assessed eventually by the whole theory to which they belong” (TJ, p.578/rev, p.506).

²⁶¹ “I do not claim for the principles of justice proposed that they are necessary truth or derivable from such truths. A conception of justice cannot be deduced from self-evident premises or conditions on principles” (TJ, p.21/ rev, p.19).

²⁶² “Its justification is a matter of the mutual support of many considerations, of everything fitting together into one coherent view” (TJ, p.21/ rev, p.19).

Por outro lado, Rawls divide os juízos morais ponderados em dois grupos: os mais fixos e os menos fixos. Os mais fixos são aqueles que “sustentamos com mais confiança”, “que estamos indispostos a descartar” etc. Rawls afirma que podemos avaliar a posição original (ou a sua teoria, daria no mesmo) em parte pela sua capacidade de acomodar nossas “convicções mais firmes”²⁶³. Rawls claramente ordena epistemologicamente as nossas convicções. Aliás, esse é o objetivo da sua teoria como um todo: mostrar quais das nossas convicções têm prioridade e devem orientar a revisão do nosso senso de justiça quando ele é incoerente. Nós poderíamos assim entender este ponto como representando a característica (1) do fundacionalismo moderado e a rejeição da característica (1) do coerentismo: as crenças não-fundacionais (juízos menos fixos) seriam justificadas em razão da sua coerência com as crenças fundacionais (juízos mais fixos), mas as crenças fundacionais não seriam justificadas em razão da sua coerência com as crenças não-fundacionais. A lógica geral do raciocínio de Rawls, que depende dessa separação entre dois conjuntos de crenças, em que um conjunto é epistemicamente prioritário em relação ao outro, representa uma ameaça ao espírito de uma posição coerentista, que, como visto pela característica (i) do coerentismo elencada por BonJour, é caracterizado por acentuar a igualdade epistêmica de todas as crenças que constituem um sistema coerente de crenças.

Neste ponto, porém, uma caracterização fundacionalista parece encontrar uma dificuldade: para o ERA, a justificação é recíproca. Nós temos de buscar por ajustes entre as nossas crenças, e revisar umas a partir das outras, em um processo em que nenhuma crença é imune à revisão. Qualquer crença pode ser revisada se ela se mostrar incompatível com um conjunto coerente de crenças, do que se segue (supostamente) que não há uma crença não-inferencialmente justificada, isto é, uma crença cuja justificação não dependa do seu ajuste a um sistema coerente de crenças. Os nossos pontos fixos, embora mais fixos que outras crenças que sustentamos, sempre são *provisórios*. É o fundacionalismo moderado incompatível com a extensa revisabilidade e a ideia associada de justificação por ajuste mútuo?

Uma posição fundacionalista, tal como é definida por Audi, parece capaz de responder negativamente a essa questão. Para tanto ela precisa invocar a sua característica (2), que permite que crenças fundacionais sejam anuladas caso se mostrem incoerentes com um conjunto coerente de crenças. Assim, juízos morais menos fixos são justificados quando coerentes com as restrições da posição original (que descrevem os juízos mais fixos), mas a posição original

²⁶³ *TJ*, p.20/rev, p.17.

não seria justificada pela sua coerência com essas crenças, embora, conforme Rawls salienta, deva ser corrigida quando incoerente com um número substantivo delas.

Uma razão forte em favor de uma leitura fundacionalista do ERA é a ideia de que alguns juízos morais ponderados, como os pontos fixos provisórios, possuem algum nível de credibilidade inicial²⁶⁴. Rawls confirma isso em *IMT*: “tenta-se ver como as pessoas ajustariam as suas várias convicções em um esquema coerente, cada convicção ponderada, seja qual for o seu nível, possuindo certa credibilidade inicial”²⁶⁵. É útil aqui voltar à discussão do capítulo passado. Goodman e Scheffler declaradamente não defenderam uma teoria “coerentista pura” em razão da objeção dos sistemas alternativos. E a ideia de credibilidade inicial foi justamente introduzida por eles para mostrar em que sentido eles não estavam defendendo esse formato de coerentismo. Para lembrar, Goodman se empenhou em refutar a posição segundo a qual a percepção sensível fornece conhecimento incorrigível. Contra essa posição Goodman e Scheffler defenderam que toda e qualquer crença oriunda da experiência imediata poderia ser anulada ou descartada se ela não coerisse com um sistema de crenças que já sustentamos. Apesar disso, eles defenderam que crenças perceptivas possuem uma credibilidade inicial para nós, e elas são o critério, *independente da coerência*, pelo qual podemos julgar sistemas alternativos. Agora, isso realmente implica a rejeição do fundacionalismo? Quando é defendido que essas crenças podem ser descartadas se incoerentes com outras crenças, o que Scheffler e Goodman fazem, tão-somente, é afirmar a característica (2) de uma posição fundacionalista moderada: de que uma crença básica pode ser anulada. Assim, seguindo Harman, Bonjour e outros, caracterizo as posições de Goodman e Scheffler (assim como o ERA em geral) como uma posição fundacionalista moderada (falibilista)²⁶⁶. O ponto básico é que a credibilidade inicial é uma credibilidade epistêmica para a crença cuja fonte não é a sua coerência com um sistema coerente de crenças. Esse ponto afirma a característica (1) do fundacionalismo moderado ao mesmo tempo em que rejeita as características (1) e (2) do coerentismo holista.

²⁶⁴ No que respeita a esse ponto, Sencerz fala em credibilidade “intrínseca” ou “não-derivativa” (1986, p.78). Particularmente, creio que a ideia de credibilidade intrínseca não se ajusta muito bem à leitura fundacionalista que Audi propõe. Isso porque a credibilidade das crenças básicas, embora assumidas como *prima facie* justificadas, pode ser abalada em caso de inconsistência com outras crenças ou em situações de expansão de crenças. A credibilidade, embora não-derivativa, não é independente das demais crenças nesse sentido.

²⁶⁵ “One tries to see how people would fit their various convictions into one coherent scheme, each considered conviction whatever its level having a certain initial credibility” (*IMT*, p.289).

²⁶⁶ HARMAN, 2005, 559–575; GAUS, 1996, p.74-85; BONJOUR, 1985, p.28. Bonjour se refere à posição de Goodman como um “fundacionalismo fraco”. O que a distinção proposta por Audi possibilita perceber é que qualquer posição epistemológica que não seja coerentista pura (que não afirme que a coerência é suficiente para a justificação) será uma posição fundacionalista. Qualquer deflacionamento da tese coerentista nessas linhas imediatamente a transforma em uma tese fundacionalista moderada. O equilíbrio reflexivo é fundacionalista porque ele é um “coerentismo” deflacionado.

Como enfatizou Audi, é preciso reconhecer o papel que o conceito de coerência desempenha em uma posição fundacionalista e, sobretudo, reconhecer que se uma crença pode ser anulada se incoerente, isso não significa que a fonte da sua justificação é a coerência²⁶⁷. Audi distingue entre *dependência negativa* da coerência e *dependência positiva*. O fundacionalismo é negativamente dependente da coerência porque se houver incoerência (entre as próprias crenças básicas ou entre as crenças básicas e as não-básicas) não há justificação, mas ele não é positivamente dependente porque, diferentemente do coerentismo, não afirma que a coerência é a *fonte* da justificação²⁶⁸. Em outros termos, para o coerentista a coerência é constitutiva da justificação ao passo que para o fundacionalista ela é uma característica consequente necessária da crença justificada. A única tese coerentista que um fundacionalista nega é a tese de que a coerência é uma base suficiente de justificação. A ideia de justificação por reciprocidade ou apoio mútuo, por exemplo, é um conceito chave em um *framework* fundacionalista, na medida em que ela permite explicar a corrigibilidade das crenças básicas. O fundacionalista moderado não precisa, e nem deve, rejeitar essa característica do ERA.

²⁶⁷ Audi nunca descreveu detalhadamente o ERA, apesar de ser possível identificar na sua tentativa de revitalizar o intuicionismo de Ross uma influência significativa do método e, de fato, uma própria versão modificada capaz de compatibilizá-lo com a filosofia de Ross (ver AUDI, 2011, p.49-51 para uma declarada tentativa de mostrar que Ross utiliza algo como o ERA). Nos comentários que fez sobre o equilíbrio reflexivo, Audi deixou claro que o compreende como uma forma de fundacionalismo moderado, antes que um método filiado à epistemologia coerentista. Cito Audi: “Equilíbrio Reflexivo é uma relação entre atitudes proposicionais, particularmente crenças; mas nós o alcançamos ajustando-as – especialmente por refinamento, eliminação e acréscimo – de um modo que resulta em um conjunto delas com conteúdos que são mutuamente coerentes, complementares em poder explanatório, e reciprocamente esclarecedores. Se o método parece apontar para uma leitura coerentista do intuicionismo ou das teorias kantianas, deve ser especificado que em ambas as abordagens, e certamente para o intuicionismo kantiano, equilíbrio reflexivo não é considerado confirmatório *ao menos que as cognições situadas nele têm alguma plausibilidade intuitiva independentemente dele*” (AUDI, 2004, p.111). A sentença destacada (por mim) indica, como Audi esclarece em outra ocasião, por que o equilíbrio reflexivo é um caso de fundacionalismo moderado (AUDI, 2004, p.224n51).

Na mesma página em que utiliza a expressão fundacionalismo moderado, Audi afirma que ofereceu razões para mostrar por que o equilíbrio reflexivo não precisa ser visto como coerentista nos capítulos 1 e 2 do seu *The Architecture of Reason* (2001). Em algumas seções desses capítulos, Audi desenvolve a distinção entre dependência negativa da coerência e dependência positiva, e entre coerência enquanto condição necessária de justificação (ou condição consequente) e coerência enquanto condição suficiente (ou constitutiva). Como venho desenvolvendo, essas distinções são absolutamente imprescindíveis para uma leitura fundacionalista do equilíbrio reflexivo. Se essas distinções são aceitas, então a implicação é a leitura fundacionalista; ao contrário, se elas forem rejeitadas, então a leitura fundacionalista não parecerá nada convincente. O próprio Audi se utiliza dessa distinção para defender a sua rearticulação da posição intuicionista de Ross. Especificamente, Audi deseja manter os aspectos epistêmicos fundamentais da posição de Ross, como o seu intuicionismo e o seu fundacionalismo, sem se comprometer com os mesmos elementos infalibilistas que usualmente se supõem pressupostos, como a tese de Ross segundo a qual há certos deveres *prima facie* que são *a priori* e necessários. Para tanto, Audi reconstrói a ideia de autoevidência para permitir que uma proposição autoevidente seja, ao mesmo tempo, não-inferencialmente justificada e revisável. Uma intuição é sempre revisável quando está em contradição ou com os resultados de uma teoria moral ou com um amplo corpo de outras intuições. Audi define esse ponto afirmando que as intuições são “negativamente dependentes da teoria”, mas não positivamente dependentes (AUDI, 1997, p.41). Como resultado, uma posição como a de Ross pode permitir o tipo de revisibilidade via ajuste mútuo, característica do ERA (AUDI, 1997, p.50).

²⁶⁸ AUDI, 1993, p.144.

Considerando essa distinção, creio que, assim como o ERE, o ERA é compatível com uma leitura fundacionalista moderada. Com essa distinção é possível manter intacta a tese fundamental do ERA: nenhum elemento, nem mesmo as restrições que compõem a posição original, é imune à revisão. Também é possível manter a inteligibilidade e a relevância das referências de Rawls ao holismo. Se uma crença (fundacional ou não) não for coerente com um conjunto de crenças, então ela não é justificada. A única qualificação a respeito do holismo que precisa ser feita a fim de acomodar essa leitura fundacionalista é a especificação de que a relação inferencial recíproca de crenças no interior do sistema não gera a justificção de todas as crenças. Essa seria uma tese inaceitável para o fundacionalismo moderado. A coerência é necessária para a justificção, mas não é suficiente: alguns juízos morais ponderados têm de ser inicialmente credíveis²⁶⁹. Enquanto o fundacionalismo moderado se mostra congruente com todos os aspectos centrais do ERA, o coerentismo holista não consegue explicar, ao menos tão bem quanto o fundacionalismo, o papel dos pontos fixos provisórios. Dessa forma, uma leitura fundacionalista, além de compatível, parece preferível.

Se fizer sentido falar em dependência positiva da coerência e dependência negativa, então é preciso operar outra distinção. Devemos utilizar a expressão juízos morais ponderados *iniciais* para designar aqueles que possuem credibilidade inicial (“mais fixos”). Esses juízos têm de ser justificados não-inferencialmente, isto é, independentemente da sua relação com outras crenças em ERA. Por outro lado, devemos denominar de juízos morais ponderados *simpliciter* aqueles alcançados em ERA, isto é, após o indivíduo refletir sobre o seu senso de justiça orientado pela posição original ou o que é a mesma coisa, considerando que a posição original é construída para articular ideias subjacentes nos pontos fixos provisórios, pelos seus juízos morais ponderados *iniciais*. Como a posição original é modelada por um processo de avanços e recuos, uma leitura fundacionalista afirma que os juízos iniciais são suscetíveis de revisão se incompatíveis com os juízos *simpliciter* em ERA. O que uma leitura fundacionalista

²⁶⁹ Daniels, talvez o mais notável dos defensores do coerentismo, argumentou que a “extensa revisibilidade” do ERA não é uma característica compartilhada por posições fundacionalistas e que de acordo com o ERA nós podemos manter os nossos juízos morais ponderados ao menos que tenhamos razões para duvidar deles (o que ele chama de conservantismo metodológico) (DANIELS, 1996, p.27-28). Equipados com a distinção acima nós podemos explicar que essa dependência da coerência é apenas negativa e que a revisibilidade, mesmo extensa, não é uma característica que distingue o coerentismo de um fundacionalismo moderado. É preciso destacar que essa “revisibilidade extensa” não parece ser vista por Rawls como igualmente aplicável a todos os juízos. A revisibilidade da concepção de pessoa é, ele escreve, uma “mera possibilidade”, algo “difícil de imaginar realisticamente” (KC, p.352). Embora ao mencionar isso Rawls esteja se referindo à possibilidade de revisar uma concepção de pessoa a partir de resultados empíricos oriundos das ciências, não creio que ele estivesse supondo que seria mais fácil imaginar ou mais provável que nós pudéssemos abrir mão de juízos fixos tais como o de tolerância religiosa e de repúdio à escravidão, juízos dos quais as concepções-modelo são derivadas, em nome de outros juízos especificamente morais que a justiça como equidade recomenda.

nega, e a posição coerentista afirma, é que esse processo de ajuste mútuo justifica os juízos *iniciais*.

Rawls afirma que cada juízo moral ponderado tem um certo nível de credibilidade inicial para nós²⁷⁰. Essa generalização deve ser refinada por uma leitura fundacionalista. A credibilidade dos juízos *simpliciter* tem de ser derivada dos juízos iniciais, e a destes últimos tem de ser, obrigatoriamente, não-derivativa. A ideia de credibilidade “inicial” não se ajusta bem à definição que eu ofereço de juízos *simpliciter*, posto que são juízos alcançados em ERA. Scheffler em um dado momento faz uma distinção entre juízos com credibilidade inicial e juízos com credibilidade inicial mais baixa (*lower initial credibility*)²⁷¹. No capítulo primeiro se viu que há vários níveis de credibilidade que são ordenados linearmente pelo nível de comprometimento: quanto maior o comprometimento com que uma crença é sustentada, maior a sua credibilidade para quem a professa. Não há por que pensar que Rawls não compartilhe desse ponto. Ele assevera que “nós sentimos com certeza” que algumas coisas devem ser respondidas de uma determinada maneira, “que estamos indispostos” a rejeitar certos juízos que sustentamos com “a maior confiança”, etc., mas que, por outro lado, estamos muito “menos seguros” com respeito a outros juízos. Os juízos iniciais podem ser vistos como aqueles do primeiro tipo, sustentados com mais comprometimento (mais fixos) e cuja credibilidade não é derivada de outros juízos, mas fundada nesse próprio ato doxástico de sustentar a crença com um alto grau de comprometimento. A credibilidade dos juízos *simpliciter* é derivativa no seguinte sentido: quando eles são questionados, S tende a citar os juízos *iniciais* para justificá-los. Quando os juízos iniciais são questionados, S tem de dizer que “ele sente” que aquela é a resposta correta e que não há nada mais para ser dito.

Alguém pode perguntar: mas se os juízos iniciais forem rejeitados ao longo do ERA, como o fundacionalista permite que sejam, isso não significa que não há juízos não-inferencialmente justificados? Como Ebertz corretamente apontou, mesmo se os juízos morais ponderados iniciais forem rejeitados ao longo do processo, isso não implica que no final não há (outros) juízos morais ponderados operando como crenças básicas. Ebertz recorre à frase de Rawls de que “há uma classe definida e limitada de fatos contra os quais princípios conjecturados podem ser testados, nomeadamente, nossos juízos ponderados em equilíbrio reflexivo”²⁷² para comprovar o seu ponto. Uma citação como essa claramente mostra, segundo

²⁷⁰ *IMT*, p.289.

²⁷¹ SCHEFFLER, 1954, p.182

²⁷² “There is a definite if limited class of facts against which conjectured principles can be checked, namely, our considered judgments in reflective equilibrium” (*TJ*, p.51/rev, p.44).

Ebertz, que os juízos morais ponderados (iniciais, eu qualificaria) operam fundacionalmente e que eles são constitutivos, enquanto operadores fundacionais, do método do ERA²⁷³.

Face uma leitura fundacionalista, essa frase de Rawls deve ser revisada para livrar-se de certas ambiguidades e para acomodar a distinção entre juízos mais fixos e menos fixos. Reconstruo-a da seguinte forma: há uma classe definida e limitada de fatos contra os quais princípios e juízos morais ponderados *simpliciter* afirmados em ERA podem ser testados, nomeadamente, nossos juízos morais ponderados iniciais. Os juízos iniciais permitem ver qual a descrição possível de todos os nossos juízos *simpliciter* é a que estamos mais dispostos a aceitar em ERA²⁷⁴, analogamente ao modo como a posição original nos permite descobrir qual a concepção de justiça que mais bem descreve o nosso senso de justiça. Pode ser o caso que, e aqui entra o elemento falibilista ou moderado desse fundacionalismo, nenhuma descrição possível possa ser aceita. Isso implicará que os nossos juízos iniciais estão nos guiando incorretamente e devem ser revisados e substituídos por outros. Nesse processo alguns juízos *simpliciter* assumirão a função de juízos iniciais, e os últimos tomarão o lugar de alguns dos primeiros ou serão completamente descartados. Contudo, como Ebertz destaca, sempre haverá um subconjunto de juízos morais ponderados funcionando como crenças básicas no processo do ERA.

Contra uma leitura fundacionalista, alguém poderia recorrer aos argumentos de Brink em favor de uma interpretação coerentista. Em resposta a uma caracterização fundacionalista, Brink começaria perguntando por que os juízos morais ponderados são, em primeiro lugar, introduzidos na discussão. De acordo com a sua formulação do ERA, Brink diria que eles são introduzidos porque são realizados sob condições cognitivamente confiáveis (as condições que definem o ponderado) e, portanto, são conducentes à verdade e ao conhecimento moral²⁷⁵. Os juízos morais ponderados (Brink não admitiria a distinção entre iniciais e *simpliciter*) não são crenças fundacionais porque a justificação para sustentar um juízo ponderado, digamos *p*, requer que alguém baseie *p* em alguma crença de segunda-ordem, *e*, sobre que tipo de crença *p* é e por que crenças do tipo *p* deveriam ser vistas como confiáveis²⁷⁶. Adaptando essa linha de raciocínio para os propósitos de Rawls, poderia se recorrer à noção de justificação prática para objetar uma leitura fundacionalista. O argumento então seria o seguinte: os juízos *iniciais* não

²⁷³ EBERTZ, 1993, p.202.

²⁷⁴ Conforme Rawls: “Here we may be looking for a way to remove our doubts. We can check an interpretation of the initial situation, then, by the capacity of its principles to accommodate our firmest convictions and to provide guidance where guidance is needed” (*TJ*, p.20/rev, p.17).

²⁷⁵ BRINK, 1989, p.134-135.

²⁷⁶ BRINK, 1989, p.122.

são justificados independentemente de outras crenças, como defende o fundacionalista, mas são justificados em virtude da sua coerência com as crenças de segunda-ordem (no caso: uma concepção de justiça válida tem de poder ser aceita por todos e fundar uma sociedade estável) que Rawls invoca principalmente na terceira parte de *TJ* e nos seus últimos escritos. Assim, esses juízos iniciais seriam justificados em virtude da sua função prática da teoria à qual eles pertencem, portanto inferencialmente.

A objeção acima é apelativa, mas é resistível ao menos por uma razão decisiva. Em que sentido as crenças de segunda-ordem de Rawls “justificam” os juízos morais ponderados? É preciso distinguir entre duas questões aqui e desfazer uma ambiguidade do uso do termo “justificação”.

Uma é a questão de saber quando *S* está justificado em sustentar que *p*, a outra é a questão de saber se *p* é, ela mesma, uma proposição publicamente justificada. Creio que quando Rawls afirma que uma teoria da justiça tem uma função prática que ela deve satisfazer para ser justificada, ele está no âmbito da segunda questão e se referindo à legitimidade total da teoria e, além da ideia de justificação pessoal (pois toda a teoria publicamente justificável também é pessoalmente justificada para cidadãos razoáveis), à ideia de justificação pública. O fundacionalismo moderado e o coerentismo, como eu venho abordando aqui, são posições sobre a estrutura do corpo de crenças de *S* quando *S* possui crenças justificadas em *t*, portanto estão sob o domínio da primeira questão e se referem à ideia de justificação pessoal (ou racional). Desse modo, o apelo à função prática não mostra que quando *S* está justificado em ERA em *t*, então, em *t*, *S* não possui crenças não-inferencialmente justificadas. O máximo que ele mostra é que *S* pode estar pessoalmente justificado em sustentar crenças fundacionais ilegítimas ou publicamente injustificáveis. Esse é o caso de um fundamentalista religioso, por exemplo.

Ao caracterizar uma posição fundacionalista, eu escrevi que quando os juízos iniciais são questionados, *S* tem de dizer que “ele sente” que aquela é a resposta correta e que nada mais há para ser dito. Brink protestaria contra essa alegação porque ainda há o que ser dito: a crença pode fazer parte de uma concepção política de justiça e é uma crença que, se fosse reconstruída em uma concepção de justiça para regular a sociedade, fundaria uma sociedade estável, é uma crença que faz parte de uma tradição de pensamento que perdura na história há bastante tempo etc. Nem todas as crenças que as pessoas “sentem” que são justificadas compartilham dessas propriedades, do que se segue que essa qualificação é necessária se quisermos avaliar a justificação da crença adequadamente.

Não creio que esse argumento seja fatal contra uma leitura fundacionalista. Conforme visto no capítulo passado, na seção dedicada à análise do ERA no contexto do *PL*, para Rawls

essas considerações a respeito do estatuto *político* de uma concepção de justiça, assim como a função social que ela desempenha, não representam as razões últimas pelas quais os cidadãos afirmam uma concepção política. Isso significa dizer que essas considerações não contribuem para a justificabilidade pessoal de uma teoria, embora contribuam decisivamente para a sua justificação pública.

A distinção entre justificação pessoal e pública aponta para duas teses diferentes de Rawls. A primeira afirma que (a) quando um indivíduo razoável que vive sob um regime democrático possui crenças justificadas em ERA, essas crenças suportam a justiça como equidade e são descritas por ela. Um cidadão razoável é aquele que possui como crenças básicas (pontos fixos provisórios) as crenças descritas por Rawls. Um cidadão não-razoável rejeita aqueles valores e age com base em valores inversos (por exemplo, o valor da desigualdade como um bem para promover as potencialidades humanas). Esse cidadão, quando em ERA, não afirmará a justiça como equidade. Isso significa que, embora justificado em ERA, a teoria da justiça que ele suporta é ilegítima (publicamente injustificável). Por isto que a legitimidade da justiça como equidade não pode ser estabelecida meramente recorrendo ao fato de que ela seria afirmada em ERA: qualquer princípio pode estar justificado em ERA.

Uma segunda tese defende que (b) a justiça como equidade é legítima porque pode fundar uma sociedade estável regida por uma concepção publicamente aceitável de justiça. Assim a justiça como equidade explica porque as crenças (sobre justiça) não-inferencialmente justificadas de cidadãos razoáveis são publicamente justificáveis, mas não faz a justificabilidade pessoal dessas crenças depender da função prática que elas desempenham²⁷⁷. Isso pode ser expresso simplesmente estabelecendo que a legitimidade de p não é uma condição necessária para S estar justificado em sustentar p , uma vez que as crenças de segunda-ordem que devem ser acrescentadas para p ser legítima (no caso, a função prática) não são vistas por S como suas razões para sustentar p ²⁷⁸. Por outro lado, a recíproca não é verdadeira: é uma

²⁷⁷ Há outras linhas eficientes de resposta a Brink. Timmons defende um contextualismo que postula a existência de crenças básicas, justamente o tipo de crença cuja existência em geral é desafiada pelo argumento de Brink. Ao responder a Brink, Timmons defende o seguinte: “Suspeita-se de uma confusão de níveis na explicação de Brink a respeito da justificação. Isto é, deveria se distinguir entre questões sobre a justificação de uma crença e questões sobre ser justificado em uma alegação epistêmica sobre uma crença – entre *ser justificado* em sustentar alguma crença e *mostrar que alguém é justificado*. Para alcançar-se o último exige-se que se tenha crenças de segunda-ordem sobre a confiabilidade e, portanto, status epistêmico das crenças de primeira-ordem. Mas é duvidoso que alguém precise ter uma crença de segunda-ordem (do tipo em questão) para ser justificado em sustentar a crença não-epistêmica de primeira ordem” (TIMMONS, 1999, p.240).

²⁷⁸ Gaus objeta Brink por uma razão análoga (embora não estivesse particularmente preocupado com o ERA). Ele argumenta que é razoável pensar que eu estou justificado em sustentar crenças perceptivas *antes* de possuir quaisquer crenças de segunda-ordem a respeito da adequação filosófica do meu aparato perceptivo (GAUS, 1996, p.91).

condição necessária para que *p* seja uma alegação legítima que *p* possa ser justificadamente sustentada por *S* (desde que *S* seja qualificado como um cidadão razoável de um regime democrático contemporâneo). A interpretação que eu ofereço neste capítulo será dramaticamente mal entendida se as distinções acima não forem respeitadas.

Alguém poderia resistir à interpretação aqui oferecida em razão de não haver uma explicação satisfatória a respeito do *status* das crenças não-inferencialmente justificadas. Em que sentido exatamente elas são básicas? Essa é uma questão evidentemente central, mas creio que em Rawls não encontramos nenhuma referência para ensaiar uma resposta. Este é o ponto em que precisaríamos ir além. Porém, o que eu posso fazer é apontar que a existência de crenças básicas ou crenças justificadas *prima facie* não compromete Rawls com a tese de que existe uma ordem de valores morais independente da concepção de pessoa²⁷⁹, de forma que uma leitura fundacionalista pode manter os traços essenciais do construtivismo. Como sugerido por Ebertz, não há porque se envolver com a existência de uma noção carregada de faculdade moral. As crenças básicas podem ser vistas simplesmente como respostas avaliativas de um indivíduo às interações sociais à sua volta²⁸⁰, cuja possibilidade é assegurada pela existência de uma capacidade moral de deliberação prática ou uma sensibilidade moral²⁸¹. Como Rawls afirma, “há questões que nós sentimos com certeza (*feel sure*) que têm de (*must*) ser respondidas de uma determinada maneira”²⁸². As crenças básicas meramente são crenças sustentadas com muito comprometimento e expressam o que consideramos *prima facie* justificado. Consequentemente, na medida em que os juízos morais ponderados iniciais são respostas a um mundo social, indivíduos imersos em diferentes mundos sociais sustentarão diferentes crenças não-inferencialmente justificadas, assim como os mesmos indivíduos em tempos distintos podem sustentar diferentes crenças básicas²⁸³. É claro que essa explicação pouco esclarece a respeito do *status* de crenças não-inferenciais, mas ela mostra que, quaisquer que sejam os

²⁷⁹ Timmons aborda diversas explicações fundacionalistas possíveis para a ideia de crenças imediatamente justificadas que não precisam contar com pressupostos metafísicos ou intuicionistas tradicionais como a ideia de percepção moral (ver TIMMONS, 1987, p.595-609).

²⁸⁰ EBERTZ, 1993, p.202.

²⁸¹ Além de Ebertz, Thomas se empenha em defender uma leitura fundacionalista do ERA. Os argumentos de Thomas são semelhantes aos meus: ele se apoia no ensaio em que Ebertz afirma que mesmo se os juízos morais iniciais forem rejeitados ao longo do ERA, isso não significa que não haverá juízos funcionando fundacionalmente, e assim aceita implicitamente a distinção entre dependência negativa da coerência e dependência positiva (THOMAS, 2006, p.198-224). O próprio Thomas utiliza o ERA interpretando-o a partir de uma epistemologia contextualista, para usá-lo assim no seu projeto realista e cognitivista em metaética. Esse projeto de Thomas se distingue do de Rawls ao menos em três aspectos: (i) Rawls insiste na necessidade do fundacionalismo para manter a relevância para a noção de *input* descritivo; (ii) esse fundacionalismo não é associado ao realismo moral; (iii) o fundacionalismo é voltado para explicar a estrutura de crenças de *S* quando *S* possui crenças justificadas em *t*, mas não para explicar por que as crenças elas mesmas podem ser vistas como um caso de conhecimento moral.

²⁸² TJ, p.19/rev, p.17.

²⁸³ AUDI, 1993, p.164.

detalhes desse conceito, ele não necessariamente precisa pressupor o tipo de doutrina intuicionista que Rawls rejeita, que é um intuicionismo que estabelece alguma relação correspondentista entre a crença não-inferencialmente justificada e alguma porção da realidade descrita por ela. Porém, se os detalhes fundacionalistas do ERA forem preenchidos, o caminho natural deve associá-lo a alguma forma não-tradicional de intuicionismo (isto é, diferente dos intuicionismos propostos por Clarke, Moore e Ross).

Grande parte das leituras não-fundacionalistas do ERA parte de uma compreensão bastante restrita das formas que o fundacionalismo pode assumir e, mais gravemente, analisa a posição epistemológica do método sem se perguntar qual a estrutura do corpo de crenças de *S* quando *S* sustenta crenças justificadas em ERA. Em geral, se assume que o ERA não pode ser fundacionalista porque ele não admite que haja alguma verdade indubitável e incorrigível ou porque para ele não há fatos morais independentes dos juízos morais ponderados. Não raro, o proponente da leitura coerentista também não diferencia projetos distintos de Rawls, como acima eu propus, e, conseqüentemente, não faz qualquer distinção entre sentidos diferentes da ideia de justificação. Ele entende que o ERA seria responsável por assegurar tanto justificação pública quanto justificação pessoal. Era assim que Daniels, nos seus ensaios do final da década de setenta, compreendia o ERA²⁸⁴.

Rawls também pode ser acusado de imprecisões. Ele afirmou que a sua teoria não é “fundacional” ao menos duas vezes, e em nenhuma delas especificou em que sentido estava utilizando esse conceito. Uma dessas afirmações está contida no *PL*:

Uma característica do equilíbrio reflexivo é que ele inclui nossas convicções ponderadas em todos os níveis de generalidade; nenhum nível, digamos de um princípio abstrato ou um juízo particular em casos particulares, é visto como fundacional. Eles todos podem ter uma credibilidade inicial²⁸⁵.

Creio que a afirmação acima não pode servir como uma fonte de negação do fundacionalismo moderado, uma vez que nela Rawls não indica o que ele quer significar por “visto como fundacional”. Dado o contexto e a referência ao *IMT* (no final da nota 8 da página 8 Rawls remete o leitor à distinção entre o ERE e o ERA tal como ela é feita no *IMT*), é plausível que Rawls estivesse simplesmente realçando que a justificação em ERA depende da relação de

²⁸⁴ Porque ele defendeu que o ERA continha um teste de *feasibility* para diferenciar concepções de justiça estáveis de não-estáveis (DANIELS, 1996, p.51). Daniels permanece, em 2011, se referindo ao ERA como coerentista, pelo que podemos ver no verbete “Reflective Equilibrium” da *The Stanford Encyclopedia of Philosophy*, que é de sua autoria (DANIELS, 2011).

²⁸⁵ One feature of reflective equilibrium is that it includes our considered convictions at all levels of generality; no one level, say that of abstract principle or that of particular judgment in particular cases, is viewed as foundational. They all may have an initial credibility (*PL*, p.8).

coerência entre diversos juízos, e que nenhum juízo é considerado incorrigível e verdadeiro. Essas teses são igualmente defendidas por uma leitura fundacionalista moderada, provida a qualificação para a noção de coerência entendida negativamente. Sobre a frase final do excerto, note que ela afirma que todos os juízos podem ter “uma” credibilidade inicial. A leitura fundacionalista não a nega e a qualifica da seguinte maneira: todos os juízos tem um nível de credibilidade, porém alguns juízos tem mais credibilidade do que outros. Algumas páginas atrás eu mencionei que essa expressão “credibilidade inicial” é por si mesma pouco esclarecedora, e que devemos admitir níveis diferentes de credibilidade ou credibilidades derivativas e não-derivativas. Simplesmente não faz sentido igualar o nível de credibilidade de todas as crenças que alguém sustenta, e Rawls obviamente não faz isso porque ele possui uma importante distinção entre juízos mais fixos e menos fixos. Por que ele precisaria dessa distinção, se estivesse assumindo o coerentismo?

A outra passagem em que Rawls se refere explicitamente ao caráter não-fundacionalista do equilíbrio reflexivo é em *JF*. Nessa obra Rawls sustenta o seguinte:

A ideia de justificação associada ao equilíbrio reflexivo pleno é não-fundacionalista no seguinte sentido: nenhum tipo de juízo ponderado de justiça política ou de um nível particular de generalidade carrega todo o peso da justificação pública²⁸⁶.

Essa segunda passagem não contraria a leitura defendida nesta seção pelas seguintes razões: (i) nela Rawls não está se referindo ao ERA, mas ao equilíbrio reflexivo geral e pleno (*general e full*); (ii) consequentemente o foco não é a justificação pessoal, à qual atribuo o fundacionalismo moderado, mas a justificação pública.

As passagens acima citadas foram escritas por Rawls após a sua virada política. Na fase anterior da sua carreira, a terminologia preferida era a de “autoevidência”. Cito também dois excertos. Em uma passagem da seção conclusiva de *TJ*, Rawls afasta o seu modelo de justificação do cartesiano. O modelo cartesiano contaria com a noção de “autoevidência” por sustentar que alguns princípios primeiros seriam “necessariamente verdadeiros”²⁸⁷. Como já está suficientemente claro, o fundacionalismo moderado repudia o projeto cartesiano assim entendido. A outra passagem, em *IMT*, lê-se como se segue:

²⁸⁶ “The idea of justification paired with full reflective equilibrium is nonfoundationalist in this way: no specified kind of considered judgment of political justice or particular level of generality is thought to carry the whole weight of public justification” (*JF*, p.31).

²⁸⁷ *TJ*, p.578/rev, p.506.

Embora que para iniciar vários juízos são vistos como fixos o suficiente para serem tomados provisoriamente como pontos fixos, não há juízos em qualquer nível de generalidade que sejam em princípio imunes à revisão. Mesmo à totalidade dos juízos particulares não é atribuído um papel decisivo; assim esses juízos não tem o status algumas vezes atribuído a juízos de percepção em teorias do conhecimento²⁸⁸.

Se nós tomarmos a segunda dentre as duas frases acima como uma explicação da afirmação de que nenhum princípio é imune à revisão, então não há nenhum obstáculo para uma leitura fundacionalista moderada. É claro que a segunda frase poderia ser lida como uma nova tese para se acrescentar à supracitada ideia de corrigibilidade, mas não vejo razões para encontrar essa possibilidade dado o que Rawls escreve no restante do artigo. Ele não esclarece ao que está se referindo quando menciona o status atribuído por teorias do conhecimento a juízos perceptivos, mas tudo indica que não é nada mais sofisticado do que isto: quando eu tenho uma percepção, a crença que eu formo sobre o conteúdo dessa percepção é necessariamente verdadeira e, assim, imune à revisão e à dúvida. Esse é o tipo de “teoria do conhecimento” que Goodman se preocupou em refutar com a ideia de crenças básicas como crenças com credibilidade inicial²⁸⁹. Pelo papel negativo que atribui à ideia de coerência, o fundacionalismo moderado não está comprometido com esse pressuposto. Conforme Goodman: um enunciado, por mais inicialmente credível que seja, será “descartado se resultar em um sistema que não satisfaz tanto quanto possível a totalidade de alegações apresentadas por todos os enunciados relevantes”²⁹⁰.

Curiosamente, no *IMT* há uma passagem bastante desconfortável para um proponente da interpretação coerentista:

Eu noto de passagem que a concepção moral de alguém pode acabar por ser baseada em primeiros princípios autoevidentes. O procedimento do equilíbrio reflexivo, por si mesmo, não exclui essa possibilidade, por mais improvável que ela possa ser. No curso de alcançar esse estado, é possível que primeiros princípios sejam formulados de um modo tão compelente que eles nos levam a revisar todos os nossos juízos prévios e subsequentes inconsistentes com eles. O equilíbrio reflexivo requer somente que o agente faça as revisões com convicção e confiança, e que continue a afirmar esses princípios quando tiver de aceitar as suas consequências na prática²⁹¹.

²⁸⁸ “Although in order to get started various judgments are viewed as firm enough to be taken provisionally as fixed points, there are no judgments on any level of generality that are in principle immune to revision. Even the totality of particular judgments are not assigned a decisive role; thus these judgments do not have the status sometimes attributed to judgments of perception in theories of knowledge” (*IMT*, p.289).

²⁸⁹ GOODMAN, 1952, p.160.

²⁹⁰ GOODMAN, 1952, p.163.

²⁹¹ “I note in passing that one’s moral conception may turn out to be based on self-evident first principles. The procedure of reflective equilibrium does not, by itself, excludes this possibility, however unlikely it may be. For

O que Rawls está dizendo nessa passagem, colocando em poucas palavras, é que a questão de se o ERA admite ou não princípios autoevidentes é contingente: depende exclusivamente do modo como o seu utilizador delibera. Um utilizador pode tranquilamente manter-se fiel às suas convicções mais compelentes, revisando todas as suas outras convicções quando conflitarem com essa suas convicções mais básicas. Nesse caso, as suas convicções mais firmes serão *autoevidentes para ele* (leia-se: “ímmunes à revisão” ou “tomadas como incorrigíveis”). Embora que esse excerto seja simpático para uma leitura fundacionalista, ele não pode ser aceito por ela sem reservas ao menos por duas razões. Em primeiro lugar, é indubitável que certas crenças podem não ser revisadas ao longo do ERA, mas isso não implica que elas são “ímmunes à revisão” ou incorrigíveis²⁹². A sua revisibilidade é contingente, é verdade, mas o mesmo se aplica à sua possível irrevisibilidade. Afirmar que um indivíduo não revisou a crença é afirmar que, até o momento, ele não encontrou razões para fazê-lo. O importante é que o ERA contenha mecanismos para promover a reavaliação de crenças básicas quando for necessário; se elas precisarão ser modificadas ou não por um dado agente, isso não alterará o modo como o ERA caracteriza a revisão ideal de crenças. Portanto, a afirmação de que o ERA “por si mesmo não exclui essa possibilidade” não pode ser aceita sem qualificação (essa primeira razão seria de bom grado aceita pelos proponentes da leitura coerentista). Porém, se com isso se quer dizer que o ERA não contém mecanismos para fazer com que o indivíduo deixe de sustentar uma determinada crença que é, do seu ponto de vista, incorrigível, então a frase é acertada. Discutirei essa questão ao abordar o descritivismo.

Em segundo lugar, é falso que o ERA poderia vir a ser coerentista se o agente deliberasse, digamos, revisando todos os seus juízos uns à luz dos outros. Aqui uma leitura fundacionalista tem de simplesmente invocar a distinção entre dependência negativa da coerência e dependência positiva. Além disso, como é destacado por Ebertz, mesmo se as crenças básicas não forem as mesmas ao longo do processo, ou forem diferentes após o

in the course of achieving this state, it is possible that first principles should be formulated that seem so compelling that they lead us to revise all previous and subsequent judgments inconsistent with them. Reflective equilibrium requires only that the agent makes these revisions with convictions and confidence, and continues to affirm these principles when it comes to accepting their consequences in practice” (*IMT*, p.289).

²⁹² Em verdade, há um sentido bastante específico em que Rawls está correto neste ponto. Tanto para uma leitura fundacionalista quanto para uma leitura coerentista, no *dado instante* em que o indivíduo está revisando as suas crenças em ERA, uma ou mais crenças devem operar imune à revisão. Nem tudo pode ser revisado em um dado momento, porque sempre é necessário contar com algo para determinar se uma revisão em particular é justificada. O coerentista não nega essa tese, pois a qualifica argumentando que a justificabilidade dessas crenças assumidas como irrevisáveis em um dado instante é verificada posteriormente pela sua coerência com o sistema de crença gerado, em um sentido relevante, por ela mesma. O fundacionalista nega justamente a plausibilidade desse último passo.

processo, isso não alterará o fato de que sempre haverá crenças básicas. As crenças básicas originais simplesmente serão modificadas ou substituídas por outras, que assumirão a sua função.

Apesar das qualificações que julgo necessárias, essa passagem é útil para demonstrar um aspecto do ERA ao qual esta dissertação atribui uma relevância absolutamente central e decisiva: a ideia de que a justificação concerne à estrutura de crenças de S quando S sustenta crenças justificadas em ERA. É nesse sentido que Rawls afirma que pode haver crenças autoevidentes, e é nesse sentido que aqui se está atribuindo o fundacionalismo moderado. Em nenhum outro escrito de Rawls essa característica é estabelecida com tamanha nitidez.

Ao longo desta subseção, não defendi que o fundacionalismo moderado é superior ao coerentismo. Esse argumento não é essencial para o meu propósito, por isso não quero fazê-lo dependente dele. Porém, noto que construir o ERA como fundacionalista permite responder algumas objeções que lhe foram apresentadas. Considere o que Brandt afirma contra Rawls:

Há um problema aqui muito similar àquele que a tradicional teoria coerentista da justificação de crenças enfrenta: a teoria defende que um sistema de crenças mais coerente é mais bem justificado do que um menos coerente, mas não há razão para pensar que essa alegação é verdadeira ao menos que algumas das crenças sejam inicialmente credíveis – e não meramente inicialmente sustentadas – *por alguma razão outra que a coerência*²⁹³.

O que Brandt defende nessa passagem é uma objeção usual ao coerentismo: como a relação de coerência entre crenças pode gerar crenças justificadas, sendo assim suficiente para a justificação?²⁹⁴ Goodman via isso como um problema. Nas suas palavras, “credibilidade não emerge dessas conexões por geração espontânea” e se algum enunciado é credível, tem de haver outro (ou outros) que são inicialmente credíveis, isto é, credíveis independentemente da sua relação com outros enunciados²⁹⁵. Sem esse postulado, é também gerado o problema dos sistemas alternativos: se a coerência é o único critério pelo qual avaliar a justificação de um corpo de crenças, o que fazer quando houver dois corpos de crenças igualmente coerentes, porém contraditórios entre si? A própria ideia de credibilidade *inicial* foi introduzida por Goodman exclusivamente para afastar a tese coerentista de que a coerência é suficiente para a

²⁹³ BRANDT, 1998, p.20. O itálico é de minha autoria.

²⁹⁴ Audi a apresenta como uma razão para rejeitar qualquer forma de coerentismo (AUDI, 2011, p.222). Trata-se de uma poderosa objeção, ao menos biograficamente falando. Foi proposta por Goodman em 1952, ecoada por Brandt na década de setenta e utilizada por Audi em 2011.

²⁹⁵ GOODMAN, 1952, p.162-163.

justificação. E como a ideia de credibilidade inicial é utilizada por Rawls, parece natural inferir que ele compartilha dessa mesma posição.

Ao tentar reinterpretar o conceito de credibilidade inicial a partir do coerentismo, Daniels teve de afirmar que apenas após o ERA nós podemos nos certificar da credibilidade daquelas crenças iniciais. O problema é que, nesse caso, a credibilidade deixa de ser *inicial* e passa a ser derivativa. E então a dificuldade é explicar como a credibilidade pode vir a existir meramente via derivação e por que Rawls utiliza essa expressão tão carregada. Claro que Daniels poderia responder, como efetivamente respondeu, que negar que a coerência tenha essa capacidade é uma objeção fundacionalista genérica²⁹⁶. Em um ensaio seu, nos é inclusive apresentado o que ele mesmo denominou de um “esboço” de estratégia para defender que a coerência pode cumprir esse papel. Como esse esboço não foi até hoje retomado por Daniels, somos inevitavelmente compelidos a questionar se a sua própria tese não é ela mesma uma objeção coerentista genérica ao fundacionalismo²⁹⁷. A questão é que a tese fundacionalista é consideravelmente mais *simples* e intuitiva do que a tese coerentista rival²⁹⁸.

3.1.3 A ideia de justificação pessoal

A exposição acima esclareceu que o ERA adota uma concepção de justificação em que o objetivo da justificação não é a verdade e o conhecimento moral. Assim nós temos um quadro negativo sobre o que o ERA não pretende realizar, mas quais são as suas pretensões, o seu *desideratum* epistêmico?

Rawls afirma que se um indivíduo razoável de um regime democrático contemporâneo seguir o ERA, então ele apoiará a justiça como equidade e, conseqüentemente, estará justificado no senso de justiça que professa. Veja que no ERA o foco não é a justificação da justiça como equidade, por mais estranho que isso possa parecer, mas a justificação do senso de justiça de quem utiliza o ERA. Mas o que significa estar justificado em ERA? Essa é a questão decisiva. Como visto, não implica possuir crenças morais verdadeiras, provavelmente verdadeiras ou que

²⁹⁶ DANIELS, 1996, p.32.

²⁹⁷ O projeto de conectar coerência e verdade, planejado por Daniels, foi executado por BonJour no capítulo 8 do *The Structure of Empirical Knowledge* (1985). Brink valeu-se dos argumentos de BonJour para concretizar, em *Moral Realism and the Foundations of Ethics* (1989), essa defesa coerentista do ERA. É merecedor de nota, contudo, que anos depois BonJour tenha abandonado esse programa com a justificativa de que ele envolve um raciocínio viciosamente circular (BONJOUR, 2003, p.59)

²⁹⁸ Especificamente comentando a controvérsia entre Daniels e Brandt, Gaus defende que o projeto de Daniels é inexequível justamente porque a credibilidade não pode ser concebida apenas derivativamente (GAUS, 1996, p.88-89). Gaus afirma assim que mesmo o coerentismo tem de assumir a existência de crenças auto-justificadas. Segundo o seu raciocínio, o coerentismo comportaria essa restrição fazendo com que todas as crenças de um dado sujeito são auto-justificadas. Nesse caso, ele sustenta, é preferível ficar com uma posição fundacionalista que já comporta mais naturalmente a inevitável existência desse tipo de crença (GAUS, 1996, p.91-97).

são evidência para a verdade moral²⁹⁹. Critérios objetivistas estão descartados. Rawls também afasta a possibilidade de que o ERA seja suficiente para assegurar justificação pública³⁰⁰. Critérios intersubjetivistas também são rejeitados. Se o conhecimento moral e o consenso estão descartados, para o que essa concepção de justificação aponta?

Rawls apresenta diversos fragmentos que indicam uma resposta para essa questão. O primeiro é a definição de ERA como “um estado alcançado quando as convicções morais da pessoa satisfazem certas condições de racionalidade”³⁰¹ ou, expressando de outra forma, quando “as convicções gerais, primeiros princípios e juízos particulares estão em linha”³⁰². O segundo é a noção de *razoável* (no léxico, razoável é algo conforme à razão), que para Rawls “substitui a busca pela verdade”³⁰³. Um conteúdo de justiça razoável é um conteúdo que pode ser foco de um consenso, é verdade, mas ele também tem uma dimensão de primeira pessoa: é um conteúdo “congruente com nosso profundo entendimento de nós mesmos e de nossas aspirações, e nossa compreensão que dada a nossa história e as tradições da nossa vida pública, é a doutrina mais razoável para nós”³⁰⁴.

O terceiro fragmento é a função das partes na posição original: “o seu objetivo é simplesmente selecionar a concepção mais racional para elas (as partes), dadas as suas circunstâncias”³⁰⁵. Não vejo que essa frase deva ser unicamente lida como “é mais vantajoso” para elas (embora seja esse o objetivo genuíno de Rawls com o uso do termo racional nesse contexto específico). Se nós a lermos a partir do ponto de vista das partes, então essa leitura *pode* fazer sentido (digo pode porque, como tentei mostrar, não há algum ato intencional no interior da posição original). Porém, o ponto de vista relevante para o ERA não é o das partes, mas os das pessoas reais³⁰⁶. Assumindo esse ponto de vista, podemos traduzir “o mais racional para eles, dadas as suas circunstâncias” como significando o mesmo que “epistemicamente mais

²⁹⁹ Daniels esboçou um programa de mostrar que crenças em ERA podem ser evidência para a verdade (DANIELS, 1996) e Brink concretizou esse programa ao tentar compatibilizar coerentismo e realismo (BRINK, 1989) a partir da mesma compatibilização no âmbito não-moral fornecida por BonJour (1985). Thomas adotou o objetivo de relacionar ERA e verdade moral, porém em termos de um fundacionalismo moderado contextualista (THOMAS, 2006). Recentemente o ERA pode ser visto no novo intuicionismo proposto por Audi (1997). Como uma questão de análise histórica, identifica-se uma tendência de interpretar o ERA como um coerentista nas décadas de 70 e início da década de 80, e uma tendência para interpretá-lo segundo o fundacionalismo nos últimos anos.

³⁰⁰ RH, p.386-387.

³⁰¹ “A state reached when people’s moral convictions satisfy certain conditions of rationality” (IMT, p.290).

³⁰² “Citizen’s general convictions, first principles, and particular judgments are at last in line” (RH, p.384n16).

³⁰³ KC, p.306, p.340, p.355.

³⁰⁴ “What justifies a conception of justice is not its being true to an order antecedent to and given to us, but its congruence with our deepest understanding of ourselves and our aspirations, and our realization that, given our history and the traditions embedded in our public life, it is the most reasonable doctrine for us” (KC, p.307).

³⁰⁵ “The aim is simply to select the conception most rational for them, given their circumstances” (KC, p.350).

³⁰⁶ “The third point of view – that you and me – is that from which justice as fairness, and indeed any other doctrine, is to be assessed. Here the test is that of general and wide reflective equilibrium” (KC, p.321).

apropriado para nós, dadas as demais crenças que sustentamos aqui e agora” (uma vez que a posição original é construída, como tentei mostrar, a partir das nossas crenças e é destinada para extrair as consequências dessas crenças). Com isso os princípios são “os mais razoáveis para aqueles que concebem a sua pessoa tal como ela é representada no procedimento de construção”³⁰⁷ (o razoável nessa sentença possui um sentido mais genérico como “mais adequado”), e a posição original nada mais é do que “um dispositivo de representação” para “servir como um meio de autoclarificação” e assim “nos ajudar a desenvolver o que nós agora pensamos”³⁰⁸.

Um quarto fragmento relevante é a parte individual do conceito geral de justificação: justificar como “um argumento endereçado” por nós “a nós mesmos quando estamos indecisos”³⁰⁹. Embora nesse fragmento, assim como nos anteriores, Rawls escreve na primeira pessoa do plural, ele deve ser lido como se estivesse escrevendo na primeira pessoa do singular. Desse modo, quando ele diz que a justificação é um argumento endereçado por nós a nós mesmos, o que ele está efetivamente querendo significar é a ideia de um argumento que um indivíduo formula para si mesmo.

Um último fragmento é a afirmação de que “uma vez que nós entendemos a teoria, nós podemos reconhecer essas interpretações como adequados desdobramentos do que nós desejamos sustentar em reflexão”³¹⁰. Assim, dizer que a teoria é justificada em ERA é, aparentemente, apenas dizer que ela é aceitável em reflexão para cada indivíduo razoável particular.

Organizo esses fragmentos como se segue. Assim como Scheffler, Rawls busca “descrever a ideia de justificação racional” (pessoal) fornecendo um método que assegura a racionalidade epistêmica *subjativa* de um senso de justiça. Para tanto o método incorpora certas restrições de racionalidade. A restrição relevante é a de coerência, obviamente: para que eu seja pessoalmente justificado no senso de justiça que possuo, ele deve ser internamente coerente. Mas, como o fundacionalismo moderado defende, ser justificado não é simplesmente ser coerente. Suponha que minhas crenças p_1 e p_2 estejam em franca contradição com o meu sistema coerente de crenças. Eu tenho várias alternativas para me tornar coerente nesse caso: posso descartar p_1 , e p_2 , posso revisá-las, posso alterar o sistema de crenças para que ele comporte

³⁰⁷ “They are most reasonable for those who conceive of their person as it is represented in the procedure of construction” KC, p.355.

³⁰⁸ “As a device of representation the idea of the original position serves a means of public reflection and self-clarification. We can use it to help us work out what we now think” (JPFNM, p.402).

³⁰⁹ TJ, p.580-581/rev, p.508-509.

³¹⁰ “Once we understand the theory, we can acknowledge these interpretations as suitable renderings of what on reflection we now wish to maintain” (TJ, p.451; rev, p.396).

consistentemente p_1 e p_2 , ou posso fazer modificações em p_1 e p_2 e no sistema de crenças ao mesmo tempo. Independentemente da minha escolha preferida, todas as opções me conduziram ao estado de coerência, mas não ao de racionalidade subjetiva ou justificação pessoal. Julgo este ponto como decisivo para o conceito epistêmico de racionalidade de Rawls. Eu devo ser coerente com aquilo que eu convictamente, sem margem para titubeios, avalio ser o correto. O meu objetivo é encontrar uma concepção de justiça que faça o menor número possível de alterações nos meus comprometimentos iniciais (juízos morais ponderados iniciais) e que promova apenas as modificações necessárias para que o meu senso de justiça como um todo (juízos morais ponderados *simpliciter*) maximize esses comprometimentos, que são justificados por um suporte independente. Nesse processo nada está determinado previamente: posso encontrar uma descrição que exija poucas revisões nesses comprometimentos, posso encontrar uma descrição que exija revisões substanciais, posso até mesmo reformular todas as descrições para que elas se ajustem aos juízos iniciais ou, mais provavelmente, posso reformular tanto as minhas convicções quanto as descrições oferecidas. As possibilidades são muitas, nada está decidido *a priori*. O essencial é que ao final do processo o meu senso de justiça seja condizente com aquilo que *me parece*, em *t*, mais profundamente ser o justo, que eu possa ver a orientação fornecida pelo ERA como aceitável para mim e que eu continue aceitando os meus juízos quando tiver de aceitar as suas consequências na prática³¹¹.

Rawls se antecipou ao resultado desse processo. Ele interpretou quais são os nossos comprometimentos mais profundos em questões de justiça: são os nossos juízos mais fixos tais como os juízos de repúdio à escravidão e de tolerância religiosa. Ele entende que as concepções-modelo que compõem a posição original expressam essas convicções que constituem a nossa identidade moral quando questões de justiça estão em disputa. Por isso a posição original: dentre todas as nossas convicções, é necessário selecionar aquelas que são mais importantes para nós. Os seus mecanismos são desenhados para que possamos extrair a consequência dessas convicções. Assim a pergunta fundamental da posição original é a seguinte: dado o que eu acredito ser o correto, o que mais eu devo acreditar em questões de justiça? Se eu acredito que o repúdio à escravidão não deve jamais ser desrespeitado e que todas as pessoas são livres e iguais, que princípios de justiça eu devo sustentar para que eu seja uma pessoa pessoalmente justificada e assim epistemicamente racional? A justiça como equidade surge como a resposta a essas questões.

³¹¹ *IMT*, p.289.

Dentre todas as apropriações existentes do equilíbrio reflexivo, a mais sistemática certamente é a de DePaul. Não posso fazer justiça aqui a tudo o que DePaul tem a dizer em seu livro, em parte em função do meu objetivo de analisar o equilíbrio reflexivo em Rawls, mas o que há de interessante na sua proposta é a utilização do equilíbrio reflexivo que parte da assunção de que a sua confiabilidade para indicar o que conta como conhecimento é impossível de ser estabelecida. É justamente essa assunção que impõe a rejeição do ERE e que os críticos tomam também como uma objeção fatal ao ERA. DePaul pergunta por que a justificação tem que ser conducente à verdade. Por que, ele quer saber, nós não podemos ficar satisfeitos meramente em possuir crenças justificadas de acordo com os nossos próprios padrões, ao invés de buscar por crenças que satisfazem um critério externo (como a verdade)?³¹² Grande parte dos esforços de DePaul é direcionada contra a tese de que, para adotar as suas palavras, o *summum bonum* epistêmico é a crença verdadeira. A partir disso ele constrói o que denomina de uma concepção experiencial da justificação (*experiential warrant*): a tese segundo a qual estamos justificados quando há uma conexão apropriada entre o caráter fenomenológico da experiência e o conteúdo das nossas crenças. A relação epistêmica não se dá entre esse conteúdo e a realidade em si, mas entre o conteúdo e o modo como a realidade *parece* para nós³¹³. Com isso DePaul chega a uma concepção subjetivista e internalista de racionalidade: ser racional é uma questão de sustentar crenças que, da perspectiva do sujeito, satisfazem os seus próprios padrões epistêmicos³¹⁴.

Nesse contexto, a questão da racionalidade das crenças de *S* depende da relação entre essas crenças. O sistema de crenças de *S* é racional, a certa altura DePaul define, se e somente se ele não contiver proposições conflitantes com princípios epistêmicos, princípios esses que são internos ao próprio sistema de crenças³¹⁵. Não creio que DePaul devesse estipular a relação de coerência meramente em termos de consistência lógica, mas essa é uma questão de menor importância. O importante agora é relacionar esse projeto com o de Rawls: segundo *TJ* um sistema de crenças (um senso de justiça) é racional se ele for coerente com certos princípios de justiça de uma teoria moral, princípios que são construídos a partir de componentes especiais desse sistema, portanto internos a ele.

É importante ter em vista duas afirmações que eu não estou fazendo nesta subseção. Em primeiro lugar, não estou equacionando racionalidade com maximização de preferências, no

³¹² DEPAUL, 1993, p.73.

³¹³ DEPAUL, 1993, p.87.

³¹⁴ DEUPAL, 1993, p.102.

³¹⁵ DEPAUL, 1993, p.107.

sentido que nós podemos encontrar em uma teoria da escolha racional. E em segundo lugar, ao caracterizar a justiça como equidade como a interpretação de um senso de justiça justificado em ERA, não estou sugerindo que para Rawls o que é racional para alguém acreditar do ponto de vista racional em primeira pessoa constitui o que é justo de um ponto de vista intersubjetivo. Aqui eu novamente remeto à distinção entre as duas teses, realizada na seção anterior. Por ora, é importante que o leitor tenha em mente que o conteúdo desta subseção representa uma parte da resposta de Rawls em favor dos seus dois princípios, e não toda ela.

O que eu estou efetivamente sugerindo é que a proposta de Rawls com o ERA é uma *concepção epistêmica de racionalidade* em um sentido análogo àquele oferecido por Foley (o próprio Foley afirma, *en passant*, que o ERA deve ser visto como um método cujo *desideratum* é oferecer o tipo de racionalidade que ele está propondo³¹⁶). Cito-o:

A teoria da racionalidade epistêmica é uma teoria de primeira-pessoa e como tal ela tenciona descrever não as proposições cuja verdade duas ou mais pessoas que estão buscando acordo via discussão deveriam acreditar. Ao invés, ela busca descrever o que uma pessoa individual deveria acreditar na medida em que ela quer ter crenças verdadeiras e não ter crenças falsas... A ideia é que uma proposição é epistemicamente racional para uma pessoa apenas se ela depende de premissas que não são controversas para ela e se a pessoa pode argumentar em favor daquela proposição de um modo que é incontroverso para ela³¹⁷.

Foley se refere à teoria da racionalidade como epistêmica porque ela busca avaliar as crenças de uma pessoa tendo em vista o fato de que crenças têm em geral a finalidade específica de se pretenderem verdadeiras e não falsas (para Foley, esse é a única finalidade puramente epistêmica). Não há problema para Rawls concordar com Foley e defender que quem está engajado no ERA tem essa finalidade em vista. Como visto no capítulo passado, é justamente isso que Rawls faz ao descrever o papel das doutrinas abrangentes no *PL*. Apesar disso, não estou insinuando que Rawls esteja comprometido com a posição segundo a qual a verdade é a única finalidade epistêmica relevante em assuntos morais. Em um sentido que não pode ser ignorado, a crença moral possui certas propriedades específicas, e, como consequência, um indivíduo engajado no ERA busca mais do que crença verdadeira. Por exemplo, ele busca também desenvolver um senso de justiça que promova o que ele considera uma vida boa e que seja congruente com a sua compreensão abrangente do mundo. Nesse caso, a verdade ou falsidade das crenças que compõem o seu senso de justiça não é um problema que o indivíduo

³¹⁶ FOLEY, 2004, p.43-44.

³¹⁷ FOLEY, 1987, p.5.

necessariamente precise julgar estar em primeiro plano. Creio que Rawls enfatiza as finalidades práticas da deliberação ética tanto quanto a sua finalidade teórica. Por essa razão, talvez seja mais apropriado se referir ao projeto do ERA como uma concepção epistêmica moral de racionalidade³¹⁸.

Como em alguns momentos eu já disse, o ERA é em Rawls um esboço ou um projeto. Rawls o mantém sempre no nível do projeto porque ele não julga precisar, para cumprir o seu objetivo, detalhar como nós devemos prosseguir para que sejamos epistêmica e moralmente racionais. Se a justiça como equidade puder ser vista como aceitável para nós, e congruente com as nossas crenças mais básicas, então o ERA já cumpriu o seu papel no interior da justiça como equidade³¹⁹.

3.1.4 A ideia de justificação pública

Embora a justificação pública não seja objeto desta dissertação, desejo nesta subseção oferecer algumas notas bastante breves sobre esse conceito. Trata-se de uma oportunidade de esclarecer, ainda que superficialmente, o que devemos entender por equilíbrio reflexivo geral (ERG) e pleno (ERP) e como devemos conceber a relação entre as ideias de justificação pessoal e pública.

O ERG é definido como se segue:

Pense em cada cidadão dessa sociedade [bem-ordenada] como tendo alcançado um equilíbrio reflexivo amplo. Uma vez que os cidadãos

³¹⁸ Além de DePaul, Foley e Scheffler, esta interpretação do ERA deve a Ernest Sosa. Para Sosa, nós começamos a reflexão filosófica (ERA) com intuições e crenças dos mais variados tipos. Essas crenças se configuram como “os pontos de partida psicológicos da reflexão e ocasionalmente mesmo o fundamento (*ground*) para os resultados da reflexão”. Elas não precisam “ter sido inferidas de qualquer coisa, nem pelo sujeito que reflete”. Para Sosa o projeto primário do ERA é o de formular uma concepção de justificação racional ou de *justificação interna*, em uma perspectiva de primeira pessoa. Interessantemente, Sosa recomenda que se deixe a arrogância filosófica de lado, pois a reflexão, “nem mesmo a reflexão mais profunda, abrangente e intensa”, não nos oferece nada além de justificação racional (SOSA, 1993, p.260-266).

³¹⁹ Em uma avaliação menos tolerante, poderia se afirmar o seguinte sobre o ERA: “O problema com essa recomendação é familiar. Ela não é errada, mas inútil. No máximo ela é um meta-conselho. Essencialmente o mesmo meta-conselho que está implícito na noção de racionalidade egocêntrica. De fato, ela essencialmente nos diz isso: leve em consideração todos os dados que você pensar que são relevantes e então reflita sobre eles, resolvendo os conflitos do modo como você julga o melhor. De outro lado, ela não nos conta que tipos de dados são relevantes, nem nos diz qual seria o melhor modo de resolver conflito nesses dados” (FOLEY, 1993, p.128). Essas palavras são de Foley, exatamente comentando o ERA tomado como um método preocupado com a racionalidade epistêmica. Contrariamente a outros (ver BONEVAC, 2004), Foley não faz esse comentário como uma objeção particular ao ERA. Antes, ele cita o ERA para constatar que, assim como qualquer concepção de racionalidade epistêmica, ele não é capaz de oferecer um guia efetivamente prático para as pessoas pensarem cotidianamente. Embora essa crítica seja acertada, ao menos como é formulada por Foley ela parece exagerada. Por exemplo, não é verdade que o ERA não diferencie tipos de dados. As definições do ponderado cumprem essa função. Também é falso que o ERA não prescreve como os conflitos entre crenças devem ser solucionados. Como eu mostrei, sempre se deve dar prioridade àquelas crenças mais fixas, básicas ou sustentadas com maior comprometimento; e se as crenças alegadamente básicas conflitarem em demasia com as demais crenças, então permite-se (de fato exige-se) que elas sejam modificadas. Além disso, que o ERA deixa para os indivíduos resolver os problemas mais fundamentais faz parte da própria concepção de racionalidade do método.

reconhecem que afirmam a mesma concepção pública de justiça política, o equilíbrio reflexivo também é geral: a mesma concepção é afirmada nos juízos ponderados de todos. Assim, os cidadãos têm alcançado um equilíbrio reflexivo geral e amplo, ou o que nós podemos referir como um equilíbrio reflexivo pleno. Nessa sociedade não apenas há um ponto de vista público a partir do qual todos os cidadãos podem adjudicar as suas reivindicações de justiça política, mas um ponto de vista que é mutualmente reconhecido por todos eles como afirmado em equilíbrio reflexivo pleno. Esse equilíbrio é inteiramente intersubjetivo: isto é, cada cidadão leva em conta o raciocínio e os argumentos dos outros³²⁰.

Cidadãos razoáveis, quando possuem um senso de justiça em ERA, necessariamente afirmam a justiça como equidade. No capítulo passado foi analisado como cada cidadão razoável é levado a aceitar a justiça como equidade por razões diferentes, razões dadas por suas doutrinas abrangentes³²¹. Dessa forma, não há um único ERA, mas diversos ERAs, isto é, diferentes conjuntos de crenças justificando pessoalmente de maneiras diversas a justiça como equidade. Mas o conteúdo da justiça como equidade, os seus princípios e os valores políticos que os sustentam, é comum a todos os ERAs (como Rawls escreve, a justiça como equidade funciona como um “módulo”), o que significa que ela é pessoalmente justificada em todos os ERAs. Assim, uma concepção de justiça é também justificada em um equilíbrio reflexivo amplo e geral. Para uma concepção de justiça ser justificada em ERG ela (i) deve ser pessoalmente justificada para cada cidadão razoável e (ii) esse fato, de que todos aceitam a mesma concepção em ERA, deve ser reconhecido por todos.

É tentador ver o ERG meramente como um estado de coisas em que há, em uma determinada sociedade, um elemento comum entre todos os ERAs. Essa interpretação na terceira pessoa do plural deve ser rejeitada. Ela tende a interpretar o ERG como uma mera consequência fortuita do ERA, enquanto que na visão efetiva de Rawls, evidente na sua afirmação de que “cada cidadão leva em conta o raciocínio e os argumentos dos outros”, o ERG abrange também uma determinada concepção de como as pessoas devem deliberar quando precisam justificar para o outro as suas crenças morais e políticas. Assim como o ERA, o ERG funciona na primeira pessoa. Porém, como agora o objetivo é especificar como um indivíduo deve justificar as suas crenças para o outro, e não mais como um indivíduo deve justificar as

³²⁰ “Think of each citizen in such a [well-ordered] society as having achieved wide reflective equilibrium. Since citizens recognize that they affirm the same public conception of political justice, reflective equilibrium is also general: the same conception is affirmed in everyone’s considered judgments. Thus, citizens have achieved general and wide, or what we may refer to as full, reflective equilibrium. In such a society, not only is there a public point of view from which all citizens can adjudicate their claims of political justice, but also this point of view is mutually recognized as affirmed by them all in full reflective equilibrium. This equilibrium is fully intersubjective: that is, each citizen has taken into account the reasoning and arguments of every other citizen” (*RH*, p.384n16; *JF*, p.31).

³²¹ *JF*, p.32.

suas crenças para si mesmo, o ponto de vista é o da primeira pessoa do plural. A questão deixa de ser a de saber quais condições as crenças de *S* têm de satisfazer para que sejam pessoalmente justificadas, e passa a ser a questão de saber que condições as crenças de *S* têm de satisfazer para que sejam racionalmente aceitáveis para os outros, isto é, para que elas sejam também intersubjetivamente justificadas ou como Rawls define, *publicamente justificadas* (*public justification*)³²². A questão não é mais a de saber como um indivíduo deve deliberar, mas de saber como os indivíduos devem conduzir as suas argumentações conjuntamente quando situações de desacordo precisam ser enfrentadas. Com a noção de razão pública Rawls oferece uma caracterização do ERG muito mais detalhada e robusta que a sua caracterização do ERA.

A ideia de justificação pública é definida como se segue:

Justificar os nossos juízos políticos para os outros é convencê-los pela razão pública, isto é, por modos de raciocínio e inferência apropriados para questões políticas fundamentais, e pelo apelo a crenças, razões e valores políticos que sejam razoáveis para os outros também reconhecer. A justificação pública procede de algum consenso: de premissas que todas as partes em desacordo, assumidas como livres e iguais e plenamente capazes de raciocinar, podem razoavelmente esperar compartilhar e endossar livremente³²³.

Na frase imediatamente seguinte a esse excerto, Rawls escreve que a justificação pública não é simplesmente uma questão de endereçar argumentos que possuem validade lógica para os outros. Isso sugere fortemente que a ideia de justificação pública supõe uma noção de bom argumento. Mas o que pode contar como bom argumento? Para Rawls, a justificação pública associada ao ERG é caracterizada pelo seu objetivo prático: alcançar um acordo razoável em questões de justiça³²⁴. Essa afirmação pouco esclarece, pois é preciso saber o que significa um “acordo razoável”.

³²² A tese de Rawls é a de que a justiça como equidade descreve o conjunto de crenças que um cidadão pode publicamente justificar para o outro. Por isso ela é publicamente justificada. Note que a ideia de justificação pública e o ERG, assim como a ideia de justificação pessoal e o ERA, podem ser isolados do conteúdo da justiça como equidade. Como muitos argumentaram, Rawls pode estar errado ao defender que a justiça como equidade descreve esse conjunto de crenças, mas isso não implica que a sua concepção de justificação pública está errada. E se nenhuma concepção política de justiça puder satisfazer as condições para a justificação pública, isso também não implicará imediatamente a necessidade de recusar a ideia de justificação pública. É possível assumirmos o ponto de vista cético de que Rawls descreve adequadamente as condições de adequação para a justificação pública, mas que nenhuma crença é capaz de satisfazer satisfatoriamente esses padrões. Gaus adota um raciocínio nessa linha. Como resultado, ele transfere os problemas de justificação pública para serem resolvidos no âmbito da justificação política e legal (GAUS, 1996, p.190-191).

³²³ “To justify our political judgments to others is to convince them by public reason, that is, by ways of reasoning and inference appropriate to fundamental political questions, and by appealing to beliefs, grounds, and political values it is reasonable for others also to acknowledge. Public justification proceeds from some consensus: from premises all parties in disagreement, assumed to be free and equal and fully capable of reason, may reasonably be expected to share and freely endorse” (*JF*, p.27).

³²⁴ *JF*, p.32.

Rawls escreve que para a justiça como equidade ser publicamente justificada “ela tem de ser aceitável não somente para nossas próprias convicções ponderadas”³²⁵. Dessa forma, quando *S* procura justificar uma crença, *p*, para uma outra pessoa razoável qualquer, digamos *S**, parece que esta condição precisa ser satisfeita:

(i) *p* tem de ser pessoalmente justificada em ERA para *S*

(i) implica que se *S* tenta justificar *p* para *S** sabendo que *p* é uma crença falsa, *p* não é publicamente justificada. *S* tem de ser sincero e justificar para o outro uma crença que ele julga que é razoável. A crença de *S* de que *p* é correta e razoável pode estar errada, mas o importante é que *S* esteja sendo sincero e não esteja tentando deliberadamente enganar *S**. Acordos em que (i) não é satisfeita não são razoáveis. Assim, a justificação epistêmica exige certos comprometimentos morais, como o respeito à reciprocidade e ao dever de civilidade.

Entretanto, (i) não é suficiente para a justificação pública. Na citação acima Rawls também defende que *S* deve justificar *p* para os outros a partir de crenças, razões e valores políticos que sejam razoáveis também para eles. Essa afirmação sugere uma segunda condição para a justificação pública:

(ii) *p* tem de ser pessoalmente justificada em ERA para *S**

Assim, se *S** aceita *p* apenas porque aceitar *p* lhe permitirá em um momento futuro viver em uma sociedade na qual os seus valores aniquilarão os valores dos outros, ou se *S** aceita *p* porque *S* lhe obriga pela força a aceitá-la, ou porque tem medo de *S*, *p* não será publicamente justificada e o acordo não será razoável. Além de pessoalmente justificada para *S*, *p* tem de poder ser aceita em ERA por *S**: “reflexão racional é o que, dentre outras coisas, distingue justificação pública de um mero acordo”³²⁶.

(i) e (ii) conjuntamente definem a ideia de que o consenso sobreposto é um acordo “razoável” alcançado pelas “razões corretas”: todos que acordam a concepção política “partem da sua própria visão abrangente e a fundamentam nas razões religiosas, filosóficas e morais que

³²⁵ *JF* p.27.

³²⁶ “It is this last condition of reasoned reflection that, among other things, distinguishes public justification from mere agreement” (*JF*, p.29).

ela fornece”³²⁷. Já que cada um afirma a concepção de justiça em ERA, todos acreditam sinceramente na sua correção, razoabilidade ou verdade.

O argumento de Rawls é dúbio quanto a essas duas alegações:

(iiia) *S* tem de efetivamente convencer *S** a aceitar *p*

(iiib) *S* tem de mostrar a *S** que, dadas as suas crenças, ele poderia justificadamente aceitar *p*

Não é claro se Rawls se compromete com (iiia) ou (iiib). Ele escreve, em alguns momentos, como se o objetivo da filosofia política fosse buscar por crenças que podem ser objeto de um consenso efetivo, mas em outros momentos ele dá a entender que ela deve buscar apenas por crenças que poderiam ser objeto de um consenso hipotético³²⁸. De uma forma geral, Rawls parece tender a (iiib), mas a questão é controversa. Para os presentes propósitos, se ater a (i) e (ii) é suficiente.

Rawls acredita que uma concepção política de justiça é construída a partir de valores políticos, definidos como aqueles valores que todas as pessoas razoáveis aceitam ou podem aceitar juntamente com as suas doutrinas abrangentes razoáveis. Quando precisam resolver os desacordos em questões essenciais de justiça, as pessoas devem valer-se desses valores políticos para, a partir deles, mostrar para as outras pessoas que elas poderiam aceitar aqueles determinados princípios de justiça e a correspondente resolução para questões de justiça que eles propiciam³²⁹. É por isso que a justificação pública sempre procede de algum consenso, ou

³²⁷ “All those who affirm the political conception start from within their own comprehensive view and draw on the religious, philosophical, and moral grounds it provides” (PL, p.147).

³²⁸ Ver, por exemplo, a afirmação de que não há garantias de que a justiça como equidade será aceita em consenso sobreposto (JF, p.37).

³²⁹ Além das ideias de justificação pessoal e pública, Rawls também utiliza a noção de justificação *pro tanto*. O que faz parte de uma concepção de justiça é, além dos princípios que ela propõe, os valores e ideais que sustentam internamente esses princípios. Na citação acima Rawls afirma que o conteúdo de uma concepção política de justiça afirmada em equilíbrio reflexivo geral constitui uma base pública para a justificação. Isso significa que quando questões de justiça estão em jogo os cidadãos podem (na verdade devem) recorrer ao seu conteúdo a fim de resolver o dissenso. Eles podem tanto reivindicar os seus princípios quanto os valores que fundamentam os princípios, mostrando para o outro que ele deve aceitar aqueles princípios porque eles são apropriadamente justificados pelos valores, ou que ele deve aceitar os valores porque eles são justificados pelos princípios, ou que ele deve aceitar alguns valores porque eles estão justificados com base em outros valores. O sentido do termo justificação empregado nas sentenças acima é o *pro tanto* (RH, p.386). Certamente não pode ser o de justificação pessoal, pois apenas está sendo afirmado que a estrutura de uma concepção de justiça é de tal modo internamente coerente e sofisticada que os seus valores e princípios podem ser arranjados de uma maneira tal que é possível extrair deles respostas para todas ou quase todas as questões prementes de justiça. Como a ideia de justificação pessoal versa sobre a relação entre o conteúdo dessa concepção e os valores morais abrangentes sustentados pelos cidadãos particulares aos quais a concepção de justiça se aplica, identifica-se facilmente por que a noção de justificação empregada nas sentenças acima não pode ser a pessoal. Pela mesma razão não pode ser a pública.

Presumivelmente, uma concepção de justiça pode ser justificada *pro tanto* sem poder ser justificada em um equilíbrio reflexivo geral. Esse é o caso de teorias morais abrangentes: elas são internamente consistentes e completas e oferecem respostas para todas ou quase todas as questões essenciais de justiça, mas como elas não são

seja, de premissas que as partes em desacordo já aceitam *antes* de iniciar o debate. Quando não há esse tipo de consenso, um indivíduo não pode “acionar” as crenças do outro que o levarão a aceitar a conclusão para o seu argumento. Presumivelmente, um bom argumento de justificação pública é aquele que tem como apoio de sua conclusão uma ou mais premissas que o outro aceita.

Note que o modo de apresentar a ideia de justificação pública a partir do diálogo de duas pessoas pode levar a mal entendidos. Se *S* e *S** forem membros de um mesmo grupo religioso, e compartilharem os mesmos valores morais, eles poderão “publicamente” justificar um para o outro uma série de crenças que eles não poderiam justificar para todos os membros da sua sociedade. Entretanto, penso que o ERG pode ser utilizado para descrever adequadamente o que estaria envolvido nessas interações justificacionais entre um número reduzido de indivíduos. O uso efetivo de Rawls entende que a justificação pública envolve todos os cidadãos da sociedade, e uma crença publicamente justificada seria uma crença que todos poderiam aceitar. Esse uso efetivo, no entanto, apenas transfere o uso dos critérios de justificação pública para um âmbito maior, o âmbito da cultura política pública. Ele poderia também estendê-la para o âmbito internacional, como efetivamente fez em *The Law of Peoples*.

E quanto ao ERP, o que significa? É possível haver uma crença que satisfaz apenas (i). Esse é o caso de crenças morais abrangentes, que embora possam ser pessoalmente justificadas para um número limitado de pessoas, não podem ser pessoalmente justificadas para todos. Conforme expõe Gaus, epistemologicamente falando o pluralismo razoável nada mais é do que o fato de que nem tudo o que é pessoalmente justificável pode ser publicamente justificável³³⁰. Rawls acredita, entretanto, que uma concepção política de justiça, assim como as crenças deriváveis dela (por exemplo, a tolerância é razoável, você deve ser tolerante etc.), satisfazem (i) e (ii). Nesse caso, essas crenças são justificadas em um equilíbrio reflexivo que é amplo e geral ao mesmo tempo, ou um equilíbrio reflexivo *pleno*.

Veja que embora possa haver crenças justificadas em ERA e não em ERG, não pode haver crenças justificadas em ERG e não em ERA. Isso porque o ERG é definido tendo o ERA como o seu componente. Isso demonstra, como alguns filósofos têm argumentado, que a noção de justificação pessoal é mais fundamental que a noção de justificação pública³³¹. Em última

políticas, elas não podem ser aceitas por todas as pessoas razoáveis. Assim, uma teoria pode ser *pro tanto* justificada sem ser pessoalmente justificada para todos os cidadãos razoáveis, e, conseqüentemente, sem ser afirmada e reconhecida por todos. Nesse caso a teoria, embora constitua um conjunto de valores e princípios a partir dos quais seja possível oferecer respostas para questões essenciais de justiça, não constitui uma base para a justificação pública.

³³⁰ GAUS, 1996, p.11.

³³¹ Por exemplo, Goldman (1997).

instância, é a noção de justificação pessoal que estipula os limites para aquilo que pode ser publicamente justificado.

Não pretendo nesta dissertação ter esclarecido satisfatoriamente a complexa relação entre as ideias de justificação pessoal e pública. Em virtude da ênfase na ideia de que a justificação pessoal é uma noção subjetiva e não intersubjetiva, o leitor pode julgar que esta dissertação esteja adotando uma versão demasiado restritiva a respeito das interações possíveis entre o ERA e o ERG. O leitor pode pensar que para esta dissertação quando um indivíduo está revisando as suas crenças em ERA, ele não deve interagir com outras pessoas, não deve considerar a opinião delas e o modo como elas reagem às suas crenças. O ERG, segue o leitor, seria uma mera consequência fortuita de indivíduos racionando isoladamente. Essa imagem parece implausível. Se há convergência a partir do emprego do ERA, é porque ao buscar o que é pessoalmente justificado o indivíduo já leva em conta a opinião dos outros. A sua busca por ERA e ERG não ocorreria separadamente, como dois processos distintos, mas funcionaria com um único processo. Se um indivíduo não consegue justificar em ERG uma crença que é justificada em ERA, esse fato pode levá-lo a alterar o resultado do ERA.

Concordo com o leitor: essa interconexão entre ERA e ERG aponta para uma considerável incompletude da caracterização da ideia de justificação pessoal oferecida nesta dissertação. É por isso que a apresento apenas como uma esquematização, algo para ser aperfeiçoado. Apesar disso, creio que a natureza subjetivista do ERA não pode ser colocada em questão. Afirmar que o indivíduo deve levar em consideração a opinião dos outros, e que a busca por crenças justificadas em ERA é moldada pela sua busca por crenças publicamente justificadas, não é o mesmo que afirmar que o indivíduo deve, para ser racional, se adequar àquilo que os outros aceitam nos seus ERAs. Ao abordar o internalismo de razões, no capítulo passado, foi mostrado que para um internalista as razões para a ação são dadas pelos desejos, crenças e demais elementos do *set* motivacional do indivíduo. Em certo sentido, porém, o internalista não precisa negar que razões são dadas por fatos. Alguém deve parar de fumar porque fumar pode provocar a morte, e a morte é em si mesma um fato indesejável da vida. O que o internalista afirma é que o fato de que fumar causa morte apenas contará como uma razão se o indivíduo escolher, em ERA, deixar de fumar porque pode provocar a morte. Creio que um raciocínio semelhante pode ser aplicado para o caso do ERA versus ERG: se o indivíduo, em ERA, não ver razão para aceitar a opinião do outro, então não será irracional para ele ignorá-la.

As observações desta subseção são de cunho alusivo. Rawls tem muito mais a dizer, a partir da sua ideia de razão pública, a respeito da questão sobre como devemos justificar as

nossas crenças uns para os outros e qual a natureza do consenso que advém desse processo de justificação pública. O conceito de justificação pública está intimamente conectado às noções de estabilidade pelas razões corretas (*stability for the right reasons*) e legitimidade. Essas noções possuem um papel epistemológico no argumento geral de Rawls que não foi analisado. Entretanto, penso que os comentários acima fornecem ao leitor uma indicação de como a interpretação do ERA defendida nesta dissertação não apenas é compatível com os demais elementos da filosofia de Rawls, como permite compreendê-los melhor.

3.2 Duas objeções reconsideradas

3.2.1 Conservantismo epistêmico

Como caracterizado no capítulo passado, o conservantismo é a posição segundo a qual alguém está justificado em sustentar a proposição p meramente por p ser uma proposição sustentada com comprometimento. Alternativamente, o conservantismo afirma que alguém está justificado em sustentar p meramente se p for uma proposição insuspeita, isto é, se ela não conflitar, em t , com outras proposições que o indivíduo aceita; isso significa que alguém está epistemicamente autorizado em sustentar p mesmo sem possuir uma razão positiva para crer em p . O ERA é conservantista? Pelo o que já foi exposto até este momento, parece evidente que sim: o que está em jogo é a maximização dos nossos pontos fixos provisórios, aquilo que meramente nos parece o correto. Porém, podemos considerar as crenças que alguém sustenta em ERA de uma dupla perspectiva. Consequentemente, podemos levantar essa questão (“a justificação em ERA implica o conservantismo?”) com dois conceitos distintos de justificação em mente, e, assim, obter duas conclusões diferentes.

Podemos falar em uma perspectiva interna. Essa é a perspectiva de quem emprega o ERA e o utiliza a fim de alcançar um senso de justiça coerente. Como eu busquei apontar ao caracterizar o fundacionalismo moderado, do ponto de vista de quem adota o ERA as crenças supostamente básicas não são apenas crenças meramente sustentadas com comprometimento ou crenças simplesmente insuspeitas³³²; antes, elas são crenças não-inferencialmente justificadas, isto é, crenças óbvias, indubitáveis e inquestionáveis para aquele que as sustentam. Nesse quadro, as crenças básicas são *prima facie* justificadas do ponto de vista interno³³³. Uma

³³² Gostaria de deixar em aberto a questão de se o ERA é conservantista em uma perspectiva interna se interpretado como uma forma de coerentismo.

³³³ Creio que se visto como uma forma de fundacionalismo moderado, como eu propus neste capítulo, o ERA também não é circular. Goodman afirmou que um juízo é legítimo se conforme a um princípio válido, e um princípio é válido se conforme a um juízo legítimo. Com base na minha distinção entre juízos morais ponderados iniciais e juízos *simpliciter*, é possível visualizar que essa circularidade não é tão dramática. Eis como eu reescrevo a definição de Goodman: um juízo *simpliciter* é justificado se ele está conforme a princípios válidos, e um princípio é válido se concorda com juízos morais ponderados iniciais. A estrutura aqui não é de circularidade porque há, da

crença meramente sustentada não é (pessoalmente) justificada para o ERA: ela tem de ser sustentada com uma atitude doxástica especial (com muito comprometimento), assim sendo uma crença básica, ou ela deve ser justificada por estar em uma relação apropriada com uma crença básica. A propriedade de ser “sustentada com alto grau de comprometimento” é uma propriedade epistêmica da crença, uma propriedade que a torna não-inferencialmente justificada para aquele que a sustenta.

Os detalhes desse fundacionalismo precisam ser considerados em um trabalho futuro, claro. Entretanto, quaisquer que sejam os seus detalhes, parece inevitável construí-lo como um fundacionalismo subjetivista mais ou menos nos moldes daquele defendido por Foley. De acordo com o fundacionalismo subjetivista de Foley, uma crença p é básica para S se e somente se é incontroverso para S aceitar p . S tem de sustentar p com mais confiança com que sustenta outras crenças que poderia utilizar para argumentar contra p , e p tem de ser tal que S , em reflexão, tem de acreditar que quando sustenta p a sua crença em p é verdadeira³³⁴. Essa definição de crença básica acentua a propriedade de ser “psicologicamente convincente”³³⁵ *para o indivíduo que a sustenta* (por isso o subjetivismo). Note como essa definição de crença básica independe da verdade ou qualquer outro critério não-subjetivista de correção de crenças. Se uma proposição não é incontroversa para S , no sentido definido por Foley, então nenhuma garantia de verdade a torna propriamente básica. Se S não estiver ciente de que a sua crença é verdadeira ou até mesmo infalivelmente verdadeira, essa crença não será básica porque S terá bons argumentos contra ela. A propriedade de ser básica é a propriedade de ser “subjetivamente persuasiva”³³⁶. Se a justificação termina em crenças justificadas não-inferencialmente, e não crenças meramente sustentadas, então ela não é conservantista. A tese central subentendida no princípio conservantista é que uma crença possui alguma presunção epistêmica em seu favor desde que ela não seja epistemicamente suspeita, isto é, desde que não haja, no interior do sistema de crenças ao qual ela pertence, um subconjunto de proposições que confirma a sua negação. Ausência de suspeita epistêmica não torna uma crença justificada para o ERA.

perspectiva interna de quem adota o ERA, um ponto fundacional em que a justificação simplesmente para (afirmar que o ERA não é circular não é o mesmo que defender que o argumento geral de *TJ* em favor dos dois princípios não é circular). Como foi ressaltado, isso não significa que os juízos morais iniciais são epistemicamente independentes dos demais juízos. Este é o ponto em que o fundacionalismo moderado se mostra compatível com o holismo e com a ideia de justificação por ajuste mútuo. O movimento essencial para compreender essa compatibilidade é construir a relação de coerência entre os juízos iniciais e os demais elementos negativamente e não positivamente.

³³⁴ FOLEY, 1987, p.68.

³³⁵ FOLEY, 1987, p.70.

³³⁶ FOLEY, 1987, p.71.

Por outro lado, de um ponto de vista externo, isto é, da perspectiva de um terceiro que avalia as crenças que alguém professa em ERA, há necessariamente conservantismo. Se nós desafiarmos ao mesmo tempo e de uma só vez a totalidade de crenças que alguém sustenta em ERA, exigindo-lhe uma razão pela qual confiar na totalidade de suas crenças, essa pessoa, por meio do ERA, não disporá de meios para oferecer uma resposta que nós como observadores julgaríamos convincente. Tudo o que ele poderá responder, utilizando o ERA, é que essas são suas crenças e que é assim que ele enxerga as coisas. Nós então diríamos que o ERA endossa o princípio conservantista porque, nesse caso, o indivíduo está justificado em ERA a despeito de a sua crença ser meramente sustentada por ele, ou ser apoiada por outras crenças também meramente sustentadas. O ato doxástico de sustentar uma crença com um alto grau de comprometimento torna uma crença não-inferencialmente justificada apenas do ponto de vista de quem sustenta a crença em questão. Para outras pessoas, essa crença não terá nenhum *status* epistêmico diferenciado.

Mas há um problema grave em se levantar a questão externa. O problema é confundir e ignorar o projeto do ERA. Como visto anteriormente, o *desideratum* epistêmico do ERA é especificar quando alguém está racionalmente justificado em sustentar p , e não se p é uma proposição verdadeira ou publicamente justificada. O conservantismo apenas emerge quando estamos com o segundo tipo de justificação em mente (se o leitor olhar em retrospectiva para o primeiro capítulo, perceberá que esse foi exatamente a objeção desenvolvida contra o ERA no *Outline*). Para ilustrar esse ponto, convém abordar a objeção clássica ao ERA proposta por Brandt.

Dentre todas as objeções conservantistas contra Rawls, a de Brandt é certamente a mais incisiva. Segundo ele o ERA constitui uma nova forma de intuicionismo que embora não trate mais as nossas convicções como apreensões de uma ordem de valores independente da mente, ordena-as, epistemologicamente, pelo nível inicial de confiança (*credence*) com que elas são sustentadas. O caráter metodológico do ERA para Brandt consiste em propor revisões necessárias para que o sistema resultante de crenças normativas “proteja” a soma total de confiança inicial melhor do que um sistema alternativo. A preferência é depositada no sistema que nós “achamos mais fácil de sustentar como um todo”. Até este ponto, a caracterização de Brandt está de acordo com a oferecida nesta dissertação. O problema é o que é dito a seguir. Para Brandt, esse tipo de abordagem é insustentável ao menos que seja mostrado que o nível de confiança inicial corresponda a um nível de credibilidade epistêmica. E é especialmente

importante que o proponente do ERA ofereça isso porque “vários fatos sobre a gênese das nossas crenças morais militam contra o mero apelo a intuições em ética”³³⁷.

O relevante aqui é ter em mente o que Brandt quer dizer com credibilidade. Penso ser justo dizer que Brandt entende por condução à verdade, do que se segue que uma crença credível é mais conducente à verdade que uma crença alternativa não-credível. Desse modo, o que há de problemático em contar apenas com o nível de confiança inicial é que ele pode nos levar “para longe da verdade ao invés de para perto dela”³³⁸. Rawls teria de mostrar, o que ele não faz, que a confiança inicial corresponde a esse tipo de credibilidade epistêmica.

Como os meus comentários anteriores indicam, há um erro nessa objeção quando ela é direcionada exclusivamente ao ERA. Ela basicamente não identifica corretamente o objetivo do ERA e o julga tendo como referência um problema que o ERA não se propôs a resolver. Se nós seguirmos Brandt, teremos de concluir que o ERA falha porque Rawls não desenvolve algo nos moldes do argumento metajustificatório oferecido por BonJour ou Brink para demonstrar que a coerência entre juízos firmemente sustentados é conducente à verdade (ou que Rawls devesse assumir algum intuicionismo moral associado com alguma forma de realismo). Mas Rawls deveria fornecer um argumento assim apenas se ele quisesse defender o conhecimento moral a partir do ERA, o que não é o caso. Tudo o que Rawls espera conquistar é a racionalidade de um conjunto de princípios para pessoas como nós, dadas as nossas crenças morais. Nós podemos ser racionais e, ao mesmo tempo, distantes do conhecimento moral assim entendido³³⁹.

A crítica de Brandt pode ser respondida indiretamente, mostrando que ela mistura finalidades epistêmicas distintas. O ERA pode operar com uma equivalência entre comprometimento inicial e credibilidade inicial se entender essa credibilidade como conducente à racionalidade pessoal. Do ponto de vista de quem adota o ERA, uma crença com credibilidade inicial é uma crença sustentada com o maior nível de comprometimento, o que implica que é uma crença *prima facie* justificada. O ERA não tem como adequadamente realizar essa equiparação, como foi defendido no primeiro capítulo, porque ele não está (tão-somente) descrevendo a ideia de justificação racional. Note que essa estratégia de resposta é consideravelmente distinta daquela adotada por Daniels. Daniels está de acordo com Brandt: os pontos fixos provisórios não podem ser credíveis meramente porque são inicialmente

³³⁷ BRANDT, 1998, p.16-21.

³³⁸ BRANDT, 1998, p.20.

³³⁹ Alguém pode querer utilizar outros sentidos para o termo “conhecimento”, como, por exemplo, saber aplicar e compreender regras socialmente validadas. Nesta dissertação eu utilizo a expressão “conhecimento moral” para designar um tipo de conhecimento proposicional acerca do mundo.

sustentados com confiança. Como ele diz, “*alguma* resposta para a questão sobre a confiabilidade dos nossos juízos morais tem de ser apresentada”³⁴⁰. O desacordo de Daniels com Brandt concerne ao momento “premature” no qual Brandt faz a sua exigência de justificação dos juízos morais ponderados. No raciocínio de Daniels, uma teoria moral aceitável em ERA pode, em (e depois do) ERA, legitimar os juízos morais ponderados iniciais que, antes do ERA, nós não tínhamos por que confiar. Com isso Daniels acaba assumindo uma distinção velada entre juízos morais ponderados iniciais, pré-ERA, e juízos ponderados não-iniciais, pós-ERA. Apenas os últimos seriam justificados³⁴¹. O problema com essa resposta é que ela não é necessária e impinge ao ERA algo que Rawls não pretendeu alcançar com ele. Além disso, Daniels se compromete com a tese de que a relação de coerência é suficiente para a justificação, já que o próprio ERA seria capaz de gerar, por meio das suas relações de ajuste mútuo, a justificação necessária para as crenças iniciais. Essa é uma tese que os coerentistas tradicionalmente possuem dificuldades em sustentar convincentemente. O próprio Daniels nunca argumentou satisfatoriamente em favor dela. Ao construir o ERA como uma forma de fundacionalismo moderado, esta dissertação esquivava-se desse problema.

3.2.2 Descritivismo

Argumentei que o ERA não pode ser conservantista. Apesar disso, talvez possa ser possível acusá-lo de conservadorismo em virtude da sua metodologia e da forma como caracteriza a ideia de justificação pessoal. Poderia ser alegado que o ERA favorece que o indivíduo assuma uma postura demasiadamente *conservadora* em relação aos seus próprios juízos morais ponderados ou que o ERA não dispõe de meios para fazer com que aqueles indivíduos profunda e dogmaticamente apegados às suas crenças morais, por mais moralmente absurdas que essas crenças possam parecer a um espectador externo, possam progredir moralmente. Aqui eu deixo de usar o conceito de conservantismo epistêmico para utilizar o termo “conservador” tal como intuitivamente entendemos a palavra. Partindo desse raciocínio conservadorista, o crítico do ERA pode alegar que ele é descritivista, isto é, que ele é insuficientemente normativo por estipular o que as pessoas devem crer a partir do que elas efetivamente creem.

Tenho duas objeções em mente. Ambas podem ser encontradas em Brandt e uma delas é explicitamente feita por DePaul. Receio que não disponho dos recursos para lidar satisfatoriamente com todas as alegações envolvidas nessas objeções. No que segue as

³⁴⁰ DANIELS, 1996, p.32.

³⁴¹ DANIELS, 1996, p.31.

abordarei, um tanto quanto simploriamente, e sugerirei, também superficialmente, como eu penso que Rawls pode lidar com elas. Minha pretensão será tão somente a de delinear a que tipo de problema essa minha caracterização do ERA conduz.

Começando por Brandt. Em *A Theory of the Good and the Right*, o projeto de Brandt é o de buscar uma concepção moralmente neutra de racionalidade epistêmica. Quando critica Rawls pelo seu conservantismo, conforme visto na subseção passada, Brandt defende que nós devemos nos transportar para fora dos nossos valores e da nossa cultura, para avaliar a sua racionalidade não de dentro, de um ponto de vista “conservador” e viciado, mas de fora. Somente nessa perspectiva nós não correríamos o risco, risco que Rawls não teria como evitar, de meramente reorganizar preconceitos morais postos pela nossa tradição moral³⁴².

Para cumprir esse projeto Brandt busca uma definição da racionalidade de uma ação, desejo ou código moral em termos do que sobreviveria ao escrutínio da lógica e dos fatos *apenas*³⁴³. Uma ação é completamente racional se e somente se os desejos e aversões do agente são racionais e todos os *inputs* cognitivos relevantes e disponíveis no momento da ação são os mais adequados. Um *input* cognitivo relevante é uma informação que, se presente, faria com que o agente não realizasse a ação. Uma informação disponível é uma informação que o agente poderia ter (não que ele tem) no instante em que age, como os resultados aceitos pela ciência do seu dia, as proposições factuais publicamente aceitas e os princípios da lógica³⁴⁴.

Brandt propõe um processo de reflexão para identificar ações racionais e “racionalizar” desejos que ele denomina de psicoterapia cognitiva (*cognitive psychotherapy*). Se trata de um processo que, grosso modo, conta “simplesmente com a reflexão a partir da informação disponível, sem a influência do prestígio de alguém, do uso de linguagem avaliativa, punição ou recompensa externa ou o uso de estados artificialmente induzidos, como relaxamento”³⁴⁵.

Dados esses pressupostos, é possível ver que Brandt rejeitaria a racionalidade alcançada pelo ERA ao menos por duas razões. Primeiramente, o que é racional para uma pessoa deve ser também racional para todos os seres racionais que compartilham a mesma estrutura psicológico-biológica e podem estar em contato com as mesmas evidências. Como consequência, o racional deve ser identificado por um processo reflexivo guiado pelos fatos e a lógica apenas, e não pelos comprometimentos morais iniciais do agente (eles são subjetivos e relativos). Essa seria a única forma, para Brandt, de nos assegurarmos de que a teoria moral não estaria sendo

³⁴² BRANDT, 1998, 21-22.

³⁴³ BRANDT, 1998, p.10.

³⁴⁴ BRANDT, 1998, p.10-13.

³⁴⁵ BRANDT, 1998, p.113.

conservadora e conivente com preconceitos morais enraizados. Em segundo lugar, a racionalidade é completamente independente das minhas crenças atuais, e deve ser remetida às crenças que eu teria se considerasse todas as informações relevantes disponíveis. Como o ERA apenas postula que o indivíduo deve buscar por descrições alternativas do seu senso de justiça na tradição moral da nossa sociedade, tal como ela é expressa pelas habituais teorias morais, ele não exaure todas as informações disponíveis, consequentemente não bloqueando posturas conservadoras.

Como um proponente do ERA poderia responder a Brandt? Ele pode começar rejeitando a premissa básica por trás de toda a filosofia moral desenvolvida por Brandt, a saber, a premissa de que os fatos e a lógica são valorativamente neutros e é possível assumir algum ponto de vista livre de pressuposições avaliativas. Daniels, Thomas e Silveira parecem argumentar nesse sentido nas suas defesas do ERA. Certamente essa alegação inviabiliza o projeto de Brandt, e mostra que grande parte das suas reservas quanto ao ERA são infundadas. Mas o que dizer da segunda razão? A sua cogência é inteiramente independente da primeira. Pelo senso comum (aparentemente) sabemos que a racionalidade deve ser remetida não às crenças reais que o agente tem, mas àquelas que ele teria se estivesse completamente informado. Esse ponto me conduz à objeção de DePaul.

Nós dizemos de uma crença que ela é racional se ela está de acordo com as evidências que o indivíduo tem para suportá-la. Para adotar um exemplo familiar, se eu acredito que os cientistas são confiáveis geradores de conhecimento, e um cientista publica um artigo em uma revista que eu acredito ser criteriosa nas suas publicações, então essas crenças são evidências que eu disponho para racionalmente acreditar na verdade do conteúdo do artigo. Porém, intuitivamente a racionalidade parece requerer boas evidências e não apenas evidências que eu tenho. Talvez eu devesse saber que essa revista é financiada por uma determinada empresa farmacêutica, e que o artigo em questão é voltado para demonstrar a eficácia de um fármaco dessa empresa. Se eu estivesse ciente dessa evidência, o mais racional seria, em condições normais, checar a veracidade da informação em outras fontes, preferencialmente em cientistas independentes. O fato de eu não verificar outras fontes mostra que a minha crença na confiabilidade do artigo é irracional, pois ela não é uma crença bem informada e baseada em todas as evidências relevantes que eu deveria ter³⁴⁶. É claro que apenas um ser onisciente possui

³⁴⁶ Há inúmeras complicações neste cenário hipotético que eu estou ignorando. Por exemplo, pode ser o caso de que a minha confiança no cientista é de tal modo forte que seria irracional procurar outras fontes para checar a veracidade do conteúdo do seu artigo. Esse tipo de complicação é absolutamente relevante, mas tratá-la nos conduziria para além do escopo da presente dissertação.

“todas” as evidências disponíveis, mas nós não precisamos deste postulado. O exemplo acima mostra bem um caso comum de irracionalidade que poderia ser evitado se eu fosse uma pessoa mais prudente na aquisição de informações. Há vários exemplos como esse no cotidiano (a viabilidade desse requerimento também não precisa exigir que o agente tenha conhecimento da ciência do seu dia, como Brandt requer).

DePaul desenvolve o seu método de *balance and refinement*, uma versão modificada do ERA, para acomodar essa característica intuitiva do conceito de racionalidade. Segundo DePaul há basicamente dois modos pelos quais podemos vir a modificar nossa compreensão moral das coisas: (i) por meio do raciocínio lógico (*reasoning experience*) ou por (ii) experiências formativas (*formative experience*). O primeiro tipo é o que eu venho enfatizando ao longo desta dissertação: envolve construir argumentos a partir de proposições já aceitas, apreciar contraexemplos para argumentos propostos, considerar racionalmente diversas concepções morais e a força dos argumentos para cada uma delas etc. O segundo tipo é mais complicado. Ele não é, estritamente falando, conhecimento de todos os fatos relevantes. Envolve algo como experiências morais efetivas ao longo da vida, tais como realmente executar uma ação correta ou errada, conviver com uma pessoa que tem uma certa virtude ou vício, interagir com pessoas que sustentam valores distintos, assistir um filme, ler um romance, apreciar uma canção ou uma pintura com intensidade etc. Esse tipo de experiência da “vida real” (DePaul contrasta com a experiência filosófica argumentativa) não faz apenas com que a pessoa altere as suas crenças já sustentadas, mas faz com que ela “veja a situação de uma forma diferente”³⁴⁷. Aqui a ideia relevante para DePaul é a de conversão moral: uma mudança não contínua no meu sistema de crenças. Quando eu assisto um filme ou presencio um crime hediondo, eu posso vir a alterar todo o meu conjunto de crenças³⁴⁸. É uma alteração espontânea, que diferentemente do modo

³⁴⁷ DEPAUL, 1993, p.158-159.

³⁴⁸ Para DePaul o conservadorismo do ERA não lhe permite incorporar um fato psicológico da moralidade: a existência de conversões morais. Uma conversão moral não é uma revisão no sistema de crenças que busca eliminar as consistências internas, mas uma alteração descontínua em que a pessoa abandona todas as suas crenças prévias. Conforme DePaul, o ERA não é capaz de ser acomodar esse fenômeno porque ele é conservador: “a reflexão filosófica sobre moralidade, em alguns casos envolvendo não mais do que a leitura de visões morais de outros filósofos, algumas vezes pode levar a pessoa a abandonar os juízos morais ponderados antigos e comprometimentos filosóficos de fundo e a adotar um amplamente novo conjunto de convicções. Em contraste, no entendimento conservadorista, quando uma pessoa muda de opinião sobre uma crença, é sempre porque ela veio a perceber que há um conflito entre as suas crenças, e ela altera a sua opinião sobre algo a fim de continuar acreditando em algo mais. De acordo com essa concepção, para o trabalho de um filósofo moral influenciar uma pessoa, o filósofo tem de tocar em algo que a pessoa já acredita muito fortemente, e usar essa crença para forçar a revisão de outras crenças” (DEPAUL, 1993, p.42). O ERA de fato não permite conversões morais? No sentido que DePaul entende, certamente não. Em razão de postular que a revisibilidade é sempre uma questão de forçar juízos mais fixos contra outros juízos menos fixos, as mudanças no senso de justiça sempre terão sido motivadas por juízos prévios. É claro que é possível que após sucessivas aplicações do ERA alguém venha a adquirir um conjunto de convicções completamente distinto, mas essa revisão não será descontínua.

como o ERA caracteriza a deliberação moral, não é guiada por uma tentativa de maximizar a coerência do meu senso de justiça a partir da identificação de uma convicção firmemente sustentada que eu já possuía antes de presenciar a experiência formativa em questão.

Isso implica que eu posso ser irracional quando as minhas crenças consistem em um estado de equilíbrio entre meus (a) juízos morais ponderados, (b) princípios de justiça e (c) consideração das principais teorias morais da nossa tradição (ou simplesmente, consideração da posição original). Esses três elementos são insuficientes para a racionalidade ao menos que eu seja uma pessoa que já esteve em contato com experiências de vida pertinentes. O exemplo de DePaul é o de Jay: um rapaz que possui uma visão ingênua e romântica acerca da guerra. Todos os livros que ele leu, filmes que assistiu e amigos com que conversou, lhe contaram apenas sobre a bravura dos combatentes, a disputa pela honra e a busca pela glória na guerra. Jay jamais teve contato com o outro lado da guerra, ele jamais experienciou um indivíduo sendo brutalmente assassinado, um capitão traindo seus soldados e as motivações por trás dos líderes do conflito. O que aconteceria se Jay procedesse segundo o ERA? Nenhum mecanismo do ERA parece bloquear as crenças românticas de Jay. Os seus juízos são ponderados e ele considerou, como prescreve Rawls, todas as considerações filosóficas das mais diversas teorias morais, teorias essas que rejeitariam a sua visão sobre a guerra. Acontece que uma das convicções mais firmes de Jay não é expressa por meio das concepções-modelo, mas é, quem sabe, a crença de que o homem sempre deve buscar satisfazer as suas potencialidades, e a guerra é a melhor oportunidade para isso. Se Jay deve dar prioridade às suas crenças mais fixas, como postula o ERA, então o seu senso de justiça é aparentemente racional.

O ponto do exemplo de Jay é que a evidência que a pessoa não possui (no caso, experiências de vida) afeta a racionalidade das suas crenças³⁴⁹. O raciocínio de DePaul é que não é racional para uma pessoa fazer juízos morais ao menos que ela possa racionalmente acreditar que os seus juízos morais permaneceriam os mesmos face todas as experiências de vida enriquecedoras³⁵⁰. Considerando isso, DePaul acrescenta dois fatores ao ERA. Em primeiro lugar, (i) Jay deve considerar experiências de vida que ele poderia ter, e atividades em que ele poderia se engajar, bem como refletir com base nos argumentos pró e contra cada uma dessas experiências. Essa experiência de vida poderia ser adquirida por meio de atividades culturais como filmes, literatura, arte e música, além de vivências morais efetivas³⁵¹. O outro fator (ii) prescreve que Jay não deve apenas realizar essas atividades, mas realmente se engajar

³⁴⁹ DEPAUL, 1993, p.165.

³⁵⁰ DEPAUL, 1993, p.160-179.

³⁵¹ DEPAUL, 1993, p.173.

e se expor a elas. Se Jay efetivamente seguir esses dois passos, então ele estará em um estado de *balance and refinement*. *Balance* porque há coerência entre as suas crenças, e *refinement* porque as experiências adquiridas apurarão a sua própria capacidade de fazer juízos morais ponderados.

O que nós temos de nos perguntar é se o ERA é suficiente para os propósitos que Rawls designou para ele. Para repetir, o ERA é utilizado por Rawls para responder à seguinte questão: *qual é a concepção de justiça mais racional para pessoas como nós?*³⁵² Pessoas como nós nada mais são do que pessoas que sustentam as convicções morais que suportam as concepções-modelo da posição original. Nesse sentido, Jay não se inclui entre as preocupações de Rawls. Perceba que Brandt e DePaul querem responder a outra questão, inteiramente diferente: *que tipo de pessoa é moralmente racional? Ou que tipo de pessoa deveríamos ser?* Ao fazer perguntas como essas Brandt e DePaul não estão preocupados conosco (no sentido acima especificado), mas com qualquer pessoa capaz de racionalidade e moralidade. Por isso eles se obrigam a rejeitar qualquer conjunto de valores já aceitos para perguntar, de um ponto de vista presumivelmente neutro, que valores nós deveríamos possuir (Brandt) ou como deveríamos proceder na reflexão moral (DePaul) caso aspiremos pela racionalidade.

Há outra disanalogia fundamental. Rawls não procura saber o que é moralmente racional para pessoas como nós, mas o que é moralmente racional *em questões de justiça social*. As concepções-modelo constituem uma interpretação dos nossos comprometimentos morais mais profundos, mas uma interpretação que busca por elementos relevantes para formular uma concepção de justiça ou, se quisermos, uma interpretação do nosso raciocínio prático realizado quando questões de justiça estão em pauta. Assim, o que Rawls afirma é que, sobre questões de justiça, nada tem mais prioridade para nós do que as ideias expressas nas concepções-modelo. DePaul frequentemente critica esse postulado porque ele é “externalista e arbitrário”, mas nada impede que ao deliberar sobre outras questões nós não devemos considerar (e de fato não consideramos) outras crenças como mais básicas. Rawls não pretende cobrir todo o domínio da racionalidade ética, apenas uma porção sua e partindo de um objetivo restrito. Talvez isso signifique que não há algo como “um domínio da racionalidade ética”, mas diversos domínios, com diferentes critérios do que conta como evidência etc. Talvez uma pessoa seja irracional em

³⁵² Aqui eu sigo a caracterização de Williams do ERA: “É algumas vezes objetado contra o método do equilíbrio reflexivo que as intuições com as quais a teoria está sendo ajustada meramente representam nossas crenças éticas locais, e que elas podem não ser corretas. Quando o empreendimento é tomado do modo como eu tenho descrito, essa não é uma objeção relevante. As intuições são vistas como representante das nossas crenças éticas, porque a teoria sendo buscada é de uma vida ética *para nós*, e o ponto não é que as intuições deveriam ser em algum sentido último corretas, mas que elas deveriam ser nossas” (WILLIAMS, 1985, p.102).

um domínio (como as crenças de Jay sobre a guerra), mas racional em outro (caso Jay estivesse discutindo questões de justiça).

Entretanto, creio que é justo afirmar que o ERA é incapaz de bloquear atitudes reflexivas conservadoras e que ele inclusive pode promover o conservadorismo³⁵³. Este é um ponto, então, que DePaul e Brandt, ao menos aparentemente, acertam. Como Thomas Kelly e Sarah McGrath recentemente argumentaram, o ERA parece “fraco demais”: se alguém impecavelmente seguir o método, pode alcançar resultados que nós julgamos amplamente insatisfatórios. O conservadorismo do ERA, seguem os autores, se deve ao fato de que o seu *output* é fortemente dependente do seu *input*. Se o último for inadequado, como é o caso de Jay, o primeiro também inevitavelmente o será³⁵⁴.

As críticas que podem ser realizadas mais genericamente contra o ERA, críticas importantes e decisivas para uma avaliação dos méritos do método, são diferentes daquelas que poderiam ser direcionadas contra o uso que Rawls faz do método. Por exemplo, por que o conservadorismo seria um problema quando o objetivo é apresentar uma concepção de justiça mais racional para pessoas *como nós*? A própria formulação da questão parece excluir a necessidade de contar com o critério contrafactual que defina o racional em termos do que julgaríamos se confrontados com toda a evidência relevante, ou do que julgaríamos se fôssemos

³⁵³ Tampouco DePaul consegue evitar essa consequência na sua versão modificada do ERA. Ele opera uma distinção entre uma experiência enriquecedora e uma experiência corruptiva. Em linhas gerais, o primeiro tipo é uma experiência que o indivíduo poderia ver como aceitável vivenciar, e a segunda é um tipo que o indivíduo veria como “corruptível” do seu senso de justiça. Pense no exemplo de uma cristã muito devota que a acredita que o erotismo é um pecado imperdoável. Talvez essa cristã tenha esse juízo porque ela nunca frequentou ambientes eróticos. O método de DePaul determina que ela deve se submeter à experiência formativa em questão, quem sabe vindo a conhecer pessoas que frequentem tais ambientes. O problema é que se ela frequentar esses ambientes, ela terá razão em argumentar que se vivenciar tal situação, o seu senso de justiça será corrompido e ela jamais poderá, em virtude dessa corrupção, ver o erotismo como um pecado. DePaul argumenta que não é racional forçar as pessoas a vivenciar certas experiências que elas veem como corruptíveis. Por isso não é possível acrescentar, como cláusula de racionalidade, que a pessoa vivencie todas as experiências formativas possíveis (DEPAUL, 1993, p.160-163). Essa distinção de DePaul vem a custo relativamente alto: implica que o critério para definir o que é uma experiência corruptível é determinado pelas próprias crenças do agente. O agente pode, e DePaul não tem como impedir esse resultado, julgar que possui experiências enriquecedoras o bastante, ou que qualquer outra experiência seria corruptível. Joey poderia reflexivamente acreditar nisso. Talvez nós possamos saber que Joey não se submeteu a todas as experiências que podia e fez um mau juízo acerca delas, mas o que Joey poderia conseguir com o método de DePaul? DePaul explicitamente rejeita que outra pessoa que não eu (um teórico moral, por exemplo) possa determinar, para mim, que experiências eu devo adquirir e como devo pesá-las. Isso seria externalista, e o externalismo é irracional porque força revisões de crenças que eu não posso aceitar com razões que são minhas (DEPAUL, 1993, p.170). Se o externalismo é irracional, de outro lado o internalismo de DePaul parece não fornecer meios fortes o suficiente para bloquear o conservadorismo.

Essa não é uma conclusão que DePaul não tenha ele mesmo alcançado. Ele admite que o seu método não protege alguém da irracionalidade. Tudo o que ele faz é levar a um sistema de crenças que o agente considera verdadeiro, por passos que o agente pensa que deveria seguir (DEPAUL, 1993, p.178-179). Se é assim, nós poderíamos perguntar, o método não é demasiadamente modesto nas suas pretensões normativas e é tão refém do conservadorismo quanto o ERA? A vantagem do método de DePaul sobre o ERA, na ótica de DePaul, é que o seu método *possibilita* atitudes não conservadoras, enquanto que o ERA incorpora o conservadorismo como uma prescrição de racionalidade.

³⁵⁴ KELLY;MCGRATH, 2010.

pessoas “realmente racionais”. Talvez nós devêssemos ser outras pessoas. Talvez sejamos ingênuos (DePaul), preconceituosos (Brandt) ou loucos fanáticos (hipótese sugerida por Sosa³⁵⁵). Tudo isso é relevante, mas ainda parece fazer sentido perguntar qual a concepção de justiça mais racional para pessoas ingênuas, preconceituosas e fanáticas como nós. A resposta à questão sobre que concepção de justiça melhor reflete os nossos juízos morais não depende da resposta à questão sobre o que os nossos juízos morais devem refletir. Mesmo que essa primeira questão nos pareça restrita e menos importante do que outras questões que sentimos que são mais fundamentais e prementes em filosofia moral, e mesmo que ela pareça ser descritiva demais e normativa de menos, ainda é uma questão que em si mesma faz sentido.

A objeção do conservadorismo conduz naturalmente à objeção de que a justiça como equidade seria meramente uma descrição dos nossos comprometimentos morais mais profundos e que por isso ela seria insuficientemente normativa³⁵⁶. Essa é uma objeção conhecida e influente, mas dado o que foi identificado neste capítulo, abordá-la inteiramente está além da alçada desta dissertação. A explicação é muito simples: o argumento de Rawls é o de que a sua teoria é, além de pessoalmente justificável, publicamente justificável. Como ele não atribui à ideia de justificação pessoal todo o peso normativo da sua teoria e não concebe o ERA como exclusivamente responsável pela justificação pública, avaliar se a justiça como equidade é descritivista envolve tratar de questões que excedem o panorama do ERA e, por consequência, os objetivos desta dissertação.

No entanto, nós podemos deslocar o foco da objeção e voltá-lo para o ERA *per se*. Como visto, é provável que o ERA seja ineficiente para bloquear atitudes conservadoras. O crítico pode se apegar a essa ideia e objetar que nós temos de concluir que o *output* do método é insuficientemente normativo porque ele é condicionado pelo *input*, que é um elemento descritivo. Assim, embora o ERA descreva o que um determinado indivíduo deve acreditar, ele compreende esse “dever” como algo especificável em função daquilo que o indivíduo já acredita. Ebertz expressa com clareza esse tipo de objeção:

Embora possa ser um modo de determinar princípios com que nós (ou outros em um grupo específico) estaríamos dispostos a viver, ele não oferece uma

³⁵⁵ SOSA, 1993, p.265.

³⁵⁶ Como visto no primeiro capítulo, Goodman já sofreu essa crítica. No que respeita a Rawls, as primeiras objeções do descritivismo foram oferecidas ainda na década de setenta, por nomes como Hare (1989), Singer (2002), Lyons (1989) e Brandt (1998). Depois, essa crítica foi atualizada por Baier (1989), DePaul (1993), Ebertz (1993), Estlund (1998), Gaus (1996), Habermas (1995), Hampton (1980), Hedrick (2010), Kelly;McGrath (2010), Little (1984), O’Neill (1997), Raz (1990), dentre outros. O que une esses diversos autores, muitos com concepções filosóficas completamente distintas e mesmo irreconciliáveis entre si, é a crença de que a justiça como equidade é demasiadamente dependente de elementos descritivos (acordo em consenso sobreposto, concepção política de justiça, teoria moral como descrição de juízos morais ponderados).

explicação de por que nós (ou qualquer um) *deveria* viver desse modo. ‘O modo como nós estamos dispostos a viver’ e ‘o modo como nós deveríamos viver’ não estão conectados de uma maneira obviamente necessária³⁵⁷.

Eu penso que esse tipo de objeção remete diretamente à restrição internalista abordada no capítulo passado. Um dos argumentos contra aquela concepção, oferecido por McDowell, a chama de “psicologista” porque ela seria refém dos pontos de partida da deliberação³⁵⁸. Assim, *qualquer* princípio moral pode ser pessoalmente justificável. Note como a incisiva objeção avançada por Haslett expressa essa crítica:

Essa metodologia [o equilíbrio reflexivo], então, oferece aos seus utilizadores não apenas uma, mas duas camadas de proteção contra qualquer refutação possível. Ela pode ser usada igualmente bem em favor de virtualmente qualquer princípio. O preço pago para esse tipo de proteção, contudo, é muito alto: aquele que, em defesa de X, não admite nenhuma refutação possível, não fornece também nenhum suporte genuíno para X. Além do mais, protegendo os seus utilizadores de qualquer refutação possível, a metodologia do equilíbrio reflexivo é o veículo ideal para a promoção de retórica e preconceitos³⁵⁹

É claro que tudo depende do que Haslett quer dizer com “suporte genuíno”. No parágrafo seguinte ela esclarece que o objetivo de um “método de justificação” é selecionar o que, de uma perspectiva interpessoal, é o correto. Como o ERA é “psicologista”, ele não pode produzir acordo e, por isso, sequer pode ser considerado um método de justificação³⁶⁰. Evidentemente Haslett utiliza o conceito de justificação de um modo incrivelmente arbitrário, excluindo, sem argumentar, a viabilidade de uma noção pessoal ou subjetiva de justificação e de racionalidade. Apesar disso, a sua objeção de que o ERA fornece aos seus utilizadores uma proteção contra qualquer refutação possível parece sólida. Em favor de seu argumento, Haslett ainda poderia ter acrescentado experimentos empíricos que demonstram que o conhecimento que subjaz à habilidade de produzir juízos corretos é o mesmo conhecimento que subjaz à habilidade para reconhecer um juízo correto. Como consequência, agentes incompetentes em um domínio não terão a habilidade necessária para reconhecer os próprios erros. Essa falta de “habilidade metacognitiva” pode ter efeitos desastrosos: ela pode inclusive levar o agente a fazer uma autoavaliação errada das suas próprias competências, bloqueando o desenvolvimento e aperfeiçoamento dessas mesmas competências³⁶¹. Assim, indivíduos que realizam maus

³⁵⁷ EBERTZ, 1993, p.209.

³⁵⁸ MCDOWELL, 1995, p.68-96.

³⁵⁹ HASLETT, 1987, p.311.

³⁶⁰ HASLETT, 1987, p.311.

³⁶¹ Um desses estudos foi realizado por Kruger e Dunning (2009, p.30-33).

juízos morais, como Jay, não apenas tendem a carecer da habilidade para reconhecer os seus erros em ERA, mas, potencialmente, o ERA pode maximizar a confiança com que eles sustentam esses erros, pondo ainda mais obstáculos para o aperfeiçoamento moral desses indivíduos.

Como eu insistentemente ao longo deste capítulo busquei mostrar, o ERA não é designado para determinar o que é certo ou errado, ou o que é intersubjetivamente justificado. Faz pouco sentido acusá-lo de não oferecer algo que não é seu objetivo prover. Mas será que o conservadorismo e o descritivismo implicam a rejeição da própria ideia de justificação pessoal e do ERA? Foley, que defende uma concepção egocêntrica de racionalidade, sofreu críticas análogas às sofridas pelo ERA. Em sua defesa, Foley argumentou que juízos avaliativos de racionalidade devem ser entendidos com referência a um objetivo e a uma perspectiva. Para avaliarmos uma dada teoria da racionalidade, devemos ter claro qual a perspectiva com que essa teoria está concernida e se da perspectiva assumida ela consegue realizar os objetivos a que se propõe³⁶².

Para Foley qualquer concepção de racionalidade terá a seguinte forma: é racional para *S* sustentar *p* porque *S* tem os recursos *R* e porque da perspectiva *P* parece a *S* que, dado *R*, sustentar *p* é um modo efetivo de satisfazer o seu objetivo³⁶³. Essa fórmula é universal porque ela se aplica não apenas a critérios subjetivos para a avaliação da racionalidade de crenças, mas a critérios intersubjetivos e objetivos. Essa fórmula possui três elementos relevantes (um objetivo, uma perspectiva e um conjunto de recursos), e uma combinação particular desses elementos constitui um *ponto de vista* específico. Dependendo de como cada elemento é especificado e combinado, tem-se um ponto de vista epistêmico distinto a partir do qual avaliar a racionalidade de certas crenças ou ações³⁶⁴. Por exemplo, nós podemos adotar um ponto de vista sociocêntrico que consiste na perspectiva de uma comunidade, o conjunto de recursos (crenças, informações, evidências, valores, instituições etc.) acessível aos membros inteligentes dessa comunidade, e o objetivo de obter crenças precisas e abrangentes. Alternativamente, podemos adotar um ponto de vista egocêntrico que consiste na perspectiva de um sujeito particular dessa comunidade, o conjunto de recursos apenas acessível a ele, e o objetivo de obter crenças precisas e abrangentes. Para Foley, uma mesma crença pode ser socioentricamente irracional, mas egocentricamente racional (e vice-versa, de acordo com a sua posição)³⁶⁵. Isso

³⁶² FOLEY, 1993, p.31.

³⁶³ FOLEY, 1993, p.34.

³⁶⁴ FOLEY, 1993, p.14.

³⁶⁵ FOLEY, 1993, p.34.

sugere, argumenta Foley, que há diferentes pontos de vista legítimos a partir dos quais lidar com questões de racionalidade, e, consequentemente, diferentes teorias da racionalidade que não são necessariamente incompatíveis entre si.

No caso do ERA, o ponto de vista é subjetivo: se considera a perspectiva de *S*, dada por suas crenças, valores, desejos, disposições de avaliação, convicções morais e não-morais, padrões de reação emocional, comprometermos e planos de vida. Considera-se que *S* tem o objetivo de realizar uma ação correta e, por conseguinte, obter crenças morais corretas e possuir desejos que sejam compatíveis com as suas crenças sobre o que é certo. Os recursos de *S* são pessoais: abrangem os dados e informações representadas nas suas crenças correntes. A questão fundamental então é a seguinte: do ponto de vista das crenças mais estáveis e profundas de *S* e considerando as informações correntes aos quais *S* tem acesso, qual crença é a mais apropriada para *S* sustentar, dado o seu objetivo de realizar uma ação correta? Não é claro exatamente que uma estratégia descritivista é inadequada para responder *essa* questão³⁶⁶. E o mérito da resposta oferecida para essa questão não está, ao menos obviamente, implicado pela capacidade do ERA em evitar o conservadorismo. Que a ideia de justificação pessoal vinculada ao ERA pode autorizar como pessoalmente justificadas certas crenças claramente incorretas não significa que ela autoriza também como racionais ou corretas de acordo com outros critérios.

³⁶⁶ Williams estava bastante ciente de que, por exemplo, se um marido não conseguir por meio de uma rota deliberativa adequada a partir do seu *set* motivacional subjetivo (crenças, disposições etc.) chegar à conclusão de que tratar mal a sua esposa é errado, então ele não tem razão para tratar bem a sua esposa e, portanto, não pode ser acusado de irracionalidade por isso. É evidente que o que conta como (pessoalmente) justificado, no caso desse marido, depende decisivamente de um componente descritivo (nomeadamente, o tipo de crenças e desejos que o marido sustenta antes que inicie a refletir sobre as suas ações). Agora, e esse parece o ponto de Foley, dizer que não é irracional para o marido tratar mal a sua esposa, não significa dizer que é um comportamento correto ou socialmente aceitável. E tudo o que o Williams afirma, no tocante ao caso desse marido, é que nós não podemos dizer que ele tinha uma razão para tratar bem a sua esposa. Nós ainda podemos afirmar que ele é “íngrato, insensível, grosso, desagradável, cruel, egoísta, sexista e muitas outras coisas desfavoráveis” (WILLIAMS, 1995, p.39)

Considerações finais

As interpretações particulares do ERA que foram apresentadas na literatura podem ser classificadas em duas tendências gerais³⁶⁷. A primeira tendência, liderada por Daniels e os críticos iniciais de Rawls, como Brandt, Hare, Singer, Lyons e outros, considerava que com o ERA a filosofia moral agora dispunha de uma tentativa de sustentação da objetividade moral e do conhecimento moral. Em especial, um tipo de sustentação coerentista e não-fundacionalista que não contaria com pressupostos epistêmicos e metafísicos indesejados. Mais recentemente, especialmente a partir da década de noventa, essa leitura foi desafiada em várias frentes. Como visto ao longo do capítulo terceiro, o seu coerentismo foi questionado por (dentre outros) autores como Ebertz, Audi e Thomas, e o seu objetivismo foi rejeitado por (dentre outros) filósofos como Sosa, Foley, DePaul e Parfit. Essa segunda tendência interpretativa, contudo, se ressentia de um maior desenvolvimento e detalhamento, principalmente se comparada à leitura sistemática apresentada por Daniels. Com esta dissertação busquei oferecer elementos para enriquecer uma formulação fundacionalista e subjetivista do ERA. Além disso, argumentei que uma compreensão subjetivista e fundacionalista do ERA se ajusta melhor à posição geral de Rawls que articulações alternativas. Estarei satisfeito, contudo, se o leitor puder ter concordado que Rawls poderia aceitar essa versão sem precisar abrir mão de nenhum aspecto fundamental da sua posição.

Embora divirjam drasticamente, as duas tendências concordam a respeito de algumas características básicas do ERA. Ambas definem o ERA como uma proposta em epistemologia que se opõe às formas clássicas de fundacionalismo. Isso significa que o ERA rejeita (i) a possibilidade de haver uma crença cuja justificação é infalível: qualquer crença moral pode ser descartada e revisada se se mostrar incoerente com um conjunto coerente de crenças. Por implicação, ambas as tendências concordam que (ii) a justificação de uma crença depende de uma relação de ajuste mútuo entre a crença e um conjunto de crenças. Há também unanimidade em relação a uma terceira característica: (iii) a ideia de que questões normativas que estipulam o que os indivíduos devem fazer e acreditar devem ser resolvidas remetendo-se a um elemento descritivo, a saber, o conjunto de crenças que as pessoas de fato já aceitam como válido. Essas três características são constitutivas do que é o equilíbrio reflexivo. Se ao menos uma delas não

³⁶⁷ A interpretação naturalista defendida por Mikhail rejeita essas duas tendências. Conforme argumentado no capítulo segundo, não creio que seus argumentos sejam suficientemente convincentes.

for contemplada por uma formulação particular do método, então não se tratará de uma versão do equilíbrio reflexivo, mas de algum outro método.

As divergências começam quando as duas interpretações tentam explicar o que significam (i), (ii) e (iii), precisamente explicar em que sentido o termo “justificação” é empregado e qual o papel do elemento descritivo. Brandt considerou o equilíbrio reflexivo essencialmente errado porque o seu proponente não ofereceria nenhum argumento satisfatório para a sua tese de que crenças justificadas em ERA são crenças epistemologicamente credíveis (isto é, crenças mais próximas da verdade), e não meramente um conjunto de preconceitos internamente coerente e sustentado com comprometimento. Ao responder a Brandt, Daniels argumentou que o processo de busca por coerência, mediado pela ideia de extensa revisibilidade dos juízos morais ponderados, geraria o tipo de justificação requerido. O emprego do ERA por indivíduos particulares causaria convergência em um certo número de crenças, e essa convergência constituiria uma evidência para a verdade moral. Assim, um indivíduo que sustenta crenças justificadas em ERA estaria mais próximo da verdade moral e, conseqüentemente, do conhecimento moral, que um indivíduo que não sustenta crenças justificadas em ERA.

Nesta dissertação não analisei em detalhes como Daniels concebe a sua defesa da objetividade do ERA, nem o mérito da sua tentativa coerentista de resposta a Brandt. Ao invés, argumentei que é viável e desejável construir outra imagem do ERA. Analisando a definição de justificação oferecida por Rawls, em que a justificação é definida como um argumento que endereçamos àqueles que discordam de nós, quando estamos tentando resolver um desacordo, *ou* um argumento que cada um endereça a si mesmo, quando está indeciso, argumentei que essa definição deveria ser isolada em duas partes. A primeira diria respeito à ideia de justificação pública, e a segunda remeteria a uma noção de justificação pessoal. Paralelamente a essa divisão, busquei defender a importância de se associar cada tipo de justificação a uma versão específica do equilíbrio reflexivo: a pública às versões geral e plena, e a pessoal à ampla. Esse é, a meu ver, o ponto crucial desta dissertação. Independentemente dos méritos da noção de justificação pessoal aqui delineada, há algo de muito errado com a leitura padrão do ERA.

O ERA funciona como um processo de revisão de crenças em que o indivíduo, na busca por equilíbrio, toma as suas crenças sustentadas com mais comprometimento e confiança (seus pontos fixos provisórios) como ideias reguladoras a partir das quais os ajustes no interior do seu sistema de crenças devem ser feitos. O ERA serve como um meio de autoclarificação pessoal que informa ao indivíduo o que ele realmente deve acreditar, dado o que ele demais acredita com extrema confiança. Como o resultado do ERA depende das crenças iniciais que o

indivíduo tem antes de iniciar a refletir, o elemento normativo “deve” é estipulado a partir de um elemento descritivo. Esse é o sentido do descritivismo do ERA. O que um indivíduo deve fazer ou acreditar, é aquilo que ele faria ou acreditaria em ERA após refletir a partir das suas crenças iniciais.

Assim compreendido o ERA deixa de ser vulnerável a objeções semelhantes a de Brandt. Se a sua função não é produzir crenças confiáveis de um ponto de vista outro que não o pessoal de primeira pessoa, qual o sentido de acusá-lo de equacionar credibilidade e comprometimento? E se o objetivo do ERA é maximizar a confiança com que um indivíduo pode sustentar o *seu* sistema de crenças, como ele não poderia ser descritivista? Os problemas que levaram muitas pessoas a rejeitar o ERA parecem se dissolver.

Contundo, é provável que essas pessoas permaneçam céticas, embora que por outra razão. Uma reação natural à versão do ERA proposta nesta dissertação pode objetar o seguinte: se as crenças mais básicas de um indivíduo são preconceitos profundamente arraigados, então a aplicação do ERA maximizará esses preconceitos e os autorizará como pessoalmente justificados. O crítico pode inclusive levantar o caso do nazista convicto para argumentar que o ERA não bloquearia as suas crenças morais antissemitas. O ERA seria, segue a objeção, insuficientemente normativo ou crítico mesmo para o domínio de justificação pessoal.

Se o nazismo era pessoalmente justificado para o nazista típico é controverso e requer análise, mas é provável que para muitos nazistas convictos ele de fato era. No que concerne a essa objeção, foi argumentado que faz tanto sentido perguntar qual a concepção de justiça mais racional para pessoas preconceituosas como nós, quanto faz perguntar se nós deveríamos ou não ser o tipo de pessoas que agora somos. Para o objetivo a que Rawls se propõe com o ERA, a saber, especificar que crenças, valores e princípios de justiça um cidadão razoável que vive em um regime democrático contemporâneo poderia racionalmente aceitar sem precisar descartar suas crenças morais, religiosas e filosóficas abrangentes, não é claro que a objeção do nazista convicto precise ser respondida. Por definição, um cidadão razoável de um regime democrático não é um nazista e respeita valores básicos como reciprocidade e civilidade, além de possuir uma disposição para encontrar termos equitativos de cooperação com os seus pares.

Contudo, essa objeção pode ser direcionada não para o uso que Rawls faz do ERA, mas para o ERA em si. Qualquer noção de justificação e racionalidade é normativa, pois não se trata de uma descrição daquilo que as pessoas acreditam, mas de estipular critérios de correção que permitam avaliar as crenças correntes de uma pessoa e assim determinar o que elas devem acreditar em um dado domínio. Se o critério de justificação pessoal for incapaz de corrigir as crenças de um nazista convicto, qual a sua função?

Para lidar com essa objeção acredito que o proponente do ERA deve se manter na defensiva: ele deve aceitar que o ERA é insuficientemente fraco para bloquear crenças morais deficientes e claramente falsas. Após se manter nessa posição, o proponente do ERA deve torná-lo mais atrativo argumentando em favor da relevância da pergunta que ele se propõe a responder, mais especificamente, em favor da importância da noção de justificação pessoal. O proponente do ERA pode adotar pelo menos duas estratégias para alcançar esse objetivo. Mantendo-se fiel à tese de que nenhuma crença será racional se alguém que emprega impecavelmente o ERA não puder alcançar uma conclusão para ela a partir do seu conjunto de crenças efetivo, ele pode argumentar que essa ideia de justificação pessoal estabelece uma condição de adequação para outros conceitos de justificação, como o conceito de justificação pública. Conforme pode ser identificado nos breves comentários sobre a ideia de justificação pública presentes ao longo desta dissertação, o meu ponto de vista é o de que essa é a posição de Rawls. Para ele, nenhuma teoria pode ser publicamente justificada, e assim gerar genuínas obrigações políticas, se ela não for pessoalmente justificada para cada cidadão. Essa seria a defesa de um papel forte para a noção de justificação pessoal.

Alternativamente, conforme visto nas últimas páginas do terceiro capítulo, o proponente do ERA pode argumentar que, apesar de a sua concepção de normatividade ser claramente insuficiente para muitos propósitos, como o caso do nazista convicto mostra, ela cumpre um objetivo que é em si mesmo relevante: ela propicia ao indivíduo um sistema de crenças que ele pode sustentar com comprometimento e confiança, um sistema com o qual ele pode viver uma vida com mais autoconfiança. Crenças pessoalmente justificadas são meramente crenças que o indivíduo pode sustentar de forma coerente, segura e sincera. A justificabilidade pessoal de uma crença nada diz a respeito da sua moralidade ou da sua correção. De fato, pode argumentar o defensor do ERA, que as crenças nazistas são pessoalmente justificadas para o nazista é a prova de que há algo de moralmente errado no seu sistema de crenças. Essa estratégia, ou algo semelhante à ela, é adotada por Foley para defender a pertinência da sua noção de racionalidade egocêntrica. Essa seria a defesa de um papel mais modesto e fraco para a noção de justificação pessoal.

Se adota a primeira estratégia, o proponente do ERA automaticamente aceita um relativismo de razões e um consequente enfraquecimento da noção de obrigação moral. Significa que um nazista convicto não tem uma *obrigação moral* para deixar de ser nazista. Embora haja claramente algo de deficiente nas suas crenças, nós não teríamos nenhuma base para dizer que ele *deve* deixar de sustentá-las. Talvez o proponente do ERA simplesmente aceite esse relativismo e recuse esses conceitos morais *thin* tradicionais (dever, obrigação, correto,

incorreto, bom, mau), defendendo então que a filosofia moral deve utilizar um vocabulário avaliativo *thick* (que use virtudes como padrões de avaliação). Porém, essa estratégia tem outra consequência: se o que é pessoalmente justificado para um indivíduo depende das razões que esse indivíduo tem, e essas razões são dadas pelos seus comprometerimentos mais profundos e estáveis, que razões ele tem para ter *esses* comprometerimentos? Que razões um cidadão de um regime democrático possui para ser *esse* cidadão? Se ele não tem nenhuma, parece que ele também não tem razões para acreditar em coisa alguma, já que todas as suas razões dependeriam de crenças que, ultimamente, ele não tem razão para sustentar.

A segunda estratégia evita esse problema porque ela é mais modesta: tudo o que ela afirma é que crenças pessoalmente justificadas são crenças que um indivíduo pode aceitar com confiança de que não está traindo as suas convicções mais fundamentais. A segunda estratégia aceita que existem outros conceitos de justificação e racionalidade e não os subordina ao conceito de justificação pessoal. Assim, mesmo se a crença nazista é pessoalmente justificada para um nazista convicto, ela é uma crença falsa ou intersubjetivamente inválida, e esse fato fornece ao nazista uma razão com base na qual ele *deve* deixar de sustentá-la, mesmo que ela seja pessoalmente justificada para ele. A dificuldade com essa estratégia é que ela parece somente remeter a uma ideia de conforto psicológico cuja conexão com uma noção de justificação, essencialmente uma noção normativa, é confusa e incerta. Não é claro que a noção de justificação pessoal realmente descreveria um conceito relevante de justificação.

Referências

- ALEXANDER, Sidney. "Social Evaluation through Notional Choice". *The Quarterly Journal of Economics*, 88 (4), 1974, p.597-624.
- AUDARD, Catherine. *John Rawls*. Toronto: McGill-Queen's University Press, 2007.
- AUDI, Robert. *A Contemporary Introduction to the Theory of Knowledge*. Third Edition. New York: Routledge, 2011.
- _____. *The Architecture of Reason: The Structure and Substance of Rationality*. Oxford: Oxford University Press, 2001.
- _____. *The Good in the Right: A Theory of Intuition and Intrinsic Value*. Princeton: Princeton University Press, 2004.
- _____. *The Structure of Justification*. New York: Cambridge University Press, 1993.
- _____. *Moral Knowledge and Ethical Character*. Oxford: Oxford University Press, 1997.
- BAIER, Kurt. "Justice and the Aims of Political Philosophy". *Ethics*, 99 (4), 1989, p.771-790.
- BAYNES, Kenneth. "Constructivism and Practical Reason in Rawls". *Analyse & Kritik*, 14, 1992, p.18-32.
- BIRD-POLLAN, Stefan. "Rawls: Construction and Justification". *Public Reason*, 1 (2), 2009, p.12-30.
- BLACHOWICZ, James. "Reciprocal Justification in Science and Moral Theory". *Synthese*, 110 (3), 1997, p.447-468.
- BONEVAC, Daniel. "Reflection without Equilibrium". *The Journal of Philosophy*, 101 (7), 2004, p.363-388.
- BONJOUR, Laurence. *The Structure of Empirical Knowledge*. Cambridge: Harvard University Press, 1985.
- BONJOUR, Laurence; SOSA, Ernest. *Epistemic Justification: Internalism vs Externalism, Foundations vs Virtues*. Oxford: Blackwell Publishing, 2003.
- BRANDT, Richard. *A Theory of the Good and the Right*. New York: Prometheus Books, 1998.
- _____. "The Science of Man and Wide Reflective Equilibrium". *Ethics*, 100 (2), 1990, p.259-278.
- BRINK, David. *Moral Realism and the Foundations of Ethics*. New York: Cambridge University Press, 1989.
- BUCHANAN, Allen. "Revisability and Rational Choice". *Canadian Journal of Philosophy*, 5 (3), 1975, p.395-408.
- CHRISTENSEN, David. "Conservatism in Epistemology". *Noûs*, 28 (1), 1994, p.69-89.
- CLARKE, Stanley. "Anti-Theory in Ethics". *American Philosophical Quarterly*, 24 (3), 1987, p.237-244.
- COHEN, Jonathan. "Can human irrationality be experimentally demonstrated?". *The Behavioral and Brain Sciences*, 4, 1981, p. 317-331.
- DANIELS, Norman (Org.). *Reading Rawls*. Stanford: Stanford University Press, 1989.
- _____. *Justice and Justification: Reflective Equilibrium in Theory and Practice*. Cambridge: Cambridge University Press, 1996.

- DANIELS, Norman. "Reflective Equilibrium". *The Stanford Encyclopedia of Philosophy* (Spring 2011 Edition). Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/spr2011/entries/reflective-equilibrium/>>.
- DARWALL, Stephen. "A Defense of the Kantian Interpretation". *Ethics*, 86 (2), 1976, p.164-170.
- DEPAUL, Michael. *Balance and Refinement: Beyond Coherence Methods of Moral Inquiry*. London: Routledge, 1993.
- _____. "Reflective Equilibrium and Foundationalism". *American Philosophical Quarterly*, 23 (1), 1986, p.59-69.
- DUNNING, David; KRUGER, Justin. "Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments". *Psychology*, (1), 2009, p.30-46.
- DWORKIN, Ronald. "The Original Position". In: DANIELS, Norman (org.). *Reading Rawls*. Stanford: Stanford University Press, 1989, p.16-52.
- EBERTZ, Roger. "Is Reflective Equilibrium a Coherentist Model?". *Canadian Journal of Philosophy*, 23 (2), 1993, p.193-214.
- ESTLUND, David. "The Insularity of the Reasonable: Why Political Liberalism Must Admit the Truth". *Ethics*, 108 (2), 1998, p.252-275.
- FOLEY, Richard. "Epistemic Conservatism". *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 43 (2), 1983, p.165-182.
- _____. *Working Without a Net: A Study of Egocentric Epistemology*. Oxford, Oxford University Press, 1993.
- _____. *Intellectual Trust in Oneself and Others*. New York: Cambridge University Press, 2004.
- _____. *A Theory of Epistemic Rationality*. Cambridge, Harvard University Press, 1987.
- FREEMAN, Samuel. *Rawls*. London: Routledge, 2007.
- _____. (Ed.). *The Cambridge Companion to Rawls*. Cambridge: Cambridge University Press, 2003.
- GAUS, Gerald. *Justificatory Liberalism: An Essay on Epistemology and Political Theory*. Oxford: Oxford University Press, 1996.
- GOLDMAN, Alvin. "Epistemic Folkways and Scientific Epistemology". In: KORNBLITH, Hilary (Org.). *Naturalizing Epistemology*. Second Edition. Cambridge: MIT Press, 1994, p.291-315.
- _____. "What is Justified Belief?". In: KORNBLITH, Hilary (Org.). *Naturalizing Epistemology*. Second Edition. Cambridge: MIT Press, 1994, p.105-130.
- _____. "Argumentation and Interpersonal Justification". *Argumentation*, 11, 1997, p.155-164.
- GOODMAN, Nelson. *Fact, Fiction and Forecast*. Cambridge: Harvard University Press, 1983.
- _____. "Sense and Certainty". *The Philosophical Review*, 61 (2), 1952, p.160-167.
- HABERMAS, Jurgen. "Reconciliation through the public use of reason: Remarks on John Rawls' Political Liberalism". *The Journal of Philosophy*, 92(3), 1995, p.109-131.

- HAMPTON, Jean. "Contracts and Choices: Does Rawls Have a Social Contract Theory?". *The Journal of Philosophy*, 77 (6), 1980, p.315-338.
- HARE, Richard. "Rawls' Theory of Justice". In: DANIELS, Norman (org.). *Reading Rawls*. Stanford: Stanford University Press, 1989, p. 81-107.
- HARMAN, Gilbert. "Positive versus Negative Undermining in Belief Revision". In: KORNBLITH, Hilary (Org.). *Naturalizing Epistemology*. Second Edition. Cambridge: MIT Press, 1994, p.317-336.
- HARMAN, Gilbert. "Three Trends in Moral and Political Philosophy". *The Journal of Value Inquiry*, 37 (3), 2003. p.415-425.
- HARMAN, Gilbert; KULKARNI, Sanjeev. "The Problem of Induction". *Philosophy and Phenomenological Research*, 52 (3), 2005, p.559-75.
- HASLETT, D.W. "What is Wrong with Reflective Equilibria?". *The Philosophical Quarterly*, 37 (148), 1987, p.305-311.
- HEDRICK, Todd. *Rawls and Habermas: Reason, Pluralism and the Claims of Political Philosophy*. Stanford: Stanford University Press, 2010.
- HOLMGREN, Margaret. "The Wide and Narrow of Reflective Equilibrium". *Canadian Journal of Philosophy*, 19 (1), 1989, p.43-60.
- KAHANE, Howard. "Nelson Goodman's Entrenchment Theory". *Philosophy of Science*, 32 (3/4), 1965, p.377-383.
- KAPPEL, K. "The Meta-Justification of Reflective Equilibrium". *Ethical Theory and Moral Practice*, 9 (2), 2006, p.131-147.
- KELLY, Thomas; MCGRATH, Sarah. "Is Reflective Equilibrium Enough?". *Philosophical Perspectives*, 24, 2010, p.325-359.
- KORNBLITH, Hilary (Org.). *Naturalizing Epistemology*. Second Edition. Cambridge: MIT Press, 1994.
- KVANVIG, Jonathan. "Conservatism and Its Virtues". *Synthese*, 79 (1), 1989, p.143-163.
- LADEN, Anthony Simon. "The House That Jack Built: Thirty Years of Reading Rawls". *Ethics*, 113 (2), 2003, p.367-390.
- LIPKIN, Robert Just. "Beyond Skepticism, Foundationalism and the New Fuzziness: The Role of Wide Reflective Equilibrium in Legal Theory". *Cornell Law Review*, 75, 1990, p.811-877.
- LITTLE, Daniel. "Reflective Equilibrium and Justification". *The Southern Journal of Philosophy*, 22, 1984, p.373-387.
- LYONS, David. "Nature and Soundness of the Contract and Coherence Arguments". In: DANIELS, Norman (org.). *Reading Rawls*. Stanford: Stanford University Press, 1989, p. 141-159.
- MCDOWELL, John. "Might There Be External Reasons?". In: ALTHAM, J.E.J; HARRISON, Ross (Eds.). *World, Mind and Ethics: Essays on the Ethical Philosophy of Bernard Williams*. Cambridge: Cambridge University Press, 1995, p.68-85.
- MIKHAIL, John. *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment*. Cambridge: Cambridge University Press, 2011.
- NEAL, Patrick. "Does He Mean What He Says? (Mis)Understanding Rawls' Practical Turn". *Polity*, 27 (1), 1994, p.77-111.
- NIELSEN, Kai. "Considered Judgments Again". *Human Studies*, 5 (2), 1982, p.109-118.

- _____. "Relativism and Wide Reflective Equilibrium". *Monist*, 76 (3), 1993, p.316-332.
- NISBETT, Richard; ROSS, Lee. "Judgmental Heuristics and Knowledge Structures". In: KORNBLITH, Hilary (Org.). *Naturalizing Epistemology*. Second Edition. Cambridge: MIT Press, 1994, p.261-290.
- O'NEILL, Onora. "Political Liberalism and Public Reason: A Critical Notice of John Rawls, Political Liberalism". *The Philosophical Review*, 106 (3), 1997, p.411-428.
- _____. "Constructivism in Ethics". *Proceedings of the Aristotelian Society*, New Series, 89, 1988-1989, p.1-17.
- _____. "Constructivism in Rawls and Kant". In: FREEMAN, Samuel (ed.) *The Cambridge Companion to Rawls*. Cambridge: Cambridge University Press, 2003, p.347-367.
- PARFIT, Derek. *On What Matters*. Volume One and Two. Oxford: Oxford University Press, 2011.
- POLLOCK, John. "A Plethora of Epistemological Theories". In: PAPPAS, George (Org.). *Justification and Knowledge*. Dordrecht: Reidel Publish Company, 1979, p.93-113.
- PUTNAM, Hillary. "Foreword to the Fourth Edition". In: GOODMAN, Nelson. *Fact, Fiction and Forecast*. Cambridge: Harvard University Press, 1983, p.vii-xvi.
- RAWLS, John. "A Kantian Conception of Equality (1975)". In: RAWLS, John. *Collected Papers*. Edited by Samuel Freeman. Cambridge: Harvard University Press, 1999, p. 254-266.
- _____. *A Theory of Justice*. Original Edition. Cambridge: Harvard University Press, 1971.
- _____. *A Theory of Justice*. Revised Edition. Cambridge Harvard University Press, 1999.
- _____. "A Study in the Grounds of Ethical Knowledge: Considered with Reference to Judgments on the Moral Worth of Character". *P.h.D Dissertation*, Princeton University, 1950.
- _____. *Collected Papers*. Edited by Samuel Freeman. Cambridge: Harvard University Press, 1999.
- _____. "Fairness to Goodness (1975)". In: RAWLS, John. *Collected Papers*. Edited by Samuel Freeman. Cambridge: Harvard University Press, 1999, p. 267-285.
- _____. *Justice as Fairness: A Restatement*. Edited by Erin Kelly. Cambridge: Harvard University Press, 2001.
- _____. "Justice as Fairness: Political, not Metaphysical (1985)". In: RAWLS, John. *Collected Papers*. Edited by Samuel Freeman. Cambridge: Harvard University Press, 1999, p. 388-414.
- _____. *Political Liberalism*. Expanded Edition. New York: Columbia University Press, 2005.
- _____. "Reply to Alexander and Musgrave (1974)". In: RAWLS, John. *Collected Papers*. Edited by Samuel Freeman. Cambridge: Harvard University Press, 1999, p. 232-253.
- _____. "Kantian Constructivism in Moral Theory (1980)". In: RAWLS, John. *Collected Papers*. Edited by Samuel Freeman. Cambridge: Harvard University Press, 1999, p. 303-358.

- _____. "The Domain of Political and Overlapping Consensus (1989)". In: RAWLS, John. *Collected Papers*. Edited by Samuel Freeman. Cambridge: Harvard University Press, 1999, p. 473-496.
- _____. "The Independence of Moral Theory (1975)". In: RAWLS, John. *Collected Papers*. Edited by Samuel Freeman. Cambridge: Harvard University Press, 1999, p. 286-302.
- _____. *The Law of Peoples*. Cambridge: Harvard University Press, 1999.
- _____. "The Idea of Public Reason Revisited (1997)". In: RAWLS, John. *Political Liberalism*. Expanded Edition. New York: Columbia University Press, 2005, p.440-490.
- _____. "Outline of a Decision Procedure for Ethics (1951)". In: RAWLS, John. *Collected Papers*. Edited by Samuel Freeman. Cambridge: Harvard University Press, 1999, p. 1-19.
- RAZ, Joseph. "Facing Diversity: The Case of Epistemic Abstinence". *Philosophy & Public Affairs*, 19 (1), 1990, p.3-46.
- _____. "The Claims of Reflective Equilibrium". *Inquiry*, 25 (3), 1982, 307-330.
- RESNIK, Michael. "Logic: Normative or Descriptive? The Ethics of Belief or a Branch of Psychology?". *Philosophy of Science*, 52 (2), 1985, p.221-238.
- ROBERTS, Peri. *Political Constructivism*. London: Routledge, 2007.
- RORTY, Richard. *Objectivity, Realism and Truth*. Cambridge: Cambridge University Press, 1991.
- SAYRE-MCCORD, Geoffrey. "Coherentist Epistemology and Moral Theory". In: SINNOTT-ARMSTRONG, Walter; TIMMONS, Mark (orgs.). *Moral Knowledge*. Oxford: Oxford University Press, 1996, p.137-189.
- SCANLON, Thomas. "Rawls on Justification". In: FREEMAN, Samuel (Ed.). *The Cambridge Companion to Rawls*. Cambridge: Cambridge University Press, 2003, p.139-168.
- _____. "The Aim and Authority of Moral Theory". In: *Oxford Journal of Legal Studies*, 12 (1), 1992, p.1-23.
- _____. *What We Owe to Each Other*. Cambridge: Harvard University Press, 1998.
- SCHEFFLER, Israel. "Inductive Inference: A New Approach". *Science*, New Series, 127 (3291), 1958, p.177-181.
- _____. "On Justification and Commitment". *The Journal of Philosophy*, 51 (6), 1954, p.180-190.
- SCHWARTZ, Adina. "Moral Neutrality and Primary Goods". *Ethics*, 83 (4), 1973, p.294-307.
- SENCERZ, Stefan. "Moral Intuitions and Justification in Ethics". *Philosophical Studies*, 50, 1986, p.77-95.
- SHAW, William. "Intuition and Moral Philosophy". *American Philosophical Quarterly*, 17 (2), 1980, p.127-134.
- SIEGEL, Harvey. "Justification by Balance". *Philosophy and Phenomenological Research*, 52 (1), 1992, p.27-46.
- SILVEIRA, Denis Coitinho. "Construtivismo Político: Além do Realismo e do Antirrealismo". *Ethic@*, 10 (1), 2011, p.1-25.
- SINGER, Marcus. "Moral Theory and Justification". *Philosophy*, 62 (242), 1987, p.517-522.
- SINGER, Peter. "Sidgwick and Reflective Equilibrium". In: SINGER, Peter. *Unsanctifying Human Life*. Oxford: Blackwell, 2002, p.27-50.

- SINNOT-ARMSTRONG, Walter. *Moral Skepticisms*. Oxford: Oxford University Press, 2006.
- SKLAR, Lawrence. "Methodological Conservatism". *The Philosophical Review*, 84(3), 1975, p.374-400.
- SOSA, Ernest. *Knowledge in Perspective: Selected Essays in Epistemology*. Cambridge: Cambridge University Press, 1991.
- STICH, Stephen. "Reflective Equilibrium, Analytic Epistemology and the Problem of Cognitive Diversity". *Synthese*, 74, 1988, p.391-413.
- _____. "Could Man Be an Irrational Animal? Some Notes on the Epistemology of Rationality". In: KORNBLITH, Hilary (Org.). *Naturalizing Epistemology*. Second Edition. Cambridge: MIT Press, 1994, p.337-357.
- STICH, Stephen; NISBETT, Richard. "Justification and the Psychology of Human Reasoning". *Philosophy of Science*, 47 (2), 1980, p.188-202.
- STRONG, Carson. "Theoretical and Practical Problems with Wide Reflective Equilibrium in Bioethics". *Theor Med Bioeth*, 31, 2010, p.123-140.
- TELLER, Paul. "Goodman's Theory of Projection". *The British Journal for the Philosophy of Science*, 20 (3), 1969, p.219-238.
- TERSMAN, Folke. *Reflective Equilibrium An Essay in Moral Epistemology*. Stockholm: Almqvist & Wiksell International, 1993.
- THAGARD, Paul. "From the Descriptive to the Normative in Psychology and Logic". *Philosophy of Science*, 49 (1), 1982, p.24-42.
- THOMAS, Alan. *Value and Context: The Nature of Moral and Political Knowledge*. Oxford: Oxford University Press, 2006.
- TIMMONS, Mark. *Morality without Foundations: A Defense of Ethical Contextualism*. Oxford: Oxford University Press, 1999.
- _____. "Foundationalism and the Structure of Ethical Justification". *Ethics*, 97 (3), 1987, p.595-609.
- TVERSKY, Amos; KAHNEMAN, Daniel. "Judgment under Uncertainty: Heuristics and Biases". *Science*, New Series, 185 (4157), 1974, p.1124-1131.
- VAHID, Hamid. "Varieties of Epistemic Conservatism". *Synthese*, 141 (1), 2004, p.97-122.
- VAN DER BURG, Wibren; VAN WILLIGENBURG, Theo (Eds.). *Reflective Equilibrium: Essays in Honour of Robert Heeger*. Dordrecht: Kluwer Academic Publishers, 1998.
- WENAR, Leif. "Political Liberalism: An Internal Critique". *Ethics*, 106 (1), 1995, p.32-62.
- WILLIAMS, Bernard. *Ethics and Limits of Philosophy*. Cambridge: Harvard University Press, 1985.
- _____. *Moral Luck*. Cambridge: Cambridge University Press, 1981.
- _____. *Making Sense of Humanity and other Philosophical Papers*. Cambridge: Cambridge University Press, 1995.