

Revista de Pesquisa em Filosofia

FUNDAMENTO

Ouro Preto - Minas Gerais - Brasil



UFOP

**Universidade Federal
de Ouro Preto**

Fundamento Revista de Pesquisa em Filosofia / Universidade Federal de Ouro Preto / Grupo de Pesquisa em Filosofia Contemporânea. n. 8, (jan–jun. 2014) – Ouro Preto:

Ed. UFOP, 2014

Semestral – Tiragem: 200 exemplares.

ISSN: 2177–6563

1. Filosofia - Periódicos. I. Universidade Federal de Ouro Preto. II. Instituto de Filosofia, Artes e Cultura. III. Grupo de Pesquisa em Filosofia Contemporânea.

Endereço para correspondência:

Revista Fundamento
Instituto de Filosofia, Artes e Cultura – IFAC
Rua Coronel Alves, 55, Centro, Ouro Preto, MG, Brasil
CEP: 35400-000

Endereço eletrônico: revistafundamento@ufop.br

Revista Fundamento

UNIVERSIDADE FEDERAL DE OURO PRETO

Reitor

Prof. Dr. Marcone Jamilson Freitas Souza

INSTITUTO DE FILOSOFIA, ARTES E CULTURA/IFAC

Diretor

Prof. Dr. Guilherme Paoliello

DEPARTAMENTO DE FILOSOFIA/DEFIL

Chefe de Departamento

Prof. Dr. Olímpio José Pimenta

Bibliotecária

Neide Nativa

FUNDAMENTO

REVISTA DE PESQUISA EM FILOSOFIA

Diretores Gerais

Prof. Dr. Mário Nogueira (UFOP)

Prof. Dr. Sérgio Miranda (UFOP)

Diretores Executivos

Aline Monteiro (UFOP)

Editor

Prof. Dr. Mário Nogueira (UFOP)

Prof. Dr. Sérgio Miranda (UFOP)

Diagramação

Aline Monteiro (UFOP)

Assistente de Revisão

Aline Monteiro (UFOP)

Comitê Editorial

Prof. Dr. Alexandre Noronha Machado (UFPR)

Prof. Dr. Hélio Lopes da Silva (UFOP)

Prof. Dr. Danilo Marcondes (PUC-Rio)

Prof. Dr. Frank Thomas Sautter (UFSM)

Prof. Dr. Guido Imaguire (UFRJ)

Prof. Dr. Marco C. Ruffino (Unicamp)

Prof. Dr. Nelson Gonçalves Gomes (UNB)

Prof. Dr. Túlio Xavier de Aguiar (UFMG)

SUMÁRIO

I - ARTIGOS INÉDITOS

- 11 A FALÁCIA NATURALISTA NA METAÉTICA CONTEMPORÂNEA:
USOS E EQUÍVOCOS
*THE NATURALISTIC FALLACY IN CONTEMPORARY METHAETHICS: USES
AND MISTAKES*
Luca Nogueira Igansi - Universidade do Vale do Rio dos Sinos
- 33 A INCLUSÃO DO OUTRO NAS SOCIEDADES COMPLEXAS
THE INCLUSIOIN OF OTHERS IN COMPLEX SOCIETIES
João Paulo Rodrigues - Universidade Estadual de Londrina
- 55 RAZÕES PÚBLICAS COMO RAZÕES PRÁTICAS
PUBLIC REASONS AS PRACTICAL REASONS
Tiaraju Molina Andrezza - Universidade do Vale do Rio dos Sinos

II - TRADUÇÃO

- 75 CÉREBROS NUMA CUBA
BRAINS IN A VAT
Hilary Putnam
Tradução de L. H. Marques Segundo - Universidade Federal de Santa Catarina

III - RESENHA

- 97 SCHOPENHAUER E OS FILÓSOFOS, DE HÉLIO LOPES DA SILVA
*REVIEW: HÉLIO LOPES DA SILVA'S "SCHOPENHAUER AND THE
PHILOSOPHERS"*
José Luiz Furtado - Universidade Federal de Ouro Preto

Artigos Inéditos

A falácia naturalista na metaética contemporânea: usos e equívocos

The Naturalistic Fallacy in Contemporary Methaethics: Uses and Mistakes

*Luca Nogueira Igansi
Universidade do Vale do Rio dos Sinos*

Resumo

Será realizada uma análise da falácia naturalista conforme Moore e sua relação com Hume para clarificá-la e atualizá-la na lógica formal contemporânea, que denunciará seus limites na metaética atual. Serão identificadas as origens da expressão falácia naturalista em Moore e refinados seu significado e uso, contrastando-os com o argumento da questão-em-aberto e a Lei de Hume. Sua aplicação é identificada em quatro aspectos: invalidamente, enquanto argumento da questão-em-aberto, por não estabelecer uma relação metafísica e prover falsos cognatos; e validamente, como erro categorial, erro de identificação equivocada e erro inferencial; este último relacionável à Lei de Hume como um erro procedimental de argumentação. Breves aplicações evidenciam os limites da falácia naturalista em que, contraditoriamente, não é uma falácia, salvo por apenas um de seus usos, assim como não refuta todas teorias naturalistas (como hedonistas e utilitaristas), além de ser argumento em prol de algumas (como as não-cognitivistas).

Palavras-chave

Falácia Naturalista. Metaética. Naturalismo Moral. Moore. Hume. Ética.

Abstract

The naturalistic fallacy according to Moore and its relation to Hume will be analyzed for an exposition both clear and updated in contemporary formal logics, which will denounce its limited scope in current metaethics. I'll identify the origins of the expression naturalistic fallacy in Moore and attempt to refine its meaning and use, contrasting its relationship to the open-question argument and Hume's Law. Its application is identified in four aspects: invalidly as the open-question argument for not establishing a metaphysical connection and providing false cognates, and validly as a categorical error, a misidentification error and an inferential error; this last one relatable with Hume's Law as a procedural argumentation error. Brief applications show the limits of the naturalistic fallacy in that, contradictorily, it isn't a fallacy but by one of its uses, as well as not rebutting every naturalistic theory (as hedonism and utilitarianism) besides being also an argument in favor of some (non-cognitivists).

Keywords

Naturalistic Fallacy. Metaethics. Moral Naturalism. Moore. Hume. Ethics.

1. Moore, Hume e uma miríade de falácias

Um dos temas mais tratados em textos de ética e metaética a partir do início do século XX é, sem dúvida, o do problema da *falácia naturalista*¹. O autor que cunhou o termo inicialmente, G. E. Moore, exerceu uma das mais significativas influências sobre a ética analítica contemporânea. Dos realistas mais radicais aos não-cognitivistas, a grande maioria das teorias éticas de hoje se referem, de alguma forma e em algum grau, às obras de Moore: de

¹ - Do original em inglês, *naturalistic fallacy*, foram feitas diferentes traduções como “falácia naturalista” e “naturalística”. A fim de consistência, utilizarei apenas da primeira.

acordo com a entrada na Stanford Encyclopedia of Philosophy de *Moore's Moral Philosophy* de Thomas Hurka, Moore foi responsável por influenciar fortemente desde o campo da arte, com sua defesa do valor intrínseco da estética, passando pelo positivismo lógico e neoaristotelianismo, até o recente fortalecimento do intuicionismo e não-cognitivismo (HURKA, 2010, p. 13-16; DALL'AGNOL, 2005, p. 356). Mas, por mais conhecido que seja o autor e seu argumento da falácia naturalista, não apenas são notáveis as diferenças nas interpretações de cada autor, como também do próprio Moore em sua mais citada obra quanto ao uso desta expressão, o *Principia Ethica*, que como Dall'Agnol frisa: “Segundo o Index do *Principia*, há 23 contextos em que a expressão ocorre no livro e, de acordo com Rohatyn (1987, p. 20), com 12 sentidos diferentes” (DALL'AGNOL, 2005, p. 150).

Não é difícil imaginar a razão de tamanha discordância. Patricia Churchland (2008), assim como Adriano Brito (2010) e Thomas Hurka (2010), aponta que a falácia naturalista é o mesmo que o *argumento da questão-em-aberto* (*open-question argument*). Brito ainda vai além e indica que este último argumento de Moore é apenas um desenvolvimento da famosa *Lei de Hume* – embora com intentos um pouco dissonantes –, na passagem em que Hume demonstra estranheza e falta de conexão entre o uso de *ser* e *dever* [*ser*]; ponto que e.g. Geoff Sayre-McCord (2012) concorda, mas sem se comprometer com a nomenclatura de *falácia naturalista*. Daniel Dennett (1995) ainda confunde a nomenclatura do argumento mooreano com a *Lei de Hume*, afirmando que a última é a formulação de Moore no *Principia*, e apresenta uma crítica similar à de Churchland no início de seu livro – embora ela, por sua vez, saiba diferenciar o argumento de Hume do de Moore.

Vou me aprofundar nesses argumentos na próxima seção, assim como procurarei apresentar e esclarecer quais usos são coerentes com a proposta original. Farei grande uso da exposição de Darlei Dall'Agnol quanto à teoria mooreana, em suas críticas e defesas, como apresentadas em sua obra *Valor Intrínseco* (DALL'AGNOL 2005). Após tal análise, procurarei investigar sua relação com a posição de Hume para, então, brevemente analisar o escopo de ação desse tal argumento.

1.1 O Contexto do Argumento

O objetivo de Moore ao apresentar sua obra *Principia Ethica* era o de salientar o caráter *sui generis* do estudo da Ética, por meio do estabelecimento de que o bom, enquanto valor intrínseco², era indefinível. É um marco do início do século XX, influência-mor da

2 - Doravante, vou me referir à interpretação de *bom* enquanto valor intrínseco apenas como *b_{vi}*, conforme Dall'Agnol, dessa forma diferenciando-o de qualquer outro uso.

ética contemporânea, por seu intento metaético de colocar como objeto da ética – com objetivo semelhante ao de Kant em seu *Prolegomena*, porém, com foco na ética em lugar da metafísica (MOORE 1959, p. IX) –, não *quais* são as coisas boas, mas, sim, *o que é bom em si mesmo* (MOORE 1959 §13 ¶3). Como o próprio autor coloca:

I have endeavoured to write Prolegomena to any future Ethics that can possibly pretend to be scientific. I have endeavoured to discover what are the fundamental principles of ethical reasoning; and the establishment of these principles, rather than of any conclusions which may be attained by their use, may be regarded as my main object (MOORE, 1959, p. IX).

Segundo o autor, há uma miríade de coisas que podem ser chamadas de *boas*, mas a Ética está preocupada sumariamente com aquelas referentes ao que é *certo* ou *errado* (§1¶1, §2¶1). Assim ele busca identificar dois aspectos principais de sentenças normativas (p. VIII-IX), ou seja, referenciais a normas e regras de cunho moral: que (i) possuem apofanticidade insulada, ou seja, cada caso é específico e independe de provas (aspecto não-moral), e (ii), a evidência para o dever moral (*moral ought*) é intuitiva, ou seja, auto-evidente, assim como causal quanto à ação em questão – o que também implica sua objetividade (cf. MOORE 1959 p. IX, HURKA 2010, p. 1-2). Assim, é impossível para a ética ou qualquer ciência exaurir todas as instâncias em que ocorre o uso de b_{vi} , assim como identificar b_{vi} com qualquer estado-de-casos em particular. Essa perspectiva acaba por atribuir a característica que define o diferencial do objeto e do modo de estudo da ética: a indefinibilidade do b_{vi} .

O b_{vi} , entendido assim por Moore, é, então, um objeto simples, que pode constar em muitos sistemas holísticos associando-se a outras coisas. Portanto, claramente para o autor, identificar b_{vi} em um juízo analítico é cometer um erro, assim como querer explicar o conceito de *amarelo* ou qualquer outra cor. Enquanto, *e.g.*, nas ciências, podemos explicar o que é o conceito *cavalo* descrevendo suas partes individuais, com as cores isso não acontece: é uma experiência subjetiva que, simultaneamente, é causal a uma experiência empírica e, portanto, objetiva. Da mesma forma se dá o b_{vi} , como o *valor intrínseco* que determinadas coisas apresentam, ou seja, estados-de-casos bons por si mesmos, independentes de quaisquer outros objetos ou outras coisas boas³.

3 - Dall'Agnol (2005) apresenta um problema na diferente interpretação entre a noção de valor intrínseco como Moore o apresenta como *per se* ou pendendo mais para uma interpretação aristotélica. Não vou entrar nesse debate, vou apenas partir da concordância com a posição de Dall'Agnol, em que a perspectiva aristotélica “fornece um esquema conceitual capaz de levar a uma análise mais plausível de valor intrínseco” (p.16).

A partir dessa posição, Moore apresenta a célebre formulação à qual outros autores frequentemente se referem ao tratar da falácia naturalista:

It may be true that all things which are good are also something else, just as it is true that all things which are yellow produce a certain kind of vibration in the light. And it is a fact, that Ethics aims at discovering what are those other properties belonging to all things which are good. But far too many philosophers have thought that when they named those other properties they were actually defining good; that these properties, in fact, were simply not other, but absolutely and entirely the same with goodness. This view I propose to call the naturalistic fallacy [...] (MOORE 1959, §10 ¶3).

Mas como vimos no início da seção, esse é apenas um dos usos do termo *falácia naturalista* que Moore faz por meio de sua obra. A primeira e mais comum interpretação dessa passagem é associada diretamente com o chamado *argumento da questão-em-aberto*, que, como vimos, por si só é como muitos autores resumem a crítica de Moore ao naturalismo, alguns reduzindo a falácia naturalista a esse argumento. Acompanhando Dall'Agnol, apresentarei que tal perspectiva é equivocada, e o argumento, no mínimo, inútil, mas que não deve ser tomado como o argumento da falácia naturalista. Vou então, posteriormente, apresentar uma formulação aceitável desta e mostrar qual sua relação com o problema do ser/dever em Hume, no conhecido argumento da *Lei de Hume* ou *guilhotina de Hume*.

É importante fazer uma breve colocação acerca da intenção de Moore quanto à sua investigação ética. Uma das interpretações errôneas da proposta da falácia naturalista é a de supor que o autor se compromete com uma teoria platônica da moral. O debate acerca desse tema vai além do escopo deste trabalho, então, a fins práticos, adotarei a perspectiva de Dall'Agnol, uma vez que será ela que discutirei nas próximas seções. Dall'Agnol coloca que isso não se segue porque o termo *naturalista*, conforme empregado por Moore, faz referência a *objeto natural*, e não à teoria naturalista *per se*. O comprometimento com a veracidade dos juízos morais faz referência ao seu valor intrínseco, e não à existência do b_{vi} enquanto uma entidade, científica ou metafísica. Isso é uma consequência das observações de Moore que apontei acima, a partir da autonomia da ética, dado o caráter *sui generis* dos juízos de caráter ético (cf. HURKA p.1, MOORE 1956 caps. 1 e 2, DALL'AGNOL 2005⁴), em que se sucede que b_{vi} , enquanto termo

4 - Dall'Agnol dedica a seção 5.1 de sua obra, em específico, para refutar a necessidade de uma teoria ontológica realista para explicar o realismo dos juízos morais, conforme a teoria de Moore. Não vou entrar no mérito específico do tema no momento, mas é suficiente dizer que o autor coloca Moore como um conceitualista com fortes influências kantianas e aristotélicas.

bom no sentido ético, nada mais é do que um operador lógico apofântico de sentenças morais.

1.1.1 Uma Questão em Aberto?

Durante essa defesa da indefinibilidade de b_{vi} , Moore apresenta o *argumento da questão-em-aberto* (*open-question argument*), que como vimos acima, muitos identificam como a falácia naturalista *per se*. A primeira instância da ocorrência do termo *questão-em-aberto* se dá em §14 ¶5 do *Principia*, em que o autor coloca que, ao definirmos um termo (no caso, o b_{vi}) mediante um juízo analítico de identidade em que o próprio significado da palavra o decide; ninguém pode pensar de outra forma exceto por uma confusão. Isso se dá pois se pensa de tal termo que não é uma questão aberta, ou seja, não está aberto a questionamentos: é uma questão fechada, como *todo solteiro é não-casado*; não há sentido em perguntar *se todo solteiro é não-casado*, pois o significado do sujeito está contido no predicado, e vice-versa. Porém, isso não pode ocorrer na ética, pois independentemente do que quisermos definir – o que é b_{vi} , o que *desejamos desejar*, etc. – sempre obteremos uma questão em aberto: a expressão *o prazer é bom*, e.g., sempre dá margem para perguntarmos com sentido *é o prazer, afinal, bom?*, claramente mostrando que não é o mesmo caso de se perguntar *é o prazer, afinal, prazeroso?*, denunciando assim a ausência de analiticidade do juízo. Assim, o que é “bom, por definição, não significa nada que seja natural; e é logo sempre uma questão em aberto se qualquer coisa natural é boa”⁵ (MOORE 1956, §27 ¶3).

Uma das críticas mais comuns contra esse argumento remete ao problema de definição *per se*. Prinz resume o problema com o seguinte argumento, colocando que “[s]ince Moore’s time, philosophers have come to recognize the existence of a posteriori identities—metaphysical facts that cannot be discovered through conceptual analysis” (PRINZ, 2007a, p. 39). Alguns dos autores que apresentam essa crítica, além de Prinz, são Hurka (2010), Dall’Agnol (2005), Lenman (2013), Sturgeon (2006), e, até certo grau, Churchland (2008). Todos concordam que certas identidades ontológicas de objetos não são sinônimas, podendo ser empiricamente (*a posteriori*) descobertas, como e.g., *água* poder ser denominada como H_2O ; *calor* como *movimento molecular* e assim por diante. Churchland (p. 186-190) e Sturgeon (p. 96-100) levam adiante que isso seria característico do discurso científico; e, entendendo-o como discurso principal das teorias naturalistas, assim também o seria, em algum grau, para o naturalismo moral. Outro lado dessa crítica seria o que Hurka denomina como o *paradoxo da análise* (2010, p. 2-3), em que, ao levarmos o argumento da questão-

5 - “Good does not, by definition, mean anything that is natural; and it is therefore always an open question whether anything that is natural is good” (tradução nossa).

em-aberto a sério, resultaria que, num legítimo espírito de Wittgenstein em seu *Tractatus*, as únicas definições possíveis seriam tautológicas, e, portanto, inúteis. Sturgeon e Dall'Agnol, assim como, em parte, Churchland, concordam com essa posição, e apontam também que para a linguagem comum (não somente a científica) isso seria um problema, dado seu caráter dinâmico; e.g. chamar um grande amigo de *irmão*, como ocorre informalmente no cotidiano, não comprometerá a definição de nenhum termo envolvido, assim como diferentes usos de *banco*, como para sentar, retirar dinheiro ou de um rio, não implica que o termo *banco* seja indefinível – o que Sturgeon chama de *falsos negativos* (2006, p. 96-98, cf. também DALL'AGNOL, 2005, p. 127).

Algumas outras críticas são como a do não-cognitivismo em geral, como aponta Sayre-McCord, que simplesmente não se comprometem à apofanticidade dos juízos morais, enaltecendo seu mero intento em expressar uma emoção, aprovação, etc.⁶ Sturgeon possui uma agenda não-cognitivista em sua argumentação, mas, ainda assim, chega ao ponto de refutar o argumento mooreano, colocando que b_{vi} pode, inclusive, não ser *reduzido* a conceitos naturais, mas isso não significa que ele não seja *per se* natural (2006, p. 97-100). Ele não se compromete com um realismo moral, pois, mesmo o discurso científico depende de uma comunidade científica, que, por sua vez, está sujeita a uma série de fatores que tornam impossível uma objetividade em qualquer grau de juízos morais. Dall'Agnol invoca também que aparentemente o único motivo de Moore aceitar esse argumento em primeira instância foi por já partir do pressuposto que b_{vi} era indefinível, cometendo assim uma falácia de *petitio principii*. Churchland, apesar de não parecer diferenciar *a posteriori* de *a priori*, apresenta um argumento interessante, fruto de seu eliminativismo materialista, em que o que é *intuicional*, assim como o que é *valor*, é necessariamente dependente de nosso funcionamento neurológico: damos valor às coisas e intuimos seu *valor intrínseco*, dada nossa configuração neurológica, fruto de nossa genética e ambiente (experiência), em um paradigma evolutivo de seleção natural: *our perceptions are permeated with value* (CHURCHLAND, 2008, p. 189) – embora ela deixe claro não querer comprometer-se com teoria normativa alguma. Vou apresentar algumas críticas de Dall'Agnol a essa posição na próxima seção, e em seguida refutar tais críticas.

Há mais uma posição que gostaria de considerar brevemente. É a apresentada por Horgan & Timmons, em 1992, procurando *reviver* o argumento da questão-em-aberto, em resposta ao crescente realismo naturalista da *Escola de Cornell*, representado em Boyd. Este autor defende um realismo científico em que a evidência empírica serve como base não-redutível e heurística para uma teoria moral consequencialista, a partir da regulação

⁶ - A ironia de muitos autores não-cognitivistas naturalistas, emotivistas e afins terem utilizado o argumento da questão-em-aberto para justificar suas posições também é notada por vários autores.

constante de crenças enquanto um fenômeno social em um *equilíbrio reflexivo* coerentista entre a teoria e sua metodologia. (cf. DALLAGNOL, 2005, p. 181-90; LENMAN, 2013, p. 6) Dall’Agnol aponta que a perspectiva de Boyd concebe termos éticos como *designadores rígidos* (no sentido de Kripke) de que há coisas relativas às necessidades naturais básicas que constituem uma definição natural de b_{vr} , ainda que não o identificando como apenas uma única propriedade natural. A crítica de Horgan & Timmons relacionaria o argumento da questão-em-aberto com um experimento-de-pensamento similar ao proposto por Putnam, em que, em uma *Terra Gêmea* na qual

[w]e can readily conceive of societies that are pervasively morally wrong-headed, where their use of evaluative concepts is systematically directed and regulated by, we would want to say, the wrong stuff. What they call “good” is not good we insist and we think this disagreement real, not simply a case of conceptual cross-purposes. (...) We may well think [they’re] mistaken but, if so, we think [their] error a substantive ethical one and not a case of conceptual confusion (LENMAN, 2013 p. 5).

Isso iria de encontro à concepção kripkeana da terminologia moral conforme Boyd a entende. O problema, como Dall’Agnol enfatiza, é que, apesar de fugir do problema de tratar de sinonímia enquanto propriedades idênticas, ainda jaz em uma questão definicional – e teorias éticas não podem ser provadas (naturalistas ou não) por elas *per se* (2005, p. 185). Em suma, de acordo com Lenman, Boyd

accepts that the reference of a kind term is indeed determined by what feature of the world turns out to occupy a particular role. But instead of this role being determined by the concepts held by users of the term, it may instead, he proposes, be determined by the ways in which their use of the term contributes to the successes they achieve by it (LENMAN, 2013, p. 6).

Além das várias teorias já supracitadas, há uma série de outras críticas realizadas, mas acredito que por ora essas sejam suficientes para ilustrar os ataques posteriores a Moore quanto a esse controverso argumento e algumas considerações sobre tais. Mas, por fim, além de todas as críticas realizadas por diversos autores desde a formulação do argumento original, talvez a mais importante ainda seja a do próprio Moore, no prefácio de sua segunda

edição do *Principia*. Conforme apontam Dall’Agnol (2005, p.133) e Hurka (2010, p. 7) a indefinibilidade do *bom* não é central para seu argumento, aparentemente influenciado por Russell e seus outros debatedores contemporâneos (Brentano, Sidgwick, etc.). Dessa forma, torna-se evidente não apenas que o *argumento da questão-em-aberto* é irrelevante para o argumento mooreano, como também – o que é mais importante para nós neste momento –, que não é central para seu argumento da falácia naturalista, uma vez que, no mesmo prefácio, ele ainda continua defendendo essa posição, apesar de não se comprometer com a indefinibilidade do b_{vi} . De uma forma mais geral, Dall’Agnol enfatiza que “uma teoria não depende substancialmente das definições que ela apresenta” (2005, p. 129).

Todavia, o argumento da questão-em-aberto tem alguma utilidade, como Prinz e Dall’Agnol reconhecem. Embora seja inútil para refutar qualquer teoria, Dall’Agnol coloca que “talvez o que [este argumento] possa fazer é desempenhar uma função heurística levando à redefinição de certos termos que não estão bem definidos” (2005, p. 133). Prinz também reconhece um valor epistemológico ao argumento:

Open-question arguments cannot establish metaphysical conclusions, but they can establish conceptual conclusions. If it is an open question whether some natural property N is good, then the concept of N cannot be part of the concept of goodness. Were there a conceptual link between the two, the question would be closed. (PRINZ, 2007^a, p. 39).

Não vou adentrar, nesse momento, na teoria de Prinz, mas a título de ilustração é suficiente dizer que ele usa dessa linha de raciocínio para defender seu *sentimentalismo construtivista*, onde juízos morais são frutos da relação conceitual entre nossas emoções (aspecto fisiológico/natural) e nossa comunidade moral (aspecto cultural), de modo a provar que – ao contrário de Mackie – todo juízo moral é verdadeiro, embora necessite de uma comunidade moral para seu critério de veracidade e aceitação.

Assim, estabelecemos que o argumento em questão não somente é inválido – ou como muitos autores colocam jocosamente, após termos *fechado o argumento da questão-em-aberto* –, mas também mostramos que ele não constitui o que Moore intentava apresentar como falácia naturalista. Reitero aqui o que coloquei acima: apesar daqueles autores terem, com sucesso, refutado o argumento da questão-em-aberto, a falácia naturalista continua intacta, e, portanto, estão fundamentalmente errados e lutam contra sombras. Faço das palavras de Rohatyn as minhas: “a influência fenomenal do argumento da ‘questão-em-aberto’ de G. E. Moore, na metaética do século

XX não é merecida” (ROHATYN, 1995 *apud* DALL’AGNOL, 2005, p. 133).

Na próxima seção irei proceder para a formulação do item central deste artigo, o argumento da falácia naturalista conforme Moore a entende, assim como posteriormente relacioná-lo ao argumento conhecido como *Lei de Hume*, frequentemente associado com o primeiro na literatura sobre o tema.

1.1.2 A Falácia Naturalista

Não há na história da filosofia um nome pior elaborado que o da falácia naturalística.

– DALL’AGNOL, 2005

Conforme exposto nas seções anteriores, não apenas há uma miríade de interpretações de vários autores acerca da *falácia naturalista*, como há também na leitura do próprio Moore. Felizmente, no documento encontrado e publicado postumamente, conhecido como o *Prefácio à Segunda Edição* para sua obra *Principia Ethica* (cf. MOORE, 1993, p. 1-33), o autor delinea três pontos principais do objetivo que possuía com o uso do termo, a saber (DALL’AGNOL 2005 p. 151):

- (i) *ou* quando se confunde B com um predicado do tipo a ser definido;
- (ii) *ou* mantendo que ele é idêntico a tal predicado;
- (iii) *ou* fazendo uma inferência *baseada* em tal conclusão.

Como vimos acima, ao início da seção 1.1, devemos entender B aqui (no original G) como b_{vi} , no sentido que o autor coloca (MOORE, 1993, p. 4-5) como relacionado ao *certo* e *errado*, mais do que tudo, e não no sentido trivial – ou segundo Dall’Agnol, “refere-se assim ao valor intrínseco de algo” (DALL’AGNOL, 2005, p. 152), como ele utiliza com o termo b_{vi} e como uso aqui b_{vi} –, que é *the sense which is of the most fundamental importance for Ethics*. (MOORE, 1993, p. 5)

Dall’Agnol procederá para *traduzir* essas diretrizes para uma terminologia lógica mais

atualizada, generalizável e aplicável. Dessa forma, ele identifica o item (i) como um *Erro Categorical* (EC), o item (ii) como um *Caso de Identificação Equivocada* (CIE) e o (iii) como um *Erro Inferencial* (EI) – o qual relacionará posteriormente com a *lei de Hume* –, o único que seria uma *falácia* propriamente dita.

A primeira formulação, (i) como Erro Categorical, consiste em associar tipos lógicos diferentes, ou seja, tomar uma classe como membro de si mesma. Baseado em Ryle, Dall’Agnol (2005, p. 153) esclarece que é como procurar a *Universidade* em prédios específicos de um campus, ou afirmações como *o número 7 é verde*. Esse erro implica lógica e conceitualmente, como é muito claro, em proposições absurdas sem sentido, pois uma classe e seus membros constituem tipos lógicos diferentes.

No segundo item temos (ii) o Caso de Identificação Equivocada, que é semelhante ao anterior, mas constitui uma confusão entre objetos de categorias diferentes ou de diferentes objetos naturais. Isso varia desde confundir uma pessoa com outra no cotidiano até dizer que *B é idêntico ao prazer*: o objetivo aqui é mostrar que b_{vi} é um tipo único, e segue a linha de raciocínio original de Moore de que b_{vi} é indefinível. Vimos que isso é um tanto problemático⁷ pela fraqueza do argumento da questão-em-aberto e falta de necessidade de compromisso com a indefinibilidade de b_{vi} ; mas Dall’Agnol aponta que há um dos sentidos que Moore apresenta em seu segundo prefácio que é proveitoso para a análise metaética: “tal e tal está identificando B com algum predicado *natural* ou *metafísico*” (MOORE 1993 p.16, grifo do autor). É importante salientar que “natural”, como o autor entende nesse contexto específico, possui três usos: (a) como do domínio das ciências (biologia, psicologia, antropologia, ciências sociais, etc.), (b) como *instinto* como antônimo de *escolha* e (c) como “normal”, “comum.”

Por fim, temos o terceiro caso, (iii) como Erro Inferencial. Talvez a única instância do argumento que possa ser denominado como “falácia” *per se* (ainda que numa concepção mais abrangente que tradicionalmente na lógica), uma vez que consiste na análise de um argumento de forma a identificar se alguma das premissas utilizadas para inferir a conclusão em questão cometeu algum dos outros dois erros que constam acima. Ou seja, o problema não é a inferência, como numa falácia tradicionalmente entendida, mas o absurdo que estaria em uma ou mais premissas influenciando, assim, uma conclusão igualmente absurda. Então, “[a]ssim como de uma contradição tudo se segue, de uma premissa absurda nada se segue” (DALL’AGNOL, 2005, p. 158).

7 - Para mais informações quanto ao problema de “definição” em Moore, cf. a obra de Dall’Agnol utilizada no presente trabalho, assim como a seção anterior.

1.1.3 O Problema do Ser-Dever

Nessa última definição da falácia naturalista entra a relação com a famosa *Lei de Hume*, o problema do ser-dever. Essa é outra posição que possui certa variedade de interpretações, mas, desta vez, é pela escassez de informações que jaz na fonte original em Hume. A famosa passagem que originou o igualmente conhecido argumento é a seguinte:

In every system of morality, which I have hitherto met with, I have always remark'd, that the author proceeds for some time in the ordinary way of reasoning, and establishes the being of a God, or makes observations concerning human affairs; when of a sudden I am surpriz'd to find, that instead of the usual copulations of propositions, is, and is not, I meet with no proposition that is not connected with an ought, or an ought not. This change is imperceptible; but is, however, of the last consequence. For as this ought, or ought not, expresses some new relation or affirmation, 'tis necessary that it shou'd be observ'd and explain'd; and at the same time that a reason should be given, for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it (HUME, 1739-40, 3.1.1.27, grifos do autor).

Adriano Brito (2010) pode ter errado em sua interpretação da falácia naturalista como o argumento da questão-em-aberto em Moore, ou, ainda, de relacionar o último com o supracitado argumento humeano; mas ele está certo ao menos em focar no argumento humeano para tratar da falácia naturalista. Ele delinea o argumento afirmando que Hume

aponta que a dedução de deveres a partir de premissas formuladas com elementos factuais não é válida e isso está de acordo com sua tese fundamental de que as distinções morais não são percebidas pela razão e não são distinções de objetos. Ora, por essa razão, juízos morais não podem ser verdadeiros ou falsos, no sentido lógico usual desses termos. Na verdade, eles não são juízos descritivos, mas avaliativos (BRITO, 2010, p. 220).

Ou, como Sayre-McCord sucintamente coloca, “*every valid argument for an evaluative conclusion either includes or presupposes some evaluative premise. And, as a result, there is no value neutral argument for an evaluative conclusion*” (SAYRE-MCCORD, 2012, p. 6).

Dall'Agnol segue essa linha de interpretação, afirmando que a única interpretação correta desse argumento humeano é que

[j]ulgamentos éticos (sejam valorações ou enunciados normativos) são categorialmente distintos das descrições. Identificar fatualidade com normatividade é um erro categorial, pois elas possuem diferentes propriedades (DALLAGNOL, 2005 p. 180).

Assim, percebemos que, *per definitionem*, a Lei de Hume está incluída na formulação geral de falácia naturalista proposta por Dall'Agnol, conforme vimos anteriormente, embora seja insuficiente para compreender sua totalidade de abrangência. Vale ressaltar, aqui, a dissonância vital de paradigma entre Moore e Hume: o primeiro possui intentos de desenvolver uma teoria ética acerca do valor de cunho antirreducionista, enquanto o segundo deseja apontar falta de causalidade entre enunciados morais descritivos e prescritivos.

Brito apresenta e critica algumas formulações de tentativas de superação desse problema, como presentes em McIntyre, Zimmerman e Searle. McIntyre formula que quando há uma proposição factual referente ao “querer”, ela se refere ao que é um consenso sobre um interesse comum baseado numa natureza comum humana e, deste, segue-se um dever de que tal seja realizado: um dever que surge de nossa natureza factual. A crítica é óbvia, pois, por mais que se estabeleça que há valor moral na vontade, o dever “envolve um elemento normativo que não aparece na frase constatativa sobre a vontade geral” (BRITO, 2010, p. 221). Uma crítica semelhante se segue com Zimmerman, que formula que um dever segue-se de um fato toda vez que “é” possa ser intercambiável com “deve” em uma proposição; há, nesta, uma obrigação concomitante com o desejo do falante. Isso, porém, gera um problema semântico da ligação entre expressar um desejo e gerar uma obrigação, pois esta última instância não é sinônimo de vontade. Novamente aqui há uma ausência de conexão entre o que é descritivo e o que é normativo, e portanto, é uma posição descartável. A terceira formulação, de Searle, é um pouco mais robusta. Searle formula que, uma vez que uma promessa é realizada, por exemplo, segue-se que há uma obrigação de ser cumprida: é possível formularmos uma descrição em que “X promete a Y pagar-lhe cinco reais”, e que, desta, segue-se uma obrigação necessária de X para com Y e, assim, uma norma gerada pela mera descrição de um fato. Apesar de astuto, o argumento é problemático por definir a “descrição da promessa” como ausente de peso normativo: isso não é o caso, pois o ato de prometer envolve um critério normativo da intenção do locutor de cumpri-la. Assim, o

argumento cai em uma *petitio principii*, partindo do que quer provar.

A partir das colocações de Brito, podemos delinear mais claramente a relação do problema da Lei de Hume com a falácia naturalista, conforme vimos acima. Comete-se um *erro categorial* ao tentar reduzir juízos normativos a sentenças descritivas, pelo fato de haver um abismo lógico entre *ser* e *dever-ser*, e, “por isso, a falácia naturalística pode ser vista como coibindo também uma inferência direta de um *dever-ser* a partir do *ser* e qualquer tentativa de reduzir o primeiro ao segundo” (DALL’AGNOL, 2005, p. 178). Estabelecendo essa relação, vou analisar algumas outras posições sobre a formulação humeana para avaliar a relevância em nosso escopo argumentativo atual.

Patricia Churchland, por sua vez, inicia sua obra *Braintrust* (2008) com uma breve colocação sobre essa questão em Hume. Ela identifica corretamente que Hume apresenta esse problema com o intento de mostrar que a resolução de problemas não se dá dedutivamente por meio puramente da razão. Mas autora realiza uma diferenciação entre *derive* (derivar), que ela identifica como o ato de *deduzir*, e *infer* (inferir): “[i]n a much broader sense of ‘infer’ than derive you can infer (figure out) what you ought to do, drawing on knowledge, perception, emotions, and understanding, and balancing considerations against each other” (CHURCHLAND, 2008, p. 6, grifos da autora). Essa perspectiva é parcialmente compatível com a interpretação que vimos sobre a moralidade em Hume e, como coloca Cohon, numa interpretação da *lei de Hume*, é que “one cannot make the initial discovery of moral properties by inference from nonmoral premises using reason alone; rather, one requires some input from sentiment” (COHON, 2010, p. 8). Aqui, Churchland está convencida de provar que, a partir da observação de certos fatos, é possível *inferir* um *dever* (ought). Porém, ironicamente, Hume foi um grande crítico da causalidade e da inferência indutiva, o que torna a formulação de Churchland inválida. A relação entre o argumento de Churchland e a crítica de Hume à indutividade mostra-se na perspectiva de Hume, baseada em que “the experience of constant conjunction fosters a ‘habit of the mind’ that leads us to anticipate the conclusion on the occasion of a new instance of the second premise” (VICKERS, 2013, p. 7).

Assim a inferência indutiva, ou *probabilística*, como Hume a chama, não possui necessidade alguma ontológica ou epistemológica. Ou seja, nenhuma objetividade, sendo nada mais do que a força individual de adquirirmos e mantermos hábitos. E assim, “[t]he objectivity of causality, the objective support of inductive inference, is thus an illusion, an instance of what Hume calls the mind’s ‘great propensity to spread itself on external objects’ (THN, 167)”. (VICKERS 2013 p. 8).

Apesar de ter identificado corretamente o que é a moralidade para Hume, Churchland equivoca-se procedimentalmente em sua explicação. A autora, com sua proposta de

materialismo eliminacionista, quer acabar justamente com o que seria o *sentimentalismo* humeano, uma vez que essa terminologia estaria inclusa no que ela chama de *psicologia popular* (folk psychology), que, segundo ela, deverá cair por terra com o progresso das neurociências. Assim, apesar de sua formulação sobre o que é moralidade envolver fatores semelhantes a Hume, fundamentalmente são incompatíveis, uma vez que para Hume “[*m*]orals excite passions, and produce or prevent actions. Reason of itself is utterly impotent in this particular. The rules of morality, therefore, are not conclusions of our reason” (HUME 1739-40 3.1.1.6). Dennett segue um infortúnio semelhante, associando à *Lei de Hume* (que ele chama de “argumento da falácia naturalista”, conforme Moore a formula, claramente equivocado) a uma mera simplicidade de raciocínio, partindo para a crítica da sociobiologia de Wilson como uma visão simplista da relação entre a natureza e a moralidade. Ele acredita que, ao mostrar a cultura (enquanto *memes*) como um fator central para a moralidade, a associação direta entre natureza e normatividade sem mediações é um reducionismo *ganancioso* (DENNETT, 1995, p. 467-8, cf. também toda a seção 16.3). Esse é um intento semelhante a Churchland ao início de sua apresentação da posição humeana, em que o erro está em realizar inferências “estúpidas” diretamente de fatos, ao contrário das devidas inferências por meio do uso das ciências, como apresentei acima. Mas essa não é uma formulação válida pelo simples fato de não ser um argumento. Dennett não identifica argumento algum, apenas propõe uma formulação fraca em bases questionáveis, e Churchland presume a correção de seu argumento pela possível semelhança de intento com Hume, mas, como vimos, há uma série de incompatibilidades justificacionais que tornam os argumentos, no mínimo, conflitantes.

Mas o que é útil nessas apresentações para nós, apesar dos erros de ambos os autores, é como a teoria de Hume lida com a ciência, cerne do naturalismo contemporâneo. Apesar de empiricista, Hume enfatiza a importância dos sentimentos para o juízo moral que, portanto, é avaliativo e não objetivo. Sturgeon trata desse problema ao apresentar sua agenda não-cognitivista quanto ao problema do conhecimento moral em relação à Lei de Hume, uma vez que, sendo uma teoria naturalista, parte do pressuposto de que todo o conhecimento é exaurível pelas ciências empíricas. Segundo Sturgeon, o fenomenalismo e o behaviorismo lógico associavam livremente dado empírico com estados psicológicos, e que portanto não haveria nenhum “abismo” entre descrição e valor. Em resumo, “*the meanings of our ethical terms fix standards of evidence a priori, guaranteeing that certain common inferences from empirical, nonethical premises to ethical conclusions are reasonable or warranted*” (STURGEON, 2006, p. 103). O autor critica essa posição, em que claramente temos, *sim*, um abismo entre conclusões éticas, que são guias de ação (*action guiding*) e premissas não-éticas, que não o são; mas isso não é um problema, uma vez que temos pressupostos

psicológicos que formam a base de processos morais de forma substantiva. O problema resultante de um fundacionalismo epistemológico pode ser superado por meio de um coerentismo semelhante a um modelo rawlseano, como ele mesmo aponta, em que a crença é aprovada e mantida ao longo do tempo por meio do diálogo entre dados empíricos e teorias normativas em geral. Isso é o que Sturgeon chama de um não-cognitivism baseado numa teoria naturalista não-reducionista, intrinsecamente relacionada com sua formulação do discurso científico para o naturalismo, em oposição ao argumento da questão-em-aberto, como vimos acima.

1.2 Aplicações da Falácia Naturalista

Gostaria de considerar outras duas possíveis aplicações realizadas por Dall'Agnol, de acordo com as formulações da falácia naturalista vistas acima. A primeira é direcionada à sociobiologia, que o autor associa – erroneamente – a Matt Ridley e Richard Dawkins; e a segunda, direcionada a Kant.

A crítica de Dall'Agnol ao argumento de Ridley foca em sua obra *The Origins of Virtue* (1998), em que Ridley procura estabelecer as origens do comportamento moral por meio de estudos acerca da genética, etologia, neurociências e afins. Dall'Agnol critica essa posição apontando que Ridley associa a cooperação humana a um instinto comportamental natural. Essa posição, embora Dall'Agnol não faça referência, é conhecida como *altruísmo recíproco*, e foi proposta por Robert Trivers, em 1971, em seu famoso artigo publicado na *Quarterly Review of Biology*, no qual, após análise de certos comportamentos aparentemente altruístas, mas simultaneamente claramente promovidos pelo auto-interesse em outras espécies de animais, Trivers procura teorizar acerca do surgimento do modelo psicológico comportamental do *modus operandi* equivalente nos seres humanos. Esse comportamento, denominado de *altruísmo recíproco*, foi rastreado por Trivers no período Pleistoceno (entre 200.000~120.000 anos atrás até cerca de 8.000 anos atrás) como favorável à sobrevivência pelo contexto em que se vivia: nas savanas gramíneas, com caça de tamanho razoavelmente superior ao nosso, fazendo com que houvesse favorecimento na seleção natural por aqueles que, por algum motivo, optaram por viver em grupo. Assim o *altruísmo recíproco*, ou meramente o comportamento cooperativo como o entendemos, faria parte de nossa natureza, uma vez que, em termos de genética, a humanidade como um todo é descendente diretamente desses grupos que sobreviveram há milhares de anos. Dall'Agnol simplesmente aponta o paradigma superficialmente e o relaciona com a teoria do *gene egoísta* de Dawkins

(2006, 2008), além de afirmar que é uma teoria que está fundamentalmente errada por motivo de erro categorial, pois confunde a moralidade, que é cultural, com instintos, que são naturais.

A próxima crítica é direcionada a Kant. Em sua formulação do agente humano, Kant realiza uma distinção ontológica entre *ego transcendental* e *ego empírico*, em sua diferenciação entre *fenômeno* e *coisa-em-si*. O que é, primeiramente, uma mera distinção lógica, em sua segunda Crítica torna-se ontológica no que ele se refere: “há algo por trás daquilo que é fenomênico” (DALL’AGNOL, 2005, p. 171). Isso gera o problema do “mito cartesiano” em que há uma substância metafísica que está por trás da substância física. Essa confusão gera o que vimos como um erro categorial. Aqui fica claro que o escopo da falácia naturalista, conforme Moore, e comprovados por Dall’Agnol, Prinz, Sturgeon, Boyd e outros, não apenas abrange a dicotomia fato/valor como também ontológica, ou como Dall’Agnol coloca (2005, p. 173), “uma falácia metafísica”.

2 Considerações finais

Pelo que pudemos ver, apesar da vasta gama de tentativas de refutação da falácia naturalista, aparentemente as únicas que obtiveram algum nível de sucesso foram aquelas que endereçaram o argumento conhecido como Lei de Hume, dados os inúmeros problemas com a teoria da questão-em-aberto, conforme vimos anteriormente. As poucas linhas, vistas acima, de Hume sobre o problema parece que foram mais efetivas do que toda a bibliografia de Moore no que tange a similar intento.

Infelizmente, de qualquer forma, a falácia naturalista não é completamente efetiva contra qualquer forma de naturalismo. Como Dall’Agnol (2005 p. 173-4) aponta, uma série de teorias naturalistas conseguem fugir de seu ataque, como o hedonismo, o utilitarismo ou a ética das virtudes. Acredito, também, que seja inefetiva contra o altruísmo recíproco, como o autor a apresenta. Isso se dá por uma série de interpretações errôneas atribuídas por Dall’Agnol ao trabalho de Ridley, como, por exemplo, identificar sua teoria como sociobiologia, interpretar o “gene egoísta” como um gene específico, da genética como determinista, que há um reducionismo moral e no erro da implicação de que moralidade, genética e instinto seriam o mesmo tipo lógico (DALL’AGNOL 2005 p. 164-174). Essas afirmações são rebatidas em uma análise um pouco mais extensa dos trabalhos de Ridley, em que ele claramente se posiciona contra a sociobiologia (personificada em E. O. Wilson) justamente por ser uma teoria que se baseia no reducionismo

e determinismo que, segundo Ridley, são conflitantes com a interpretação-padrão atual do equilíbrio de importância entre natureza (genética, fisiologia, etc.) e cultura (experiências, meio social, etc.) (Cf. RIDLEY 1998, p. 18-9, 40, 103, 106; 1999, p. 306; 2003, p.180-1, 214-16, 219, 229, 250, 234, 125); ponto que é reforçado ainda mais em uma análise da teoria basilar do gene egoísta conforme Dawkins, que, da mesma forma, procura opôr-se à Wilson (2006, p. 94). O debate inteiro e suas implicações fogem ao escopo do atual trabalho, mas é suficiente colocar que o intento de Ridley e Dawkins é apresentar uma teoria descritiva por meio da simplificação de conceitos científicos sobre o fenômeno moral humano. Assim, não há teoria moral normativa para que se cometa alguma forma de falácia naturalista. Além do caráter antirreducionista que não procura resumir ou interpelar descrição de fatos básicos com valores ou normas, as teses apresentadas não são teorias morais, e sim paradigmas científicos acerca do comportamento humano. Entretanto, a teoria de Wilson, por sua vez, assim como a de Spencer, que Moore procura atacar, caem vítimas, sim, da falácia naturalista.

No mais, acredito que a problemática acima representa, da mesma forma, a que circunda as teorias morais naturalistas pós-humeanas (predominantes contemporaneamente no âmbito do naturalismo moral), em que há um prezar pelo empreendimento metaético de identificar e descrever o fenômeno moral, afastando-se de juízos de valor e normativos. Mas essa, novamente, é uma temática para outro momento. Encerro reconhecendo que a falácia naturalista possui grandes limites, suficientes para apenas servir como aviso para inferências simplistas e imprudentes – não diferentemente de como Churchland e Dennett a propõe, ainda que de forma equivocada –, e não somente aplicável à esfera do naturalismo. Assim, ela não constitui suficientemente o carrasco do naturalismo que muitos imaginam ou defendem ser. Ela, assim como a lógica como um todo, representa um algoz do argumento falacioso em um âmbito geral, em prol da manutenção da validade argumentativa para um debate sadio e coerente. Todavia, é apenas uma pequena faceta, e, definitivamente, não merece a alcunha que vestiu neste último século.

3 Referências

BRITO, A. Naves de. Hume e o Empirismo na Moral. *Philosophos* (UFG), Goiânia, v. 6, n.1, p. 42-64, 2001.

_____. Falácia Naturalista e Naturalismo Moral: do “é” ao “deve” mediante o “quero” in *Kriterion* vol. 51 nº 121. Belo Horizonte: Editora UFMG, 2010.

_____. *Ação & Valor na Moral naturalizada*. Relatório de pesquisa CNPq. Porto

Alegre, 2013.

_____. Naturalismo Moral in *Manual de Ética – Questões de Ética Teórica e Aplicada*. Org. João Carlos Brun Torres. Petrópolis: Vozes, 2014. No prelo.

CHURCHLAND, Patricia. *Braintrust*. Princeton: Princeton University Press, 2008.

COHON, Rachel. Hume's Moral Philosophy in *The Stanford Encyclopedia of Philosophy*, Fall 2010 edition, Edward N. Zalta (ed.). Stanford: Stanford University Press, 2010. Disponível em <<http://plato.stanford.edu/archives/fall2010/entries/hume-moral/>> Acessado em 26/09/2013

DALLAGNOL, Darlei. *Valor Intrínseco – Metaética, ética normativa e ética prática em G. E. Moore*. Florianópolis: Editora da UFSC, 2005.

DAWKINS, Richard. *O Gene Egoísta*, trad. Rejane Rubino. São Paulo: Companhia das Letras, 2008.

_____. *The Selfish Gene*. 30th anniversary edition. New York: Oxford University Press, 2006.

DENNETT, Daniel. *Darwin's Dangerous Idea*. London: Penguin Books, 1995.

_____. *Freedom Evolves*. USA: Penguin Books, 2004.

_____. Darwin's "strange inversion of reasoning". In *PNAS: Proceedings of the National Academy of Sciences of the United States of America*, Medford, v. 106, p.100661-100665, 16 June 2009. Disponível em <<http://www.pnas.org/content/106/suppl.1/10061>>. Acesso em: 20/10/2013.

HUME, David. *Investigações sobre o entendimento humano e sobre os princípios da moral*. Trad. José Marques. São Paulo: Editora UNESP, 2004.

_____. *A Treatise on Human Nature : Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects*. Leeds Electronic Text Centre, 1740. Disponível em <www.davidhume.org>, acessado em 11/02/2014.

_____. *An Enquiry Concerning the Principles of Morals*. Leeds Electronic Text Centre, 1777. Disponível em <www.davidhume.org>, acessado em 11/02/2014.

_____. *Enquiry concerning human understanding*. Leeds Electronic Text Centre, 1777. Disponível em <www.davidhume.org>, acessado em 11/02/2014.

_____. *Dialogues concerning natural religion*. Leeds Electronic Text Centre, 1779. Disponível em <www.davidhume.org>, acessado em 11/02/2014.

HURKA, Thomas. Moore's Moral Philosophy in *The Stanford Encyclopedia of Philosophy*, Summer 2010 edition, Edward N. Zalta (ed.). Stanford: Stanford University Press, 2010. Disponível em <<http://plato.stanford.edu/entries/moore-moral/>> Acessado em 17/09/2013

LENMAN, James. Moral Naturalism. in *The Stanford Encyclopedia of Philosophy* Edward N. Zalta (ed.) Stanford: Stanford University Press, 2013.

MOORE, George Edward. *Principia Ethica*. Cambridge: Cambridge University Press, 1959.

_____. *Principia Ethica – Revised Edition*. Cambridge: Cambridge University Press, 1993.

_____. *Ensayos Éticos*. Barcelona: Planeta-Agostini, 1994.

PAPINEAU, David. Naturalism, in *The Stanford Encyclopedia of Philosophy*. Spring, Edward N. Zalta (ed.) Stanford: Stanford University Press, 2009.

PRINZ, Jesse. *The Emotional Construction of Morals*. Oxford: Oxford University Press, 2007a.

RIDLEY, Matt. *Genome: The autobiography of a species in 23 chapters*. New York: ed. Harper Perennial, 2006.

_____. *Genome: The autobiography of a species in 23 chapters*. New York: HarperCollins Publishers Inc., 1999.

_____. *O que nos faz humanos: Genes, natureza e experiência*, trad. Ryta Vinagre. 2ª edição. Rio de Janeiro: ed. Record, 2008.

_____. *Nature via Nurture: Genes, experience, and what makes us humans*. New York: HarperCollins Publishers Inc., 2003.

_____. *The Origins of Virtue: Human Instincts and the Evolution of Cooperation*. Londres: Penguin, 1998

SAYRE-MCCORD, Geoff. Moral Realism in *The Stanford Encyclopedia of Philosophy* Edward N. Zalta (ed.) Stanford: Stanford University Press, 2009.

_____. Metaethics in *The Stanford Encyclopedia of Philosophy* Edward N. Zalta (ed.) Stanford: Stanford University Press, 2012.

STURGEON, Nicholas L. Ethical Naturalism in *The Oxford Handbook of Ethical Theory*. David Copp (ed.) Oxford: Oxford University Press, 2006.

TRIVERS, Robert. The evolution of reciprocal altruism in *The Quarterly Review Of Biology*, v. 46, p.35-57, Mar. Cambridge: Cambridge University Press, 1971.

VICKERS, John, The Problem of Induction, in *The Stanford Encyclopedia of Philosophy* (Spring 2013 Edition), Edward N. Zalta (ed.). Stanford: Stanford University Press, 2013. Disponível em <<http://plato.stanford.edu/archives/spr2013/entries/induction-problem/>> Acessado em 26/09/2013.

A inclusão do outro nas sociedades complexas

The Inclusion of Others in Complex Societies

João Paulo Rodrigues
Universidade Estadual de Londrina

Resumo

Para Habermas, a fonte da legitimação do direito moderno encontra-se no processo democrático da legislação, e esta recorre para o princípio da soberania do povo. Portanto, é preciso saber que direitos os cidadãos de uma sociedade complexa devem atribuir uns aos outros para a convivência mútua. *A inclusão do outro* e *Facticidade e Validade* apresentam o seguinte ponto: o desejo pela questão das conclusões do conteúdo universalista dos princípios republicanos, para as sociedades complexas. Primeiro, será apresentado o direito moderno e o seu papel de mediador da integração social entre o mundo da vida e o sistema, e que a legitimidade do direito moderno está fundada no princípio da democracia. Logo após, será explicado o significado de “inclusão do outro”, que busca o mesmo respeito por todos e traz à tona a responsabilidade da solidariedade universal entre os cidadãos, revelando que as fronteiras da comunidade estão abertas a todos.

Palavras-chave

Direito Moderno. Habermas. Inclusão do Outro.

Abstract

For Habermas the font of legitimation of modern law is in the democratic process of legislation, and this call to the principle of people's sovereignty. Therefore, is must know that laws the citizens of a complex society should assign to each other to the mutual coexistence. *The Inclusion of Other* and *Facticity and Validity* have the following point: the desire for the issue of the conclusions of universalistic content of republican principles, for the complex societies. First, will present the modern law and your paper of the mediator of social integration among world of life and system, and that the legitimacy of modern law is founded on the principle of democracy. Soon after, will be explained the meaning of "inclusion of other", seeking the same respect for everyone and brings out the responsibility of universal solidarity among citizens, revealing that the boundaries of the community are open to everyone.

Keywords

Modern Law. Habermas. Inclusion of Other.

Introdução

Habermas tem como objetivo apresentar uma teoria da sociedade com propósito prático, que vise a autonomia dos indivíduos e a emancipação da sociedade moderna, esta que foi desacoplada em mundo da vida e sistema e já não pode mais fundamentar o direito por meio da tradição cultural e sua racionalidade prática. Nessa sociedade moderna, aparecem dois tipos de racionalidade: a comunicativa, que substituiu a razão prática após o giro linguístico, utilizada pelos indivíduos localizados no mundo da vida; e a estratégica, utilizada pelos indivíduos que se encontram no sistema.

Diante dessa sociedade moderna, que vive dia após dia o risco de dissenso entre a racionalidade comunicativa, empregada pelos indivíduos do mundo da vida, e a racionalidade estratégica, utilizada pelos agentes situados no sistema, tenta-se resolver os seguintes problemas: Como é possível ocorrer uma integração social entre os cidadãos

que se utilizam desses dois tipos distintos de racionalidade? Quem estaria disposto a se responsabilizar por tal integração social? O que legitimaria esse possível responsável? A fonte da legitimação do direito moderno se encontra no processo democrático da legislação, e esta recorre, por sua vez, para o princípio da soberania do povo. Então, se a soberania do povo é quem vai garantir a legitimação do direito moderno, por meio da democracia, como garantir a inclusão de todos os cidadãos nas atuais sociedades pluralistas?

O objetivo principal deste trabalho é investigar se a teoria do discurso, por meio do direito moderno, consegue aplicar a inclusão de todos os cidadãos nas sociedades complexas e formular as intuições morais nas mesmas. Mas para que isso seja possível, é preciso, em um primeiro momento, mostrar como ocorre o processo de juridicização na sociedade moderna, por meio da teoria da sociedade de Habermas. Posteriormente, é necessário apresentar a especificidade do direito moderno, que é fazer com que os indivíduos que se encontram no mundo da vida e no sistema obedeçam à norma, tanto pela sua facticidade quanto pela sua validade e, depois, descobrir em que se funda a legitimação desse direito. Tal caminho será necessário para explicitar, finalmente, o significado do conceito “inclusão do outro”, mostrando que as fronteiras da comunidade estão abertas a todos, inclusive aos que são estranhos um ao outro e querem continuar sendo estranhos.

Teoria da sociedade

Habermas mostra, em seu livro *Teoria y Práxis* (1987a, p. 13), que seu objetivo sempre foi desenvolver uma teoria da sociedade com propósito prático, já que o próprio autor trata de início a modernidade¹ como um projeto inacabado². Esse seria o fio condutor que permitiria a evolução das variadas ramificações de seu pensamento. Seu desejo constitui também em contribuir para a realização das metas emancipatórias da modernidade.

Para que o objetivo de desenvolver uma teoria da sociedade seja alcançado, Habermas

1 - “Moderno é considerado o novo que rompe com a tradição” e “a modernidade é caracterizada justamente por esse espírito de uma ruptura irreparável com a tradição, que a distingue das “modernidades” que a precederam” (PINZANI, 2009, p. 115). “O conceito de modernização refere-se a um conjunto de processos cumulativos e de reforço mútuo: [...] à expansão dos direitos de participação política, das formas urbanas de vida e da formação escolar formal; à secularização de valores e normas etc. A teoria da modernização efetua sobre o conceito weberiano de “modernidade” uma abstração plena de consequências. Ela separa a modernidade de suas origens – a Europa dos tempos modernos – para estilizá-la em um padrão, neutralizado no tempo e no espaço, de processos de desenvolvimento social em geral” (HABERMAS, 2000, p. 5).

2 - “A modernidade permanece inconclusa porque ainda não contemplou inteiramente a institucionalização de suas estruturas normativas pós-convencionais. Isso demonstra que a modernidade é antes de tudo projeção, em longa medida, da reconstrução operada no âmbito da ontogênese da consciência moral que, no entanto, ainda não encontrou espaço para a sua institucionalização social” (BANNWART, 2008, p. 48).

procura reconstruir a razão prática de Kant por meio de sua teoria do agir comunicativo. Habermas concebe a razão prático-moral como um programa emancipatório, ou seja, como um projeto que visa à liberdade e à justiça. Segundo Velasco (2003, p. 20), Kant já havia dito que emancipar significa ter a capacidade de decidir por si mesmo (se tornar um ser autônomo), e o objetivo principal desse processo de apresentação é o de criar as condições para que o indivíduo possa exercer a sua capacidade efetiva sem qualquer impedimento. Já no plano político, o ponto de orientação normativa que Habermas apresenta é a de autogoverno, e seu objetivo seria a configuração de uma sociedade livre de dominação.

Para Habermas (2003, v. 1, p. 17), a modernidade acaba identificando a razão prática como uma faculdade subjetiva partindo da formação de um sujeito singular, tornando-se também uma razão de marca normativista, permitindo ao indivíduo moderno uma alternativa de solução aos problemas que norteiam sua comunidade e a si mesmo. Porém, acaba ocorrendo uma falta de leis provenientes da ausência de um conteúdo orientador da ação e de um patamar normativo. Habermas não concorda com esse horizonte e substitui, a partir do giro linguístico³, a razão prática pela razão comunicativa, fundando o conceito de racionalidade ao *medium* linguístico (MOREIRA, 2004, p. 100). Nas palavras de Habermas (2003, v. 1, p. 19): “eu resolvi encetar um caminho diferente, lançando mão da teoria do agir comunicativo: substituo a razão prática pela comunicativa. E tal mudança vai muito além de uma simples troca de etiqueta”. A razão comunicativa está inserida no *telos* do entendimento por meio do *medium* linguístico⁴ em que, a partir dos atos de linguagem reproduzidos comunicativamente às formas de vida, busca-se o entendimento com alguém sobre algo no mundo, ou seja, a ação comunicativa diz respeito aos atos ilocucionários⁵,

3 - Do original *linguistic turn*. “Giro linguístico foi a mudança de paradigma que ocorreu no pensamento filosófico ao longo do séc. XX. Aqui a linguagem deixa de ser um objeto de estudo entre outros e passa a ter uma referência inevitável e fundamental onde se abordam todos os problemas filosóficos. Razão e linguagem se tornam idênticos de tal modo que a linguagem se torna a única forma racional de se conhecer a realidade. Nossa relação com o mundo passa a ter um caráter simbolicamente mediado, visto que a linguagem desempenha um papel fundamental. A linguagem não é mais um meio de conhecimento, ela passa a ser a condição de possibilidade de conhecimento” (VELASCO, 2003, p. 171). A partir do momento em que ocorre a passagem do paradigma da filosofia da consciência para o paradigma da filosofia da linguagem, “os sinais linguísticos, que serviam apenas como instrumento e equipamento das representações, adquirem, como reino intermediário dos significados linguísticos, uma dignidade própria. As relações entre linguagem e mundo, entre proposição e estados de coisas, substituem as relações sujeito-objeto. O trabalho de constituição do mundo deixa de ser uma tarefa da subjetividade transcendental para se transformar em estruturas gramaticais. O trabalho reconstrutivo dos linguistas entra no lugar de uma introspecção de difícil controle. Pois, as regras, segundo as quais os signos são encadeados, as frases formadas e os enunciados produzidos, podem ser deduzidas de formações linguísticas que se apresentam como algo já existente” (HABERMAS, 1990, p. 15).

4 - “O que torna a razão comunicativa possível é o *medium linguístico*, através do qual as interações se interligam e as formas de vida se estruturam. Tal racionalidade está inscrita no *telos* linguístico do entendimento, formando um *ensemble* de condições possibilitadoras, e, ao mesmo, limitadoras” (HABERMAS, 2003, v. 1, p. 20).

5 - Atos locucionários, ilocucionários e perlocucionários são conceitos adotados por Austin e são entendidos da seguinte forma: ato locucionário é quando o ato de fala tem sentido e referência definidos, já que estes atos têm significado. Nas palavras de Austin (1990, p. 85) “esse ato de “dizer” algo nesta acepção normal e completa chamo de realização de um ato locucionário”, enquanto que ato ilocucionário acontece quando, ao dizer algo, realiza-se uma ação, pois estes têm força

visto que os participantes dessa ação comunicativa buscam fins ilocucionários.

Para fundamentar a validade dos enunciados e juízos morais, Habermas e Apel desenvolveram a “Ética do Discurso”, que seria uma extensão da ação comunicativa. A Ética do Discurso tem como finalidade separar a norma socialmente vigente da moralmente válida, pois em um discurso os indivíduos argumentam sobre normas e tentam verificar quais são moralmente corretas. Habermas mostra, em seu livro *Consciência Moral e Agir Comunicativo* (1989, p. 112), que a ética do discurso pretende revelar os pressupostos que tornam racional a argumentação, chegando assim às seguintes conclusões: durante o discurso, todas as pessoas capazes de se comunicar são interlocutores válidos, e seus interesses sobre normas devem ser levados em conta e defendidos por eles mesmos ou por outra pessoa.

Para que uma norma possa encontrar a aprovação de todos os participantes de um discurso prático, dois princípios devem ser aceitos: o *princípio da universalização ‘U’*, visto que, a partir das mencionadas regras do Discurso, uma norma controversa só poderá encontrar assentimento entre os participantes de um Discurso prático, se o princípio da universalidade ‘U’ for aceito: “Se as consequências e efeitos colaterais, que previsivelmente resultam de uma obediência *geral* da regra controversa para a satisfação dos interesses de *cada indivíduo*, podem ser aceitos *sem coação* por todos” (HABERMAS, 1989, p. 116).

Apesar de se ter mostrado como o princípio da universalidade pode ser fundamentado por meio da derivação pragmático-transcendental, a partir de pressuposições argumentativas, a própria *Ética do Discurso* pode ser reduzida ao *princípio do discurso ‘D’*, segundo o qual: “D: São válidas as normas de ação às quais todos os possíveis atingidos poderiam dar o seu assentimento, na qualidade de participantes de discursos racionais”⁶ (HABERMAS, 2003, v. 1, p. 142).

A modernidade apresenta também outras duas formas de razão: a razão instrumental, que tem um fundamento justificado nos fins pela ação dos meios, visto que as ações sociais

ilocucionária, ou seja, um ato ilocucionário é “a realização de um ato *ao* dizer algo, em oposição à realização de um ato *de* dizer algo” (AUSTIN, 1990, p. 89). Já nos atos perlocucionários o que interessa é obter certos efeitos ou consequências sobre os sentimentos, pensamentos, ou ações de alguém ao se dizer algo (AUSTIN, 1990, p. 89). Em resumo, distingue-se: “o ato locucionário [...] que tem um *significado*; o ato ilocucionário que tem uma certa *força* ao dizer algo; e o ato perlocucionário que consiste em se obter certos *efeitos* pelo fato de se dizer algo (AUSTIN, 1990, p. 103).

6 - “O predicado “válidas” refere-se a normas de ação e a proposições normativas gerais correspondentes; ele expressa um sentido não-específico de validade normativa, ainda indiferente em relação à distinção entre moralidade e legitimidade. Eu entendo por “normas de ação” expectativas de comportamento generalizadas temporal, social e objetivamente. Para mim, “atingido” é todo aquele cujos interesses serão afetados pelas prováveis consequências provocadas pela regulamentação de uma prática geral através de normas. E “discurso racional” é *toda* a tentativa de entendimento sobre pretensões de validade problemáticas, na medida em que ele se realiza sob condições da comunicação que permitem o movimento livre de temas e contribuições, informações e argumentos no interior de um espaço público constituído através de obrigações ilocucionárias. Indiretamente a expressão refere-se também a negociações, na medida em que estas são reguladas através de procedimentos fundamentados discursivamente” (HABERMAS, 2003, v. 1, p. 142).

individualizadas são mediadas por interesses subjetivos, tomando uma visão utilitarista; e a razão estratégica, que se funda na orientação da ação para o êxito, a partir de uma avaliação das condições dadas. Exemplo: o êxito da Economia é medido pelo meio dinheiro e o êxito da Política é medido pelo meio poder⁷.

Então, enquanto a linguagem natural for utilizada como fonte de integração social, trata-se de ação comunicativa; agora, caso a linguagem seja utilizada como meio para a transmissão de informações, trata-se aqui de ação estratégica. Ora, na ação estratégica, o efeito da coordenação das ações irá depender da influência dos atores uns sobre os outros e por meio da situação das ações, transmitidas sob atividades não-linguísticas (HABERMAS, 1990, p. 71). Em outras palavras, a ação estratégica se apoia na razão teleológica de planos particulares de ação; já a ação comunicativa encontra a sua força na razão motivadora de atos de entendimento comunicativo.

Ao pensar em uma teoria da sociedade, Habermas propõe uma sociologia no lugar de uma antropologia. Para que haja uma realização satisfatória da ação comunicativa, é necessário que os interlocutores compartilhem as mesmas experiências e vivências “pré-reflexivas” na qual se adote um sentido a tudo que se diz. Habermas chama isso de *Mundo da Vida*. Esse conceito se refere “ao ambiente imediato do agente individual, o ambiente simbólico e cultural que forma a camada profunda de evidências, certezas e realidades que não são normalmente colocadas em questão” (VELASCO, 2003, p. 47). É nesse horizonte comum de compreensão que os sujeitos podem atuar de modo comunicativo.

Contrário ao *mundo da vida* está o *sistema*. O sistema tem um equilíbrio que se autorregula por meio da especificação funcional dos diferentes subsistemas que apareceram após o desacoplamento presente na teoria da sociedade de Habermas. No sistema, as ações de cada indivíduo são determinadas por cálculos de interesse, que maximizam a utilidade. É, também, um conjunto social formado por diversos mecanismos anônimos dotados de lógica própria que, na sociedade moderna, se cristalizou em subsistemas sociais diferenciados e regidos por regras estratégicas, e por meios materiais ou técnicos: o subsistema Estado e o subsistema Economia (VELASCO, 2003, p. 48). Entretanto, na sociedade moderna, o sistema acaba sendo o elemento que mais se expande, e cabe observar uma mediação do sistema no mundo da vida; Habermas chamará este fenômeno de *colonização do mundo*

7 - “A ação instrumental orienta-se por *regras técnicas* que se apoiam no saber empírico. Estas regras implicam em cada caso prognoses sobre eventos observáveis, físicos ou sociais; tais prognoses podem revelar-se verdadeiras ou falsas. O comportamento da escolha racional orienta-se por *estratégias* que se baseiam num saber analítico. Implicam deduções de regras de preferência (sistemas de valores) e máximas gerais; estas proposições estão deduzidas de um modo correto ou falso. A ação racional teleológica realiza fins definidos sob condições dadas; mas, enquanto a ação instrumental organiza meios que são adequados ou inadequados segundo critérios de um controle eficiente da realidade, a ação estratégica depende apenas de uma valoração correta de possíveis alternativas de comportamento, que só pode obter-se de uma dedução feita com o auxílio de valores e máximas” (HABERMAS, 1987b, p. 57).

da vida. Isso ocorre quando a reprodução simbólica do mundo da vida começa a se fundar sobre a base do sistema, o que acaba trazendo danos para o Estado social, pois o mundo da vida se reproduz quando se apreendem as condições de uma ação formalmente organizada e entendida como relação expressada pelo direito (MOREIRA, 2004, p. 53).

Há uma preocupação dessa colonização quanto aos recursos próprios do sistema, tais como o dinheiro e o poder, que se introduzem no mundo da vida e acabam substituindo a comunicação entre os sujeitos por meios não verbais, monetarizando e burocratizando as relações humanas. Tal colonização do mundo da vida faz com que o direito também se instrumentalize. Esse direito instrumentalizado se utiliza, então, da racionalidade estratégica para estabilizar as expectativas de comportamento para uma dominação legal dos indivíduos do mundo da vida. Assim, a solidariedade perde o seu papel de integração social, pois o dinheiro, o poder e o direito instrumentalizado acabam tomando o lugar de mediador que antes era da solidariedade, fazendo com que a interação social seja feita pela integração funcional da racionalidade estratégica (DURÃO, 2006, p. 104).

Legitimação do direito moderno

Apartir disso, Habermas tenta fortalecer o mundo da vida por meio da ação comunicativa. Mas como esclarecer a possibilidade de reprodução social nesta sociedade moderna em um fundamento frágil como é o das pretensões de validade transcendentais? Pode existir ordem social partindo dos processos de formação de consenso sempre ameaçados por uma tensão entre facticidade e validade das normas? Como é possível a integração social na sociedade moderna por meio da solidariedade da racionalidade comunicativa? Aqui o direito moderno⁸ aparece como possível candidato para resolver isso, visto que Habermas, partindo da teoria do discurso fundamentada na racionalidade comunicativa, alterará em *Facticidade e Validade* o papel do direito, formulando uma nova teoria do direito, tentando assimilar a tensão entre facticidade e validade das normas (MOREIRA, 2004 p. 103).

Para assimilar a tensão entre facticidade e validade das normas jurídicas na sociedade

8 - Portanto, é preciso saber qual é o tipo de direito que interessa para Habermas. Seria o direito: (1) como uma manifestação do direito temporalmente limitada e condicionada, ou seja, o direito moderno; (2) determina o tipo de direito como sendo direito normatizado ou positivo; (3) como uma ordem normativa que é justificada “somente apelando para um sistema coerente que possibilita a produção de normas segundo um procedimento exatamente determinado por regras precisas” (PINZANI, 2009, p. 141); (4) o direito moderno se depara com normas jurídicas interpretadas apenas por uma instância autorizada a fazê-lo e da qual sua interpretação é vinculante; e (5) as normas positivas do direito moderno são caracterizadas por serem instaladas por uma instância legítima ordenada da força necessária. Resumindo, Habermas só entende o Direito em seu nível pós-convencional, ou seja, no direito moderno, no qual “as estruturas da consciência moderna materializam-se no sistema jurídico” (MOREIRA, 2004, p. 36).

moderna, Habermas formulará uma nova teoria discursiva do direito, fundamentada na razão comunicativa, para que possa ocorrer uma integração social entre os indivíduos situados no mundo da vida e no sistema. Então, ao se utilizar da razão estratégica, os agentes situados no *sistema* passam a obedecer à lei por temor da coerção, ou seja, pela facticidade, ou então, ao fazer uso da razão comunicativa, os atores agem motivados pelo puro respeito à lei, já que estão convencidos de sua validade. Desse modo, o direito moderno funciona como transformador linguístico, pois traduz a linguagem da razão estratégica usada pelos indivíduos situados no sistema para a linguagem da razão comunicativa dos indivíduos do mundo da vida e vice-versa. Assim, o direito moderno desenvolve uma tensão entre facticidade e validade para se realizar a função de integração social na sociedade moderna (DURÃO, 2006, p. 105).

A solidariedade desempenha um papel central para o direito, como um consenso de valores distribuídos intersubjetivamente que orientam as ações dos indivíduos. De acordo com Pinzani (2009, p. 146), tal solidariedade só é possível por meio de costumes, lealdades e confiança recíproca, no qual são solucionados os conflitos que aparecem na interação social. Porém, a solidariedade é apenas um dos três recursos utilizados na integração social; os outros dois são o dinheiro e o poder.

Sociedades modernas são integradas não somente através de valores, normas e processos de entendimento, mas também sistematicamente, através de mercados e do poder administrativo. Dinheiro e poder administrativo constituem mecanismos da integração social, formadores de sistema, que coordenam as ações de forma objetiva, como que por trás das costas dos participantes da interação, portanto não necessariamente através da sua consciência intencional ou comunicativa (HABERMAS, 2003, v. 1, p.61).

Aqui surge, mais uma vez, a tensão entre os indivíduos situados no mundo da vida, que usam a solidariedade como integração, e no sistema, que, por sua vez, usam o dinheiro e o poder como integração. Porém, a solidariedade acaba sendo o meio de integração social mais fraco, após a colonização do mundo da vida. Assim, acaba ocorrendo, na sociedade complexa, uma falta de solidariedade que só pode ser ocupada pelo direito.

A força social e integradora da solidariedade, que não pode ser extraída apenas de fontes do agir comunicativo, deve desenvolver-se através de um amplo leque

de esferas públicas autônomas e de processos de formação democrática da opinião e da vontade, institucionalizados através de uma constituição, e atingir os outros mecanismos da integração social – o dinheiro e o poder administrativo – através do *medium* do direito (HABERMAS, 2003, v. 2, p. 22).

O direito moderno, “em reação ao processo de racionalização característico da modernidade” (PINZANI, 2009, p. 146), acaba recebendo uma dupla função no âmbito dessa sociedade complexa. Primeiramente, o direito, “através de uma prática de autodeterminação, que exige dos cidadãos o exercício comum de suas liberdades comunicativas” (HABERMAS, 2003, v. 1, p. 62), deve assegurar a solidariedade social na sociedade complexa. Com isso, os conflitos acabam sendo resolvidos, agora juridicamente, e não mais pela ética, como era antigamente. Isso acontece pelo fato do consenso só poder ser obtido por meio de procedimentos regularizados juridicamente⁹.

Por segundo, o direito possui a tarefa de se opor à colonização do mundo da vida, funcionando como um objeto que une os agentes do mundo da vida e do sistema e acaba se tornando “uma correia de transmissão abstrata e obrigatória, através do qual é possível passar solidariedade para as condições anônimas e sistematicamente mediadas de uma sociedade complexa” (HABERMAS, 2003, v. 1, p. 107). Essa dupla função do direito traz consigo consequências políticas, pois a universalização de indivíduos institucionalizados pública e juridicamente forma o complemento fundamental para a juridificação da integração social cujo núcleo de tal cidadania é constituído pelos direitos de participação política que só podem acontecer em uma democracia (HABERMAS, 2003, v. 1, p. 105).

O direito moderno acaba sendo o único instrumento capaz de resolver os riscos de dissenso entre os indivíduos situados no mundo da vida e no sistema após a colonização do mundo da vida, pois tal direito é somente legítimo quando ocorrer a democracia, esta que terá a função de reduzir a complexidade social. A democracia é, então, “a única forma que uma ordem jurídica legítima pode tomar”, pois “não há direito democrático sem democracia” (PINZANI, 2009, p. 147).

Agora, para apresentar a legitimação do direito moderno, Habermas parte da própria legalidade kantiana (normas do direito são, ao mesmo tempo e sob aspectos diferentes, leis da coerção e leis da liberdade), fundada no princípio da democracia (institucionalização

⁹ - “Solidariedade é formalizada em uma espécie de lealdade procedimental que supera a fraqueza motivacional dominante nas modernas sociedades secularizadas, já que sem a retaguarda de cosmovisões metafísicas ou religiosas imunes à crítica, as orientações práticas só podem ser obtidas em última instância, por meio de argumentações, e estas últimas precisam de regras rígidas” (PINZANI, 2009, p. 146-7).

jurídica do princípio do discurso 'D'), pois, ao empregar a racionalidade comunicativa ao direito, Habermas acaba construindo uma teoria discursiva do direito.

O projeto de uma teoria discursiva do direito se orienta pelas seguintes perspectivas: primeiro, rompe-se com a razão prática à medida que a razão comunicativa não aparece como informativa para a ação e não tem como recurso uma esfera normativa; segundo, seria a institucionalização do direito por meio de um procedimento que vem da relação recíproca entre direitos humanos e soberania popular; terceiro, a relação entre direito e moral como uma relação de co-originariedade (MOREIRA, 2004, p. 138).

Normas morais regulam relações interpessoais e conflitos entre pessoas naturais, que se reconhecem reciprocamente como membros de uma comunidade concreta e, ao mesmo tempo, como indivíduos insubstituíveis. Eles se dirigem a pessoas individuadas através de sua história de vida. Ao passo que normas jurídicas regulam relações interpessoais e conflitos entre atores que se reconhecem como membros de uma comunidade abstrata, criada através das normas do direito (HABERMAS, 2003, v. 1, p. 147).

Para Habermas, as normas jurídicas e as normas morais são co-origenárias na medida em que uma não é legisladora para a outra, ou seja, não se pode fundamentar uma ao se normatizar a outra (MOREIRA, 2004, p. 148). Porém, apesar de todos os pontos em comum, Habermas (2003, v. 1, p. 141) diz que a moral e o direito diferem-se *prima facie*, pois a moral na sociedade moderna representa, somente, uma forma do saber cultural. Já o direito toma para si obrigatoriedade também no nível institucional, mostrando que o direito é, ao mesmo tempo, um sistema de ação e não somente um sistema de símbolos. Mas, uma ordem jurídica só pode ser legítima quando não contrariar princípios morais.

O princípio da democracia tem a função de amarrar um procedimento que normatize legitimamente o direito moderno. O princípio da democracia é a institucionalização jurídica do princípio do discurso 'D'.

Ele [o princípio da democracia] significa, com efeito, que somente podem pretender validade legítima as leis jurídicas capazes de encontrar o assentimento de todos os parceiros do direito num processo jurídico de normatização discursiva. O princípio da democracia explica, noutros termos,

o sentido performativo da prática de autodeterminação de membros do direito que se reconhecem mutuamente como membros iguais e livres de uma associação estabelecida livremente (HABERMAS, 2003, v. 1, p. 145).

Sendo assim, o princípio da democracia tem a finalidade de orientar a produção do próprio *medium* do direito, para que, além de institucionalizar não apenas uma formação política racional, possa também proporcionar o *medium* jurídico legítimo em que se expresse uma vontade intersubjetiva de sujeitos autônomos de Direito. O vínculo interno entre Estado de direito e democracia pode ser explicado a partir da dependência mútua entre a autonomia privada (que trata do direito privado) e a autonomia pública do cidadão (soberania do povo). Ao defender um vínculo interno entre autonomia privada e autonomia pública, a Ética do Discurso tem como pretensão proporcionar uma justificação do Estado de direito democrático nas quais direitos humanos e soberania popular desempenham papéis singulares, mas também complementares.

Habermas acaba defendendo então um *paradigma jurídico procedimentalista* do direito, pois mostra que o direito só será legítimo, podendo exercer por completo a sua função de integração social, quando seus destinatários forem ao mesmo tempo os criadores das normas de uma ordem jurídica. Pode-se observar então que no paradigma jurídico procedimentalista do direito a sociedade possui não apenas processos produtivos, tal sociedade possui também processos comunicativos (PINZANI, 2009, p. 147-8). Agora resta saber: o que fundamenta o direito moderno para Habermas? Como o direito moderno se torna normativo? O processo de normatização do direito moderno é obtido a partir do processo democrático da legislação, e esta recorre, por sua vez, para o princípio da soberania do povo.

Os membros de uma dada comunidade jurídica têm de se atribuir direitos para que possam se constituir membros de uma comunidade jurídica autônoma. Assim, a ideia de que o ordenamento jurídico se constitui enquanto uma instância externa aos cidadãos, heterônoma, cede lugar à ideia de uma produção efetiva de seres livres que têm, no ordenamento jurídico, a manifestação de sua vontade livre, ou seja, o Direito é, ao mesmo tempo, criação e reflexo da produção discursiva da opinião e da vontade dos membros de uma dada comunidade jurídica (MOREIRA, 2004, p. 138-9).

Será por meio do princípio da democracia que Habermas fundamentará o Direito,

estabilizando a tensão entre autonomia privada e autonomia pública, que se dá através do procedimento legislativo. Desse modo, a apresentação da co-originariedade da autonomia privada e autonomia pública só aparece quando a soberania do povo assume uma figura jurídica, ou seja, quando se decifra o modelo de auto-legislação proveniente da teoria do discurso, que ensina os destinatários a serem, ao mesmo tempo, os autores de seus direitos. Aqui a soberania do povo assume figura jurídica, pois a substância dos direitos humanos introduz-se “nas condições formais para a institucionalização jurídica desse tipo de formação discursiva da opinião e da vontade, na qual a soberania do povo assume figura jurídica” (MOREIRA, 2004, p. 139).

Ora, se para Habermas (2003, v. 1, p. 114) “o direito moderno tira dos indivíduos o fardo das normas morais e as transfere para as leis que garantem a compatibilidade das liberdades de ação”, tais leis obtêm sua legitimidade por meio de um processo legislativo que, por sua vez, se apoia no princípio da soberania do povo, fazendo com que se chegue à conclusão de que o processo democrático da criação do direito constitui a única fonte pós-metafísica da legitimidade. Mas de onde o direito retira sua legitimidade? A teoria do discurso apresenta uma resposta simples, apesar de ser pouco provável em um primeiro momento: o processo democrático, possibilitador de uma livre flutuação de temas e de contribuições, de informações e de argumentos, acaba por assegurar um caráter discursivo à formação política da vontade, estabelecendo assim, “a suposição falibilista de que os resultados obtidos de acordo com esse procedimento são mais ou menos racionais. *Prima facie* há duas considerações que falam a favor do princípio da teoria do discurso” (HABERMAS, 2003, v. 2, p. 308).

Tanto mais legítimo será o Direito quanto mais preservar o espaço da liberdade privada. A autonomia do cidadão se mede através da liberdade negativa que cabe a cada indivíduo. E as prescrições só são legítimas à medida que são a justaposição do arbítrio dos contratantes. Sendo assim, as prescrições do ordenamento jurídico só são válidas quando as partes receptoras tiverem seu espaço de liberdade, sua autonomia, preservados. No momento em que a normatividade jurídica invade o espaço de liberdade do indivíduo, cessa a legitimidade da produção normativa (MOREIRA, 2004, p.144).

Na medida em que a comunidade jurídica se constitui por meio de um entendimento obtido pelo discurso, eles saem do campo de destinatários para entrar no campo de coautores

da normatividade que vem do direito moderno. Porém, apesar dos cidadãos serem autores do sistema jurídico, tal produção discursiva precisa de um processo de institucionalização. Sendo assim, “a normatividade do Direito não é fechada sobre si mesma, antes precisa comprovar-se na fatualidade das decisões democráticas” (MOREIRA, 2004, p. 145), visto que tal normatividade emana discursivamente da vontade dos cidadãos. Em outras palavras, é normativa apenas quando gera obrigatoriedade.

Portanto, é somente com a juridificação da liberdade comunicativa que se torna possível realizar, legitimamente, o direito e a democracia. A integração social toma forma totalmente reflexiva a partir do momento em que o direito moderno serve-se do risco permanente de dissenso, dado que o mesmo preenche a sua cota de legitimação com a ajuda da força produtiva da comunicação, fazendo com que ele se transforme num incentivo capaz de dar vida aos discursos públicos institucionalizados juridicamente, e faz com que a liberdade possa ser finalmente realizada.

A inclusão do outro nas sociedades complexas

Para finalizar, será demonstrado então que Habermas defende o conteúdo racional de uma moral que tem por base o respeito mútuo e uma responsabilidade pela solidariedade recíproca. É claro que a modernidade fica desconfiada de tal universalidade, que assimila e iguala todos os indivíduos, e que não compreende o sentido dessa moral, fazendo com que desapareça a relação existente entre a alteridade e a diferença, esta que continua tendo a sua validade por um universalismo bem compreendido. Habermas formula, na *Teoria do Agir Comunicativo*, os princípios básicos dessa moral, de um jeito que os mesmos formassem boas condições de vida que quebrassem a alternativa entre “comunidade” e “sociedade”, visto por Habermas como se fosse algo falso.

O livro *A inclusão do outro*, do autor Jürgen Habermas, surgiu depois da publicação de *Facticidade e Validade*. O que esses dois livros têm em comum seria “o interesse pela questão das consequências que hoje resultam do conteúdo universalista dos princípios republicanos” (HABERMAS, 2002, p. 7), mais especificamente: para as sociedades complexas, no qual os contrastes multiculturais se tornam questões urgentes; para os estados nacionais, que se ligam em unidades supranacionais; “e para os cidadãos de uma sociedade mundial que foram reunidos numa involuntária comunidade de risco, sem ter sido consultados” (HABERMAS, 2002, p. 7).

Conforme Velasco (2003, p. 123-4), Habermas pretende estabelecer uma distinção

entre os elementos que configuram a *cultura política* e as diversas *formas de vida* que indivíduos livremente podem abraçar. Trata-se aqui de evitar que a definição de identidade coletiva acabe sendo utilizada como mecanismo de exclusão do diferente e se torne, como acontece com certa facilidade, uma vontade consciente de homogeneidade que provoque a marginalização interna de grupos sociais inteiros. Daí sairia a sua convicção de que, para resolver esse problema, as atuações políticas próprias de uma democracia devem se dirigir para a “inclusão do outro”, de certa maneira que, a partir da independência da procedência cultural de cada um, as vias de acesso da comunidade política sempre permaneçam abertas. Mas, para que isso ocorra, é indispensável que as instituições públicas sejam desprovidas, em seu maior grau possível, de conotações morais densas e adotem características procedimentais do direito moderno que garantam a neutralidade. Leva-se isso em consideração, pois, em um Estado constitucional democrático, a maioria não pode prescrever para as minorias a própria forma de vida cultural, na medida em que divirja da cultura política comum do país, como sendo a cultura dominante. Sendo assim, somente por meio das próprias instituições dessa forma de Estado, seria possível estabelecer relações de respeito mútuo entre sujeitos com diferentes bagagens socioculturais.

Há uma correspondência entre a teoria da sociedade e a teoria da moral e do direito, que seria um universalismo que possui uma fragilidade quanto às diferenças. Tal respeito mútuo (para *todos e cada um*) não é apenas para aqueles que são semelhantes, mas também à pessoa do outro ou dos outros em suas diferenças (alteridade). A responsabilidade pela solidariedade com o outro, como sendo uma pessoa igual a nós, refere-se a um “nós” flexível em uma comunidade que se conserva firmemente a tudo o que é material e possui uma amplitude constante de suas fáceis fronteiras. A constituição dessa comunidade moral é feita basicamente da ideia negativa do fim do preconceito e do sofrimento, além da inclusão de todos os marginalizados em uma relação de respeito recíproco. Porém, tal comunidade não pode ser considerada um coletivo que impõe a obrigação de que todos os indivíduos afirmem a índole própria de cada um. Habermas apresenta o conceito de inclusão não como um aprisionamento dentro de si mesmo e um bloqueio frente ao alheio, mas sim, a “inclusão do outro” diz respeito ao significado de que as fronteiras da comunidade estão abertas a todos os indivíduos, principalmente para os indivíduos que são diferentes aos olhos dos outros e desejam continuar sendo diferentes (HABERMAS, 2002, p. 7-8).

Após os mandamentos morais não serem mais legitimadas pela posição divina, elas deixam de ser justificadas soteriológicamente e ontoteologicamente. Sendo assim, a Ética do Discurso também tem um preço a pagar com isso, já que a mesma não pode mais manter o teor moral íntegro das instituições religiosas e não pode mais guardar o sentido realista de validação inerente às normas morais. Ora, se a práxis moral deixa de se orientar por meio

das instituições religiosas, acabam derivando disso duas ruins consequências: primeiro, o saber moral se solta dos motivos subjetivos da ação; segundo, o que é compreendido por moral correta transforma-se em algo diverso da concepção de um bem-viver apresentado pelas instituições religiosas (HABERMAS, 2002, p. 47-8).

A ética do discurso, conforme Habermas (2002, p. 48), ordena argumentações de autoentendimento e argumentações de fundamentação normativa (ou de aplicação). Porém, a mesma não reduz a um tratamento confuso a moral, já que pretende dar mérito à justiça e à solidariedade. O acordo, alcançado por meio do discurso, é dependente do “sim” ou do “não” de todos os participantes, além de ser preciso superar o egocentrismo, pois uma práxis argumentativa pretende se regular pelo convencimento recíproco. Todavia, quando a moral se desprende às questões da vida bem-sucedida, ela acaba por possuir também seu lado motivacional. Ora, se não existe nenhuma punição para quem for contra a moral religiosa, não há mais um motivo para seguir os mandamentos morais. A partir do momento em que os discursos racionais obtêm novamente o seu fundamento, por meio do ponto de vista moral, a ética do discurso irá forçar a separação intelectualista entre juízo moral e ação.

Aqui surge a necessidade de se complementar a moral com um direito coercitivo e positivo, como foi visto anteriormente. Segundo Habermas (2002, p. 49), com a apresentação da diferença entre o dever e o que é almejado pela ética, o dever consegue a sua validade e se torna normatividade. Mudando a perspectiva de Deus para o homem, pode-se compreender agora por “validade” que as normas morais obterão o assentimento de todos os participantes do discurso prático, que testarão em conjunto a possibilidade da respectiva práxis ser de interesse mútuo. No momento em que se encontra esse assentimento, fica em destaque a razão falível dos sujeitos em conselho, já que são convencidos reciprocamente de que encontraram o devido reconhecimento da norma, e a liberdade dos indivíduos que legislam, pois são compreendidos, ao mesmo tempo, como autores e destinatários das normas, reconhecendo, assim, tanto a falha do espírito humano quanto à capacidade de construção que o mesmo projeta.

Os posicionamentos morais revelam a sua validade quando se traz à memória as afirmações fundamentais com que os mandamentos morais obtiveram justificação, por meio da religião, estabelecendo um mundo racional. Sabe-se que, se o realismo moral não se defende mais pela religião e pelo direito natural, “a validação da obrigatoriedade de asserções morais não pode mais ser assimilada pela validação da verdade de asserções descritivas. Alguns dizem como são as coisas no mundo, outros dizem o que devemos fazer” (HABERMAS, 2002, p. 49).

O teor de uma moral do respeito sem distinções e da responsabilidade solidária por

cada um é justificada por Habermas (2002, p. 53) por meio da ética do discurso, sendo somente alcançada com a reconstrução racional dos conteúdos de uma tradição moral religiosa. Caso o princípio do discurso 'D' conservasse sua ligação com essa tradição da origem, tal genealogia iria se intercalar ao objetivo de justificar o teor cognitivo dos juízos morais em toda a sua totalidade. Portanto, é preciso fundamentar também o próprio ponto de vista moral, por meio da teoria moral. O princípio do discurso 'D' é a saída da coação, que se faz presente nos participantes de certas comunidades morais quando os mesmos, transferidos para as sociedades pluralistas, são atraídos no dilema de continuar discursando sobre juízos e posicionamentos morais portando seu tipo de razão, tal como faziam antes. As pessoas, em nível global ou local, participam de conflitos de conduta que elas mesmas, ainda que seu *ethos* não tenha mais valor, compreendem como conflitos morais, e que são solucionados partindo de alguma fundamentação. Será visualizado agora um percurso estilizado de maneira ideal e tipificada, demonstrando como tal moral ocorreria sob condições reais.

Habermas (2002, p. 53) parte da ideia de que os participantes pretendam resolver seus conflitos sem a utilização de violência ou acertos ao acaso, por meio de um acordo recíproco. Seria proposto, agora, determinar uma auto-compreensão ética que fosse aceita universalmente, porém, encontraria seu fracasso devido às condições de uma sociedade pluralista. Apesar de suas fortes convicções valorativas serem certificadas por meio da crítica preservada na prática, os participantes percebem que existem concepções diferentes sobre o que seja o bem. Mas ainda será suposto que tais participantes queiram, mesmo assim, chegar a um acordo recíproco, sem querer substituir o convívio moral, que já é ameaçado por acordos entre partes cujas opiniões diferem. Frente à fragilidade de um acordo fraco sobre os conteúdos das normas, os participantes percebem que estão abandonados a uma circunstância qualquer, que ainda assim é comunicativa e obtêm a sua estrutura por meio de um acordo mútuo linguístico. Há aqui certos aspectos estruturais em comum, o que faz com que os participantes possam se questionar sobre a existência ou não de conteúdos normativos que forneçam uma base para orientações universais (HABERMAS, 2002, p. 53-4).

As formas de vida comunicativas bem qualificadas não são capazes de fundamentar o porquê da violação de uma orientação particular de valores de determinados indivíduos, muito menos do porquê de os mesmos deverem fazer parte de relações de reconhecimento igual, ilimitadamente inclusivas no contexto de uma igualdade universal. Porém, uma posição universal, que tem o interesse em evitar falsas abstrações, necessita aproveitar discernimentos favorecidos por meio da teoria do discurso. O indivíduo somente adquire a sua individualidade quando se integra a uma sociedade, resultando na aprovação de uma moral válida tanto para o indivíduo irrepresentável quanto para quem faz parte da sociedade,

ligando a justiça com a solidariedade, o que faz com que sejam tratados com igualdade os desiguais, esses que são conscientes de um pertencimento em comum na sociedade. Agora, o aspecto conforme o qual todas as pessoas são iguais não encontra a sua validade a partir de outro aspecto, como se os mesmos fossem absolutamente diferentes de todos os outros. “O respeito reciprocamente equânime por cada um, exigido pelo universalismo sensível a diversificações, é do tipo de uma inclusão *não-niveladora e não-apreensória* do outro *em sua alteridade*” (HABERMAS, 2002, p. 55).

Surge, assim, a necessidade de se justificar a passagem para uma moral pós-convencional. As obrigações que têm por base a ação comunicativa e se ajustam pela tradição não vão, através delas mesmas, para fora da comunidade. Mas, para Habermas (2002, p. 55), a forma reflexiva da ação comunicativa é diferente, já que argumentações apontam por si sós para além de toda individualidade. Ora, isso é possível levando em consideração que nos pressupostos programáticos de discursos racionais, o teor normativo de suposições que fazem parte da ação comunicativa é ampliado a uma comunidade que insere todos, não excluindo ninguém que possua a capacidade de dar contribuições relevantes. Aqui aparece a saída na qual os participantes perderam o suporte ontoteológico e necessitam criar as próprias orientações normativas através dos próprios. Parte-se da ideia de que os participantes poderão recorrer apenas às coisas que possuam em comum, essas que se reduzem à provisão de qualidades formais presentes na situação de conselho e que são distribuídas pelos participantes de maneira performativa, abrangendo todos os participantes no empreendimento cooperativo por meio de um conselho que se encontra na prática.

Verificando a fragilidade dessa base, percebe-se que o conteúdo neutro de sua subsistência comum representa, ao mesmo tempo, uma chance, tendo em vista o constrangimento que surge pelo fato do pluralismo de cosmovisões. É necessário achar uma fundamentação conteudística-tradicional de um comum acordo normativo básico, se o mesmo tipo de comunicação que está de acordo com as reflexões práticas comuns resultasse em certo aspecto no qual haveria a possibilidade de fundamentar normas morais e no qual haveria a necessidade de ser convincente para todos os participantes. A carência desse bem universal encontraria sua superação de forma permanente apenas por meio do caráter próprio da práxis de reuniões em conselho. Habermas dá três passos para se alcançar uma fundamentação do ponto de vista moral, a partir da teoria moral (HABERMAS, 2002, p. 55-6).

Primeiro passo: partindo da ideia que o único expediente possível para o ponto de vista do julgamento imparcial de questões morais seria a práxis de reuniões em conselho, aparece a necessidade de se alterar a referência a conteúdos morais pela referência autorremissiva

à forma dessa práxis. Será por meio da compreensão dessa situação que o princípio do discurso ‘D’ aparecerá. Assim, o acordo conquistado pela base das condições discursivas é compreendido por um comum acordo aprovado por razões epistêmicas, ou seja, não se compreende esse acordo como um acerto ocasional motivado por meio de uma visão egocêntrica racional. Porém, ‘D’ abre um caminho no qual há todo o tipo de argumentação que tem por objetivo o comum acordo discursivo. “Com ‘D’ não se supõe de saída que uma fundamentação de normas morais seja sequer possível fora do contexto de um acordo substancial” (HABERMAS, 2002, p. 56).

Segundo passo: quando se introduz ‘D’ de forma condicional, ele acaba por apresentar a própria condição a ser seguida por normas válidas, desde que as mesmas encontrem a possibilidade de serem fundamentadas. Deve-se, então, explicar o conceito de norma moral. Os participantes de um discurso sabem intuitivamente como tomar parte em argumentações e, apesar de terem familiaridade apenas com a fundamentação de sentenças assertivas, e não tomarem conhecimento se as reivindicações de validação moral são possíveis de serem julgadas de modo igual, estão dispostos a pensar o que seria fundamentar normas. Porém, para se utilizar ‘D’, é necessária uma regra para a argumentação que aponte como podem se fundamentar as normas morais (HABERMAS, 2002, p. 56).

Por fim, o terceiro passo seria o de que os próprios participantes se dão por satisfeitos com tal regra de argumentação, desde que a mesma se mostre útil e não os levem a resultados que não sejam obtidas por meio da intuição. Há aqui a necessidade de mostrar que normas, que possuam a capacidade de conquistar concordância geral (Direitos Humanos), estão marcadas por uma práxis fundadora orientada assim mesma. Então, falta apenas um último passo fundador (HABERMAS, 2002, p. 57).

A linguagem é encontrada em todas as culturas e sociedades, e não há nenhum outro tipo de solução de problemas igual a esse. Sendo assim, essa difusão universal da linguagem e a falta de uma outra saída para ela, não seria fácil achar uma contestação à neutralidade de ‘D’. Porém, a partir da abdução de ‘U’, pode haver, mesmo que de forma camuflada, uma compreensão prévia etnocêntrica não distribuída por outras culturas, juntamente com uma concepção do que é bom. Quando a suposição de que um comprometimento eurocêntrico, que entende uma moral operacionalizada por ‘U’, poderia perder força caso existisse a possibilidade de tornar permanente a explicação para o ponto de vista moral, dito de outro modo, caso esse ponto de vista moral tivesse o poder de explicar sobre o que deve ser feito quando o participante se envolve em uma práxis argumentativa (HABERMAS, 2002, p. 57-8). Habermas (2002, p. 58) se dá por satisfeito com o seguinte vestígio fenomenológico: a argumentação acontece por meio da intenção de um convencimento mútuo, no que

diz respeito à legitimação das reivindicações de validação que os participantes mostram e defendem a favor deles. Com a prática da argumentação, começa a se instalar, em cooperação recíproca, uma concorrência por argumentos melhores, partindo da união dos participantes desde o início, que se orientam por um acordo recíproco. Ora, a concorrência, que pode levar a resultados racionalmente aceitáveis e convincentes, se fundamenta sobre a força de convencimento dos próprios argumentos, além do que, um argumento bom ou ruim pode ser colocado em discussão. Assim, uma afirmação aceita racionalmente encontra a sua base sobre razões ligadas a certas características do mesmo processo de argumentação (pressuposições pragmáticas) presente na ética do discurso.

Se cada um que se envolver em uma argumentação tiver que fazer ao menos essas pressuposições pragmáticas, então nos discursos práticos, (a) por causa do caráter público e inserção de todos os envolvidos e (b) por causa da igualdade de direitos de comunicação para todos os participantes, só poderão ter espaço as razões que levem em conta, de forma equânime, os interesses e as orientações de valor de cada um; e por causa da ausência de (c) engano e (d) coação, só poderão ser decisivas as razões para o assentimento de uma norma discutível. Por fim, sob a premissa de uma orientação segundo o acordo mútuo, presumida reciprocamente em todos os envolvidos, essa aceitação “não coativa” só pode dar-se “em comum” (HABERMAS, 2002, p. 58-9).

Em sentido moral, por enquanto, não há uma normatividade no teor dos pressupostos argumentativos gerais, tendo em vista que a possível questão da inserção é entendida por meio da condição de acesso irrestrito ao discurso, retirando a possibilidade de uma norma de ação vinculativa universal. As liberdades de comunicação no discurso, distribuídas igualmente, e os discursos favorecidos pela obrigação da sinceridade são compreendidos como deveres e direitos argumentativos morais. Além do que, o mesmo processo argumentativo se deve à falta de coação. As regras que fazem parte do jogo argumentativo estabelecem “o intercâmbio de argumentos e de posicionamentos de “sim”/“não”; elas têm o sentido epistêmico de possibilitar a justificação de asserções, e não o sentido prático imediato de motivar ações” (HABERMAS, 2002, p. 59).

Conforme a ética do discurso, o núcleo de fundamentação do ponto de vista moral reside no fato de que, apenas por meio de uma regra argumentativa, há a possibilidade de mudar o teor normativo desse jogo de linguagem epistêmico para a escolha de normas acionais,

propostas em discursos práticos juntamente com a reclamação de validade da moral. Por si mesmo, não há como a obrigação moral ser dependente de, por exemplo, uma imposição transcendental de pressupostos argumentativos que não se possam evitar, já que ela também se une às normas inseridas no discurso prático e que traz à tona as razões agrupadas nas reuniões em conselho. Isso é realçado, tendo em consideração que ‘U’ é aceito pelo teor normativo de pressupostos argumentativos unidos a um conceito de fundamentação de normas (HABERMAS, 2002, p. 59).

Tal fundamentação divide, conforme Habermas (2002, p. 59-60), o peso dos esforços para ser aceito, juntamente com um questionamento genealógico disfarçado, por meio de suposições caras à teoria da modernidade. Confirma-se, com a reflexão, que ‘U’ traz uma substância normativa excedente em sociedades complexas, tendo em vista que se mostram a partir da forma de um resíduo de si próprio poupado de argumentação, e com a forma da ação que segue o caminho do acordo recíproco.

Sobra, então, a questão da aplicação da norma, já que o ponto de vista moral é validado em sua plenitude apenas por meio do princípio da adequação e levando em consideração os juízos morais singulares. Concluído que discursos de fundamentação e aplicação são transmitidos com sucesso, mostra-se que questões práticas são divergentes, partindo do ponto de vista moral, já que questões morais referentes ao convívio correto são divididos entre questões pragmáticas da escolha racional e questões éticas do bem viver. Fica evidente também, retrospectivamente falando, que ‘U’ funciona a partir de um princípio discursivo mais extenso, de início com o objetivo de um questionamento de ordem moral. ‘D’ pode ser trabalhado, igualmente, como favorecedor de outras questões, como, por exemplo, para reuniões em conselho de um legislador político como também para discursos jurídicos (HABERMAS, 2002, p. 60).

Referências Bibliográficas

AUSTIN, John Langshaw. **Quando Dizer é Fazer:** palavras e ação. Tradução Prof. Danilo Marcondes de Souza Filho. Porto Alegre: Artes Médicas: 1990.

BANNWART JÚNIOR, Clodomiro José. **Estruturas Normativas da Teoria da Evolução Social de Habermas.** 2008. 275 f. Tese (Doutorado em Filosofia). Universidade Estadual de Campinas, Campinas, 2008.

DURÃO, Aylton Barbieri. A Tensão entre Faticidade e Validade no Direito Segundo Habermas. **ethic@**, Florianópolis, v.5, n.1, p. 103-120, Jun. 2006.

HABERMAS, Jürgen. **Teoria y Praxis**. Trad. Salvador Mas Torres e Calos Moya Espi. Madrid: Tecnos, 1987a.

_____. **Técnica e ciência como “ideologia”**. Trad. Artur Morão. Lisboa: Edições 70, 1987b.

_____. **Consciência Moral e Agir Comunicativo**. Trad. De Guido Antônio de Almeida. Rio de Janeiro: Tempo Brasileiro, 1989.

_____. **O discurso filosófico da modernidade**. Trad. Luiz Sérgio Repa e Rodnei Nascimento. São Paulo: Martins Fontes, 2000.

_____. **A inclusão do outro: estudos de teoria política**. Trad. George Sperber e Paulo Astor Soethe. São Paulo: Edições Loyola, 2002.

_____. **Direito e Democracia: entre facticidade e validade**. Trad. Flávio Beno Siebeneichler. 2ª ed. Rio de Janeiro: Tempo Brasileiro, 2003. 2 v.

MOREIRA, Luiz. **Fundamentação do Direito em Habermas**. 3ª ed. Belo Horizonte: Mandamentos, 2004.

PINZANI, Alessandro. **Habermas**. São Paulo: Artmed Editora, 2009.

VELASCO ARROYO, Juan Carlos. **Para Leer a Habermas**. Madrid: Alianza Editorial, 2003.

Razões públicas como razões práticas

Public Reasons as Practical Reasons

Tiaraju Molina Andreatza
Universidade do Vale do Rio dos Sinos

Resumo

Neste artigo analiso o modelo de razão pública de John Rawls. O artigo tem duas partes. A primeira analisa como podemos diferenciar razões públicas de outras razões. Proponho que um fato é uma razão pública se a explicação de por que ele conta como uma razão pode ser baseada exclusivamente em valores políticos. A segunda parte analisa quando a sentença “S tem uma razão pública para realizar ϕ ” é um enunciado verdadeiro sobre o que uma pessoa, S, tem razões para fazer. Defendo a interpretação de que S tem uma razão para fazer ϕ apenas se a realização de ϕ por S promover algum desejo que S teria em reflexão apropriada.

Palavras-chave

Razões Públicas. Razões Práticas. Condição de Aceitabilidade. Razoabilidade. Reflexão Apropriada. Valores Políticos.

Abstract

In this paper my aim is to examine the public reason account of John Rawls. The paper has two main sections. In the first part I analyze how we can distinguish public reasons from other reasons. I propose that a fact is a public reason if the explanation of why it count as reason can be exclusively based on political values. In the second part I examine when the sentence “S has a public reason to ϕ ” is a true utterance about a person’s, S, reasons. I propose that S has a public reason to ϕ only if S’s ϕ -ing would promote some desire that S would have in due reflection.

Key-words

Public Reasons. Practical Reasons. Acceptability Condition. Desire-based Reasons. Reasonableness. Due Reflection. Political Values.

1. Situando a questão

Teorias contratualistas em filosofia moral têm, como um de seus conceitos centrais, o de razões para ação. Scanlon, por exemplo, defendeu que um juízo de certo e errado não é nada mais do que um juízo sobre o que nós temos razão para fazer (SCANLON, 1998, p.3), e Rawls defendeu que a tarefa de cidadãos razoáveis é identificar, dado o fato do pluralismo razoável, quais razões eles devem oferecer uns aos outros quando questões fundamentais de justiça estão em jogo - e propôs que as razões apropriadas para cumprir esse papel são *razões públicas* (RAWLS, 1997, p.766/795). Devido a essa centralidade, não é surpreendente que Scanlon tenha reservado um capítulo inteiro de *What We Owe to Each Other* (1998) para esclarecer o que são razões, e, mais recentemente, todo um livro (*Being Realistic About Reasons*, 2014). Por outro lado, não se encontra a mesma diligência tanto nos escritos do próprio Rawls quanto na imensa literatura que os discutem. Rawls, de fato, reservou uma conferência do *Political Liberalism* e um ensaio posterior para apresentar a sua doutrina da razão pública, mas em nenhum momento explica em que sentido falar de razões públicas é falar de *razões*. Neste artigo, pretendo elaborar uma explicação que Rawls poderia ter oferecido, estivesse ele preocupado em oferecer uma. Essa explicação, defenderei, é congruente com a posição defendida no *Political Liberalism*.

Em um sentido usual, quando fazemos um enunciado normativo da forma “S tem uma razão para ϕ ”, em que S é uma pessoa qualquer e ϕ está para um verbo designando uma

ação, estamos identificando a existência de uma relação normativa tripartida entre um fato, um agente e uma ação¹. Por exemplo, o fato de que há praias e faz calor no Rio de Janeiro é uma razão para S, que gosta de vestir roupas leves e jogar futebol na praia, visitar o Rio de Janeiro. Razões, como visto nesse exemplo, são fatos que contam para certos agentes em favor de certas ações. Nós também podemos entender razões simplesmente como proposições ou considerações verdadeiras e permitir que não apenas fatos não-normativos possam funcionar como razões, mas também fatos normativos. Assim, que a honestidade é uma virtude pode ser, em princípio, uma razão para ser honesto. Se nós quisermos falar em razões públicas como razões para alguém fazer algo - qual outro sentido seria? - então assumirei que precisamos enquadrá-las nesse mesmo esquema relacional tripartido.

Teorias sobre a natureza de razões para agir oferecem critérios para identificar sob que condições enunciados do tipo “S tem uma razão para ϕ ” são verdadeiros ou apropriados. Uma maneira útil de abordar razões públicas como razões práticas é procurar identificar qual é a condição de verdade de enunciados em que razões públicas figuram como um de seus termos. Imagine que estejamos exigindo de uma pessoa, S, que ela se oponha à escravidão. Para justificar a demanda, dizemos que o fato de que a escravidão permite que umas pessoas possuam outras como sua propriedade é uma razão para se opor à escravidão². Esse fato, penso, seria um exemplo de razão *pública*. Rawls tem o seu próprio: o fato de que a separação entre Estado e igreja protege a religião do Estado e o Estado da religião é uma razão pública para separar Estado e igreja (RAWLS, 1997, 795). A questão que orientará este artigo é então esta: por que o fato de que a escravidão permite que umas pessoas possuam outras como sua propriedade é uma razão pública para S se opor a ela?

Essa é uma questão ambígua que permite pelo menos duas interpretações - e duas respostas diferentes. Rawls acredita que a sua teoria da justiça - ou qualquer teoria moral - fornece princípios ou padrões por meio dos quais é possível explicar por que certos fatos contam como razões em favor da justificação de certas demandas (RAWLS, 2005, 122). A ideia é que teorias morais oferecem princípios para selecionar, dentre todas as características factuais de ações e eventos, quais são relevantes para o juízo moral e por quê. Por exemplo, a teoria

1 - Scanlon defende que enunciados de razões estabelecem uma relação entre quatro termos: “This suggests that “is a reason for” is a four-place relation, $R(p, x, c, a)$, holding between a fact p , an agent x , a set of conditions c , and an action or attitude a . This is the relation that holds just in case p is a reason for a person x in situation c to do or hold a ” (SCANLON, 2014, p.31). A ideia de relação entre três termos, que uso como modelo, é defendida por Schroeder: “This relation is a three-place relation between the thing that is the reason, the agent for whom it is a reason, and the action-type that it is a reason to perform (SCHROEDER, 2007, p.17). Adoto o esquema do Schroeder e não o do Scanlon por conveniência, já que a introdução de um elemento a mais complicaria o quadro, sem necessariamente acrescentar à discussão mais restrita deste ensaio.

2 - Esse exemplo é de fato citado por Rawls, embora eu o tenha alterado para colocar em destaque que uma ação particular está sendo exigida: “To argue that slavery is unjust we appeal to the fact that it allows some persons to own others as their property” (RAWLS, 2005, p. 122).

da justiça de Rawls pode explicar que o que há de moralmente significativo no fato de que umas pessoas tenham outras como sua propriedade é a violação da liberdade e da igualdade de quem é escravizado. Um utilitarista, por sua vez, poderia explicar que o que é moralmente significativo nesse fato é que quando uma pessoa é propriedade de outra o seu bem-estar é reduzido. Esse é o primeiro sentido em que nós podemos abordar a questão acima.

O segundo sentido é muito diferente. Quando dizemos para S que o fato de que a escravidão permite que umas pessoas tenham outras como sua propriedade é uma razão para ele se opor a ela, nós atribuímos uma razão para ele e fazemos um enunciado sobre as suas razões. O problema, nesse segundo sentido, é saber por que esses fatos selecionados para contar como razões por uma teoria moral são razões práticas para S - ou porque eles figuram em um enunciado verdadeiro sobre as razões de S. O que precisa ser esclarecido no segundo sentido, e que não é posto em questão no primeiro, é o sentido em que enunciados de razão pública são enunciados de razões práticas. Essa questão é suscitada pela consideração de que uma sentença sobre as razões públicas de S não parece muito diferente de outros enunciados sobre razões que podemos fazer a respeito de S, como o enunciado de que ele tem uma razão para visitar o Rio de Janeiro. A explicação de por que o fato de que a escravidão permite que umas pessoas possuam outras como sua propriedade é uma razão pública para S se opor a ela tem de ser a mesma que oferecemos para explicar por que é verdadeiro que o fato de que há praias no Rio de Janeiro é uma razão para S visitar o Rio de Janeiro.

Com uma terminologia algo imprecisa, mas geralmente útil, poderíamos caracterizar o primeiro sentido da nossa questão como um sentido de primeira ordem - porque a resposta para ela consiste em especificar o que nós temos razões públicas para fazer - e o segundo sentido como um sentido de segunda ordem - porque a resposta para ela tem de mostrar por que razões públicas são razões. A estrutura deste artigo reflete esses dois sentidos. Assim, primeiro identificarei o critério contratualista que Rawls articula para estipular quais são as razões públicas que nós temos e como nós podemos diferenciar razões públicas de outras razões - denominarei esse critério de *condição de aceitabilidade*. Na seção seguinte sugerirei que é possível interpretar o seu contratualismo como comprometido com a tese de que razões públicas são razões práticas porque elas respondem aos desejos do agente para o qual elas são razões recomendando ações que promoveriam os desejos que o agente teria se ele fosse adequadamente reflexivo - denominarei essa posição como uma *teoria de razões baseada em desejos*.

2. O que conta como razões públicas

Parece plausível sumarizar a resposta de Rawls à questão de primeira ordem na seguinte condição de aceitabilidade, como a denominou:

Condição de aceitabilidade: um fato p é uma razão pública para fazer ϕ se a explicação de por que p conta como uma razão para fazer ϕ pode ser formulada exclusivamente com base em considerações que todas as pessoas razoáveis poderiam aceitar sem abrir mão de suas concepções de bem particulares.

Por exemplo, se a única explicação disponível de por que o fato de que a escravidão permite que umas pessoas tenham outras como sua propriedade é uma razão para se opor a ela for a de que a escravidão é contrária à lei de Deus, então esse fato não contaria como uma razão pública. Isso porque o que torna esse fato moralmente relevante de acordo com essa explicação - ser contrário à lei de Deus - não poderia ser aceito por todas as pessoas razoáveis como uma consideração verdadeira sem que elas precisassem abrir mão de suas concepções de bem particulares. Ela pode selecionar o que conta como razões, mas não o que conta como razões públicas. Considerações que dependem exclusivamente de doutrinas morais, filosóficas e religiosas não satisfazem a condição de aceitabilidade.

O conceito decisivo da condição de aceitabilidade é o de pessoa razoável. Rawls remete o conteúdo dos termos racional e razoável a um artigo de Sibley, “Rational versus Reasonable” (1953). Nesse artigo, Sibley distingue dois sentidos em que, ordinariamente, avaliamos a conduta de alguém. Nós dizemos de uma pessoa que ela é irracional quando ela não é intelectualmente virtuosa. Por exemplo, quando ela sacrifica um fim em nome de outro que ela sabe que é menos valoroso, quando ela seleciona meios irrealis para atingir seus fins, quando ela deixa de colocar em prática as suas decisões, quando ela não se informa sobre os seus objetivos, quando ela adota meios ineficazes para atingir seus fins etc. Racional aqui denota um tipo de performance intelectual, e, e essa é a parte que Rawls salienta estar especialmente de acordo com Sibley, quando uma pessoa é racional nós não sabemos quais fins ela adotará, apenas que perseguirá os seus fins inteligentemente - não sabemos se ela é boa ou má, justa ou injusta (RAWLS, 2005, p. 49). Sibley esclarece que o razoável, por outro lado, envolve mais do que racionalidade ou inteligência práticas. Razoabilidade é um tipo de disposição moral ou desejo para agir justificadamente e para “raciocinar em comum” com o outro que nós podemos encontrar em uma pessoa razoável (SIBLEY, 1953, p. 557-560)

Uma pessoa razoável pode ser definida da seguinte forma:

Pessoa razoável: S é razoável somente se S tem (i) um desejo para realizar ações em favor das quais S sinceramente acredita que há razões e (ii) S aceita que a explicação de por que essas razões contam em favor das suas ações tem de poder ser aceita por outras pessoas similarmente motivadas sem que elas precisem abrir mão de suas concepções de bem particulares.

Como indicado por (i) e (ii), essa definição tem dois aspectos. Começo por (i). Para entendê-lo nós precisamos levar em conta as referências de Rawls ao princípio de motivação moral proposto por Scanlon em “Contractualism and utilitarianism” (1982). No artigo em questão Scanlon esboça a sua visão, que em *What We Owe to Each Other* ele acabaria por desenvolver mais pormenorizadamente, de que uma ação incorreta é uma ação cuja realização é proibida por um conjunto de princípios que ninguém poderia razoavelmente rejeitar (SCANLON, 1998, p. 153). Note que qualquer ação pode ser alvo de uma objeção: se alguém me pede dinheiro com o objetivo de usá-lo para apostar na loteria e eu não dou, a minha ação poderia ser objetada por essa pessoa citando como razão que ela gosta de apostas, mas não quer arriscar perder o próprio dinheiro. Essa razão não é razoável, portanto a objeção não é razoável - isto é, não torna errado o meu ato de não dar o dinheiro. E ela não é razoável porque o seu gosto por apostas não é uma razão para que eu doe dinheiro para ela. O ponto central é que o critério de rejeição razoável está baseado em uma concepção do que conta como boas e más razões para se oferecer em favor da recusa de um ato. Agir justificadamente é, em uma primeira aproximação, realizar ações suportadas por razões (SCANLON, 1998, p. 197-198).

Para Scanlon uma pessoa razoável é alguém que tem o desejo de justificar as suas ações para os outros com base em razões que eles não poderiam razoavelmente rejeitar (SCANLON, 1982, p. 116), e, nas palavras de Rawls, é uma pessoa movida pelo desejo por um mundo social no qual ela pode cooperar com os outros com base em termos (razões) que todos podem aceitar (desde que os outros estejam similarmente motivados) (RAWLS, 2005, p. 50). Esse desejo ou disposição para realizar ações em favor das quais há razões é o primeiro aspecto da pessoa razoável - e algo que tanto Rawls quanto Scanlon aceitam.

O aspecto (ii), que é apenas sustentado por Rawls, é a aceitação e disposição em reconhecer as restrições que os limites do juízo (*burdens of judgment*) impõem no tipo de razões que são adequadas para se rejeitar razoavelmente uma demanda de justiça política.

Os limites do juízo são uma explicação da fonte do desacordo entre pessoas razoáveis: por exemplo, como muitos conceitos são vagos e de difícil interpretação em casos difíceis, é compreensível que pessoas razoáveis tenham opiniões divergentes em uma série de questões morais e não morais importantes³. Rawls afirma que uma pessoa razoável reconhece esses limites, o que significa que ela sabe que apesar de a sua visão de mundo particular seja verdadeira e as de outras pessoas razoáveis sejam falsas, as circunstâncias tornam plausível que outras pessoas razoáveis afirmem uma visão oposta à sua. Como resultado desse reconhecimento, uma pessoa razoável entende que superar esses limites através de argumentos (ou outros meios) é impraticável e indesejável, pois representaria um desrespeito à liberdade e à igualdade do outro. Por isso, pessoas razoáveis estão dispostas a comportar-se de acordo com razões que, de alguma forma, *respeitam* a concepção de bem de outra pessoa⁴. *S* respeita alguém, nesse sentido, se ele não insiste em oferecer em favor de suas demandas e suas ações apenas as razões tais como a sua doutrina abrangente as identifica, e, no lugar delas, ou em apoio a elas, fornece também razões que podem ser aceitas por outros sem que eles precisem abdicar de suas concepções de bem particulares - se elas são razoáveis (RAWLS, 2005, p. 61).

O aspecto (ii) dessa definição expressa que pessoas razoáveis são aquelas que aceitam a condição de aceitabilidade⁵. Esse aspecto demarca uma diferença drástica entre os modelos contratualistas de Rawls e Scanlon. Considere os momentos em que Scanlon analisa as implicações do contratualismo para a avaliação de ações que representam a destruição de belezas naturais como o Grand Canyon. A explicação de por que o fato de que o Grand Canyon seria destruído por alguma ação é uma razão para se opor a essa ação poderia ser a consideração de que destruir o Grand Canyon significaria privar as pessoas do tipo de atividade enriquecedora que faz com que a vida valha a pena ser vivida - no caso, experimentar belezas naturais. Penso que Scanlon autorizaria essa consideração como uma legítima explicação de por que certos fatos contam como razões⁶. Essa consideração, porém,

3 - Rawls elabora uma lista que contém seis fontes óbvias de erro, e revela que há espaço para muitas outras (RAWLS, 2005, p. 56-57).

4 - Para uma defesa dessa interpretação ver Quong (2011, p. 253).

5 - A principal motivação para a introdução de uma restrição a partir dos limites do juízo é garantir a efetividade de uma teoria da justiça. Se uma demanda de justiça fosse *contrária* a uma concepção de bem particular, os cidadãos abandonariam a demanda e manteriam a concepção de bem. Não é claro que muitos autores reconheceram esse ponto. Bird, por exemplo, interpreta Rawls como defendendo que se uma pessoa evangélica acreditar que os juízos morais implícitos na ideia de neutralidade política são falsos, o liberal responderá que isso não é uma razão para que eles sejam rejeitados por ela porque eles não estão sendo recomendados porque são verdadeiros, mas porque são razoáveis (BIRD, 1996, p. 69-70). Entretanto, o que Rawls defende é que uma concepção política de justiça “must be confirmed as true by the comprehensive doctrine” (RAWLS, 1997, p. 801). Uma concepção política é defendida como razoável, ao invés de verdadeira, porque ela sinaliza apenas o que pessoas razoáveis compartilham como verdadeiro sobre justiça - sem defender que o que pessoas razoáveis compartilham como verdadeiro é o que é verdadeiro sobre justiça.

6 - Para confirmação dessa leitura, considere o argumento de Scanlon em favor da prioridade dos valores de certo errado

depende de suposições sobre o que é uma vida que vale a pena ser vivida, suposições que para Rawls são abrangentes e por isso não poderiam ser aceitas por todas as pessoas razoáveis sem que elas precisassem abrir mão de suas concepções de bem particulares. Nesse caso, uma explicação harmoniosa com os limites do juízo poderia ser a de que o fato de que uma ação representaria uma ameaça ao Grand Canyon é uma razão para se opor a ela porque ela representa a violação de regulamentos ambientais que preservam patrimônios naturais. Essa consideração, que de alguma forma depende da autoridade de regras legítimas, é pública porque pode ser aceita por qualquer pessoa razoável, independentemente da sua concepção de bem.

Nós temos, assim, uma explicação de por que certos fatos contam como razões públicas. A ideia central é que razões públicas são fatos selecionados como relevantes em questões de justiça a partir de considerações *compartilhadas* entre pessoas razoáveis - isto é, razões que todas as partes podem aceitar a partir dos seus pontos de vista. A ideia de razões públicas como razões compartilhadas entre pessoas razoáveis não é aceita por todos os teóricos da razão pública. Gaus, por exemplo, acredita que não existem considerações que todas as pessoas razoáveis poderiam aceitar - pessoas razoáveis, ele entende, não compartilham valores políticos comuns e, se compartilham, interpretam diferentemente as consequências significativas desses valores para que o apelo a eles seja de alguma valia (GAUS, 2011, p. 283). Para que uma demanda seja justificada, é suficiente que cada indivíduo tenha as suas próprias razões individuais para aceitá-la, em que as razões que uma pessoa tem (conforme terminologia utilizada por Gaus) são as razões que ela acreditaria que tem após uma quantidade significativa de raciocínio (esse requerimento visa apenas corrigir aqueles erros de performance intelectual levantados por Sibley)⁷. Se, após uma quantidade satisfatória de raciocínio, você acredita que o mundo foi criado pelo monstro do espaguete e que o desejo do monstro do espaguete é o de que todos sejam livres e iguais, eu posso justificar uma demanda de tolerância para você oferecendo como razão o fato de que o monstro do espaguete deseja que todos sejam livres e iguais. Eu sei que o monstro do espaguete não existe, mas, e eis o cerne do liberalismo político para Gaus, eu tenho de reconhecer que o que conta como boas razões para você é diferente do que conta como boas razões para mim (GAUS, 2011, p. 283).

Rawls não é um cético quanto à existência de considerações compartilhadas que

sobre outros valores, como a amizade (SCANLON, 1998, p. 159-167), ou como ele concebe a relação entre o que ele denomina de *impersonal* e *personal reasons* (SCANLON, 1998, p. 219-220).

7 - Gaus entende que “to deny that others have reasonable evaluative standards, justified beliefs, and reasons that I do not share, and with which I disagree, deeply misconceives our real predicament as severely cognitively bounded reasoners” (GAUS, 2011, p. 292). Alguém poderia entender que a ideia de limites do juízo expressa essa convicção de que um indivíduo deve reconhecer que os outros têm razões que ele não compartilha. Mas não é esse o caso. O que a ideia dos limites do juízo implica é que eu devo reconhecer que os outros podem acreditar de boa fé que têm razões diferentes das minhas, não que se deve reconhecer que eles têm razões diferentes das minhas. Isso seria incoerente. Quando um ateu entende que o fato de que Deus criou os homens não é verdadeiro, ele acredita que *esse fato não é uma razão* para se opor à escravidão.

explicam por que certos fatos contam como razões⁸. Diferentemente de Gaus, ele acredita que, quando oferecemos um argumento para o outro, nós devemos oferecer, além de uma conclusão que o outro pode aceitar, também razões que ele, assim como nós, pode aceitar (RAWLS, 1997, p. 800). Há duas motivações para sustentar essa tese. A primeira é a de que *S* não realiza o seu desejo de razoabilidade se o outro aceita uma justificativa que *S* sabe que não é uma boa justificativa (SCANLON, 1982, p. 116). Se o seu amigo pede a sua ajuda para realizar uma mudança e você responde, mentindo, que está doente e por isso não pode ajudá-lo, mesmo que o seu amigo aceite a justificativa, você se sentirá mal porque sabe que foi uma má justificativa. A segunda segue do próprio conceito de justificação do qual Rawls parte, em que justificar é endereçar um argumento ao outro partindo de premissas aceitas em comum e premissas que sinceramente se espera que possam ser aceitas pelo outro na medida em que ele é racional e razoável (RAWLS, 1997, p. 786). Se eu sei que não existe monstro do espaguete, usar essa crença como uma razão é colocar como objetivo da minha relação com o outro alguma outra coisa - talvez persuasão? - que não a justificação.

A ideia de que uma pessoa razoável tem o desejo de justificar as suas ações para os outros com base em razões que eles não poderiam razoavelmente rejeitar é, para Rawls, apenas uma maneira de dizer que uma pessoa razoável está preocupada em se relacionar com os seus pares de acordo com valores morais e políticos que são também reconhecidos por eles, ou podem ser reconhecidos por eles (RAWLS, 1997, p. 766). Nós podemos considerar os valores políticos propostos por Rawls, bem como os seus princípios de justiça, como os elementos de considerações aptas a explicar por que certos fatos contam como razões públicas. Eles são aptos para esse papel porque apenas considerações baseadas em valores políticos satisfazem a condição de aceitabilidade. Assim, valores políticos são o conteúdo de considerações que explicam por que certos fatos contam como razões (no sentido de primeira-ordem). Como razões podem ser entendidas como proposições ou considerações verdadeiras, nós devemos também permitir que os próprios valores políticos sejam vistos como razões públicas em determinados contextos. Por exemplo, na medida em que razões públicas são elementos que figuram na justificação de demandas de justiça, parece apropriado oferecer como razão para se opor à escravidão a proposição de que todas as raças têm um igual direito à igualdade política.

8 - Essa é também a posição defendida por Nagel: "This means that it must be possible to present to others the basis of your own beliefs, so that once you have done so, they have what you have, and can arrive at a judgment on the same basis. That is not possible if part of the source of your conviction is personal faith or revelation" (NAGEL, 1987, p. 232). A ideia de razões compartilhadas é uma característica constitutiva de uma doutrina da razão pública também para Macedo: "It holds out the possibility, however, that people for whom disagreement about our highest ideals and the truth as a whole is a permanent condition, might nevertheless share a political community based on a reasonableness they can share" (MACEDO, 1997, p. 29). Dado como Rawls, Nagel e Macedo entendem a ideia de liberalismo político e razões públicas, a teoria de Gaus seria mais uma teoria cética quanto a existência dessas razões do que uma teoria alternativa.

3. Razões públicas como razões práticas

Gostaria, agora, de abordar o sentido de segunda-ordem que podemos ter em mente quando perguntamos por que é verdadeiro que o fato de que a escravidão permite que umas pessoas possuam outras como sua propriedade é uma razão para S se opor a ela. Parece-me que, para Rawls, razões públicas são razões práticas no sentido de que elas recomendam ações cuja realização promoveria algum desejo que o agente para o qual elas são razões teria sob certas condições de reflexão adequada. Rawls teria, assim, alguma versão de uma teoria de razões baseada em desejos.

Para ver por que Rawls aceitaria esse tipo de teoria nós temos que retornar mais uma vez para o artigo de 1982 de Scanlon, “Contractualism and utilitarianism”. No artigo em questão Scanlon defendeu, o que depois ele veio a rejeitar em seus escritos futuros sobre o assunto (SCANLON, 1998, p. 7-8), a tese de que a fonte da motivação moral é um desejo de justificar as ações para outros em termos que eles não poderiam razoavelmente rejeitar. Esse é o desejo, ele explica, que é acionado pelo reconhecimento de que uma ação é errada, e que explica por que a moralidade tem o tipo de importância e força motivacional que geralmente atribuímos a ela (SCANLON, 1982, p. 116). Um dos pontos principais do artigo de Scanlon é rejeitar que nós deveríamos preferir o utilitarismo porque ele identifica a correta fonte da motivação moral - no caso, fatos sobre o bem-estar individual. Essa explicação oferecida por Scanlon não passa de um conjunto de enunciados empíricos, não normativos, sobre o que move as pessoas a agir quando elas julgam que algo é correto ou incorreto. Mas, em sua leitura desse artigo, Rawls também acredita que ela é mais do que isso:

Scanlon's principle is more than a psychological principle of motivation (though it is that) since it concerns the fundamental question *why anyone should care about morality at all*. The principle *answers this* by saying that we have a basic desire to be able to justify our actions to others on grounds they could not reasonably reject—reasonably, that is, given the desire to find principles that others similarly motivated could not reasonably reject (RAWLS, 2005, p. 49n2. Grifos nossos).

Inequivocamente, o desejo de razoabilidade tem um papel normativo a cumprir aqui. Rawls diz que S deve se importar com ϕ *porque* ele tem um desejo de razoabilidade. É natural assumir que, ao decidir se deve se importar com ϕ , S deve levar em consideração se ele tem razões para se importar com ϕ - isto é, se há fatos que contam em favor da sua atitude de se

importar com ϕ . Isso significa que se S deve se importar com ϕ é *porque* S tem razões para se importar com ϕ . Se deve se importar com ϕ *porque* ele tem um desejo de razoabilidade, como Rawls está defendendo, então uma interpretação do que Rawls está dizendo é a de que o *desejo* de razoabilidade *explica* por que é verdadeiro que S tem uma razão para se importar com ϕ . Em outra passagem, Rawls fala que S deve considerar o que “responde” ao seu desejo de razoabilidade e o que esse desejo “requer”, o que, penso, também sugere essa interpretação (RAWLS, 2005, p. 85n33).

Acredito que essa é uma interpretação natural dessas passagens. Segundo essa interpretação, Rawls estaria defendendo a seguinte condição necessária para se dizer verdadeiramente de S que ele tem uma razão pública para realizar ϕ :

S tem uma razão pública para fazer ϕ somente se S tem um desejo de razoabilidade que seria promovido pela realização de ϕ por S .

Rawls defenderia essa condição porque ele estaria partindo da visão segundo a qual:

S tem uma razão para fazer ϕ apenas se a realização de ϕ por S promover algum desejo que S tem⁹.

De acordo com esse tipo de posição, o fato de que as pessoas tenham outras como sua propriedade é uma razão para S se opor à escravidão porque, se opondo à escravidão, S promoverá - ou responderá ao - o seu desejo de razoabilidade. Alguns fatos contam como razões públicas em favor de certas ações no sentido de que eles recomendam ações cuja realização promoveria algum desejo do agente para o qual eles são razões - é nesse sentido que o fato de que há praias no Rio de Janeiro conta em favor da ação visitar o Rio de Janeiro para alguém que gosta de praias.

Não acredito, porém, que essa teoria baseada em razões siga necessariamente das passagens indicadas. O que se segue, efetivamente, é que o desejo de razoabilidade é parte da explicação de por que S tem uma razão para se importar com ϕ , mas disso não segue que

⁹ - Essa definição de uma teoria de razões baseada em desejos é oferecida por Scanlon: “X has a reason to do A just in case doing A would promote the fulfillment of some desire that X has” (SCANLON, 2014, p. 3). Schroeder oferece a seguinte definição: “if r is a reason for x to do a , that is because there is some p such that x has a desire whose object is p , and the truth of r is part of what explains why x 's doing a promotes p (SCHROEDER, 2007, p. 29).

a explicação de por que *S* tem uma razão para se importar com ϕ depende inteiramente do seu desejo se importar com ϕ . Por exemplo, teorias de razões baseadas em valores defendem que um fato é uma razão para *S* realizar ϕ somente se, ou porque, ϕ é uma ação que tem valor intrínseco, ou se a realização de ϕ por *S* promove algo que tem valor (PARFIT, 2011, p. 45). Filósofos que defendem esse tipo de posição, como Scanlon e Parfit, podem aceitar que *S* tem razão para visitar o Rio de Janeiro porque ele deseja vestir roupas leves e jogar futebol na praia e visitando o Rio de Janeiro ele promoveria esses desejos. Se *S* odiasse praia e futebol, então o fato de que há praias na cidade do Rio de Janeiro não seria uma razão para *S* visitá-la. Assim, o desejo de *S* é parte da explicação de por que *S* tem uma razão para visitar o Rio de Janeiro. Porém, esses filósofos podem argumentar, *S* tem razão para visitar o Rio de Janeiro não porque ele tem um desejo por praia e futebol, mas porque a ação de visitar o Rio de Janeiro é uma ação valiosa - por exemplo, porque uma vida em que *S* faz o que ele deseja fazer é uma vida de felicidade e realização pessoal¹⁰. Em outras palavras, não se segue da constatação de que certos fatos contam como razões para *S* porque ele tem certos desejos que fazem inteiramente parte da explicação de por que esses fatos são razões para *S*.

É devido a essa consideração que não é possível afirmar que as passagens acima indicadas sejam conclusivas em favor da interpretação que eu ofereço - também não são, evidentemente, conclusivas em favor de uma interpretação rival. Nós temos de nos perguntar, contudo, qual é a interpretação mais natural daquelas passagens. Eu responderia que é uma teoria de razões baseada em desejos.

Há, penso, uma fonte extra de evidências que pesa em favor da minha sugestão. Creio que é uma constatação trivial, identificada por uma série de pessoas que escreveram sobre o assunto¹¹ e pelo próprio Rawls, que o *Political Liberalism* não pretende oferecer argumentos para pessoas que não aceitam valores políticos e que não têm o desejo de razoabilidade¹². Nós já vimos, na seção anterior, que, para Rawls, a justificação é um argumento endereçado ao outro, que parte de premissas que ambas as partes *aceitam* como válidas. Rawls claramente entende que nós não devemos justificação para o irrazoável (RAWLS, 2005, p. 64n19), mas não é claro se ele também acredita que nós não podemos justificar demandas para o irrazoável.

10 - Essa observação é feita por Scanlon em favor de uma teoria de razões baseada em valores: "It is very plausible to say that what explains the difference between Ronnie's reasons and Bradley's is the fact that Ronnie enjoys dancing and Bradley does not. But this leaves open the question of why the fact that Ronnie enjoys dancing makes it the case that the fact that there will be dancing at the party gives him a reason to go" (SCANLON, 2014, p. 47).

11 - Friedman (2000, p. 28), Laden (2003, p. 381-384) e Quong (2011, p. 233) enfatizam essa constatação.

12 - Por exemplo, quando responde por que alguém deveria honrar os limites da razão pública, Rawls responde que não honrá-los representaria a violação do princípio liberal de legitimidade (RAWLS, 2005, p. 216-217). O princípio liberal de legitimidade determina que o poder político é apenas legítimo quando ele é exercido de acordo com uma constituição cujos elementos essenciais todos os cidadãos podem razoavelmente endossar. Esse princípio é apenas uma maneira de expressar o razoável. Se a pergunta "por que honrar os limites da razão pública?" é feita por uma pessoa que rejeita o razoável, responder que a sua conduta seria ilegítima é enunciar algo que ela já sabe - é enunciar justamente o que ele está pondo em questão.

Como o irrazoável não aceitaria os valores políticos nos quais a justificação de uma demanda de justiça repousaria, parece que demandas de justiça não são justificáveis para ele. Se nós aceitarmos a plausível visão de que uma demanda é justificada para *S* somente se *S* tem razões para aquiescer à demanda, então dizer que uma demanda de justiça política não é justificável para *S* implica afirmar que *S* não tem uma razão pública para realizar a ação requerida por essa demanda. Uma teoria de razões baseada em desejos se ajusta com facilidade a esse resultado - de fato, ela explica por que razões públicas não são razões para pessoas irrazoáveis. Por outro lado, uma teoria baseada em valores conduz à conclusão de que pessoas irrazoáveis têm razões públicas para agir de acordo com demandas de justiça - já que ações requeridas por demandas de justiça promovem valores como reciprocidade e igualdade - e, por conseguinte, de que demandas de justiça *são* justificadas para pessoas irrazoáveis.

É possível que alguém queira manter a interpretação de que Rawls tenha uma teoria de razões baseada em desejos, mas sem aceitar a implicação de que pessoas que não têm o desejo de razoabilidade não têm razões públicas para agir de acordo com uma concepção política de justiça. Esse resultado poderia ser obtido, em princípio, com a seguinte qualificação para a formulação oferecida de uma teoria de razões baseada em desejos: o que *S* tem razão para fazer não é determinado pelos desejos que ele tem, como está sendo proposto, mas pelos desejos que ele teria se fosse apropriadamente reflexivo. Essa qualificação parece consistente com a importância que Rawls atribui para a ideia de reflexão apropriada (*due reflection*) no *Political Liberalism*¹³. Ela também parece resultar em uma teoria de razões baseada em desejos mais atrativa, pois intuitivamente parece que em alguns casos certos desejos do agente - pense em desejos cruéis ou em desejos extravagantes, como o desejo de contar até dez antes de falar uma palavra com a letra *p* - não fornecem razões para agir, e em outros casos o agente parece ter razões para fazer o que ele não deseja fazer. Alguém poderia utilizar essa versão qualificada de uma teoria de razões baseada em desejos para argumentar que razões públicas são razões para pessoas que não são razoáveis porque, embora elas atualmente não tenham um desejo de razoabilidade, elas *teriam* se fossem apropriadamente reflexivas.

A noção de reflexão apropriada sempre aparece no *Political Liberalism* como referência ao método do equilíbrio reflexivo (*reflective equilibrium*)¹⁴, de modo que ele parece um bom candidato para qualificar uma teoria de razões baseada em desejos. De acordo com a melhor interpretação do equilíbrio reflexivo, trata-se de um método que descreve como

¹³ - Rawls defende que nós temos uma teoria política aceitável somente quando os fatos que ela seleciona para contar como razões - como o fato de que a escravidão permite que umas pessoas tenham outras como sua propriedade - são coerentemente conectados por princípios aceitáveis para nós em reflexão apropriada (RAWLS, 2005, p. 28).

¹⁴ - Para uma interpretação do equilíbrio reflexivo como um modelo de racionalidade prática ver DePaul (1993, p. 101-121).

alguém deve proceder para decidir o que pensar e o que fazer em relação a um determinado assunto. Scanlon defende que é melhor compreendê-lo como descrevendo uma reflexão em três etapas (SCANLON, 2014, p.76-77). Em uma primeira etapa, *S* deve selecionar o que conta como suas convicções ou juízos ponderados (*considered*), que são basicamente juízos realizados em condições favoráveis ao julgamento (juízos realizados em momentos de estresse ou sob influência de coerção, por exemplo, não contam como ponderados). No segundo passo, *S* deve buscar por considerações gerais - como princípios - que os expliquem (Rawls sempre cita os juízos ponderados de repúdio à escravidão e de tolerância religiosa e estipula que esses juízos podem ser explicados com a ideia mais geral de que todos os cidadãos devem ser tratados como livres e iguais). É provável que *S* não encontre nenhum princípio que explique todos os seus juízos ponderados. Em um terceiro passo ele deverá decidir como responder à divergência entre os juízos ponderados e os princípios e ideias gerais: ele poderá revisar ou descartar alguns juízos, ou os princípios ou ambos, até obter um conjunto de juízos e um conjunto de princípios que estejam em “equilíbrio” ou em uma relação de ajuste mútuo. Presumivelmente, nós devemos também incluir os desejos de *S* como elementos que têm de estar em equilíbrio reflexivo com suas crenças e princípios.

A partir do equilíbrio reflexivo nós podemos oferecer a seguinte reformulação:

S tem uma razão para fazer ϕ se a realização de ϕ por *S* promover algum desejo que *S* teria como resultado do equilíbrio reflexivo¹⁵.

Uma característica notável do equilíbrio reflexivo é que um agente pode ser apropriadamente reflexivo independentemente do conteúdo das suas crenças ou do objeto dos seus desejos. Isso porque tudo o que o método exige é que as crenças e desejos de *S* estejam em uma certa relação adequada - basicamente uma relação de coerência. Em uma nomenclatura proposta por Parfit, parece apropriado caracterizar o equilíbrio reflexivo como oferecendo uma concepção procedimental de reflexão racional, e não uma concepção substantiva. A diferença entre as duas consiste no fato de que concepções procedimentais estipulam *como* alguém deve deliberar mas não inserem o reconhecimento de certas considerações morais e prudenciais como parte do que significa ser um agente

15 - Essa é uma definição possível para uma teoria baseada em desejos oferecida por Scanlon: “*x* has reason to do *A* if doing *A* would promote the fulfillment of a desire that *X* would have if *X* were fully aware of the relevant non-normative facts and thinking clearly” (SCANLON, 2014, p. 3). Parfit interpreta Rawls desta forma: “Rawls accepts a desire-based subjective theory, claiming that we ought rationally to try to achieve the aims that, after fully informed and procedurally rational deliberation, we would most want to achieve” (PARFIT, 2011, p. 347).

apropriadamente reflexivo - ou racional. Se *S* seguiu os passos do método corretamente, ele é procedimentalmente racional independentemente do que ele conclui como resultado da sua reflexão. Concepções substantivas de racionalidade - como, por exemplo, a defendida pelo próprio Parfit - são diferentes porque elas estipulam o reconhecimento de certos fins e crenças, e a aquisição de certos desejos, como parte do que significa ser um agente (substantivamente) racional (PARFIT, 2011, p. 62-63).

Não é claro como um ideal procedimental é capaz de garantir que pessoas irrazoáveis adquiririam, como resultado do equilíbrio reflexivo, o desejo de razoabilidade. O equilíbrio reflexivo de fato implica que as crenças e desejos com que *S* inicia o processo provavelmente serão diferentes das crenças e desejos com que ele termina, do que, naturalmente, se segue que muitas pessoas que não têm o desejo de razoabilidade poderiam vir a adquiri-lo como resultado da reflexão apropriada - e, conseqüentemente, sobre essas pessoas é verdadeiro que elas têm razões públicas para agir mesmo se, antes da reflexão, elas não tinham o desejo de razoabilidade. Apesar disso, parece que o resultado do método será sempre controlado pelos pontos de partida da reflexão. Isso explicaria por que o resultado do equilíbrio reflexivo não seria o mesmo para diferentes indivíduos¹⁶. Considere a discussão de Sibley, e a tese de que o razoável não pode ser deduzido do racional. Sibley acredita que quando uma pessoa é racional nós sabemos que ela perseguirá os seus fins inteligentemente, mas não sabemos se esses fins serão razoáveis. Rawls, especificamente, salienta que está de acordo com *essa* tese (RAWLS, 2005, p.49n1). O uso do equilíbrio reflexivo não garante que o resultado será necessariamente razoável, se os pontos de partida da reflexão também não o forem¹⁷. O que Rawls parece estar aceitando juntamente com Sibley é a tese de que o que é contrafactualmente verdadeiro a respeito de *S* (que desejos e crenças ele teria em reflexão apropriada) é dependente do que é factualmente verdadeiro a respeito de *S* (que crenças e desejos ele tem).

Gostaria de encerrar esta seção indicando uma tese que não estou atribuindo a Rawls, ao propor que ele tem uma teoria de razões baseada em desejos: a tese motivacional segundo a qual crenças - como a crença de que algo é requerido por uma concepção política de justiça - são insuficientes para explicar a motivação moral, que sempre requer, além de uma crença, um desejo. O desejo de razoabilidade, que é a fonte da motivação moral para Rawls, é por ele descrito como um *conception-dependent desire*: um desejo que tem como objeto um princípio - o princípio de realizar ações que outras pessoas podem aceitar como

16 - Essa é uma implicação do método enfatizada por Scanlon (2014, p. 79).

17 - Para uma defesa desta tese sobre o equilíbrio reflexivo, ver Kelly e McGrath (2010) e Daniels em seu último ensaio sobre o equilíbrio reflexivo (DANIELS, 1996, p. 167). Para a defesa da tese contrária de que uma concepção procedimental de racionalidade é capaz de gerar um desejo para agir razoável e moralmente independentemente dos pontos de partida da reflexão, ver os argumentos de Smith a partir do equilíbrio reflexivo (SMITH, 1994, p. 164-174).

justificadas - que o agente que tem o desejo concebe como parte de um certo ideal - o ideal de cidadania democrática (RAWLS, 2005, p. 84). Até onde posso ver, a visão de Rawls não é a de que o agente aceita o ideal de cidadania apenas se ele tem o desejo de razoabilidade, mas parece ser a de que ele tem esse desejo como resultado de formar a crença de que ele é requerido para que ele possa se ver como um agente moral. Rawls parece acreditar que esse reconhecimento é propiciado pela participação em uma cultura política pública na qual a virtude da cidadania é amplamente compartilhada (RAWLS, 2005, p. 85n33).

4. Considerações finais

Creio que uma teoria de razões baseada em desejos é uma interpretação possível, e natural, de algumas passagens centrais do *Political Liberalism*, além de se ajustar a outros aspectos da posição de Rawls, como a tese de que demandas de justiça política não são justificáveis para pessoas irrazoáveis. Porém, pelas razões discutidas, não acredito que essa seja a única interpretação possível. Dado tudo o que Rawls escreveu, ele pode muito bem estar defendendo uma teoria de razões baseada em valores.

O modelo de explicação oferecido ao longo deste ensaio buscou caracterizar razões públicas como fatos, como o fato de que a escravidão permite que umas pessoas tenham outras como sua propriedade, que contam em favor de certas ações relevantes para questões de justiça política para os agentes para os quais elas são razões. Na primeira parte, foi visto que os fatos que contam como razões públicas são fatos cuja explicação de por que eles contam como razões pode ser baseada exclusivamente em valores políticos - em outras palavras, cuja explicação satisfaz a condição de aceitabilidade. Na segunda parte, foi proposto que razões públicas são razões porque elas recomendam ações cuja realização promoveria algum desejo que o agente teria como resultado de uma reflexão apropriada. Note-se que essa explicação de por que certos fatos são razões públicas para *S* é uma que nós poderíamos oferecer para esclarecer outros enunciados normativos a respeito das razões de *S*, com isso sendo uma explicação bem sucedida do sentido em que razões públicas são razões práticas. Nós podemos dizer que o fato de que há praias no Rio de Janeiro é uma razão para *S* visitar o Rio de Janeiro porque *S* deseja usar roupas leves e jogar futebol na praia, e indo ao Rio de Janeiro ele promoveria esses desejos - do mesmo modo, o fato de que a escravidão permite que umas pessoas tenham outras como sua propriedade é uma razão para *S* se opor à escravidão porque *S* deseja ser razoável, e se opondo à escravidão ele promoveria esse desejo.

Referências

- BIRD, C. 1996. Mutual Respect and Neutral Justification. *Ethics*, v. 107, n.1, p. 62-96.
- DANIELS, N. 1996. *Justice and Justification: Reflective Equilibrium in Theory and Practice*. Cambridge: Cambridge University Press.
- DEPAUL, M. 1993. *Balance and Refinement: Beyond Coherence Methods of Moral Inquiry*. London: Routledge.
- FRIEDMAN, M. 2000. John Rawls and the Political Coercion of Unreasonable People. In: DAVION, V; WOLF, C (org.). *The Idea of Political Liberalism*. New York: Rwoman & Littlefield Publishers, p. 16-33.
- GAUS, G. 2011. *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World*. New York: Cambridge University Press.
- KELLY, T; MCGRATH, S. 2010. Is Reflective Equilibrium Enough?. *Philosophical Perspectives*, v.24, p. 325-359.
- MACEDO, S. 1997. In Defense of Liberal Public Reason: Are Slavery and Abortion Hard Cases?. *The American Journal of Jurisprudence*, v. 42, n. 1, p. 1-29.
- NAGEL, T. 1987. Moral Conflict and Political Legitimacy. *Philosophy & Public Affairs*, v. 16, n. 3, p. 215-240.
- PARFIT, D. 2011. *On What Matters*. Volume One. Oxford: Oxford University Press.
- QUONG, J. 2011. *Liberalism without Perfection*. Oxford: Oxford University Press.
- RAWLS, J. 2005. *Political Liberalism*. Expanded Edition. New York: Columbia University Press.
- _____. 1997. "The Idea of Public Reason Revisited". *The University of Chicago Law Review*, v. 64, n.3, p. 765-807.
- SCANLON, T. 1982. Contractualism and utilitarianism. In: SEN, A; WILLIAMS, B (eds.). *Utilitarianism and Beyond*. Cambridge: Cambridge University Press, p.103-128.
- _____. 2014. *Being Realistic about Reasons*. Oxford: Oxford University Press.
- _____. 1998. *What We Owe to Each Other*. Cambridge: Harvard University Press.
- SCHROEDER, M. 2007. *Slaves of the Passions*. New York: Oxford University Press.

SIBLEY, W. M. 1953. The Rational versus the Reasonable. *The Philosophical Review*, v. 62, n. 4, p. 554-560.

SMITH, M. 1994. *The Moral Problem*. Oxford: Blackwell Publishing.

Tradução

Cérebros numa cuba *

Brains in a Vat

Hilary Putnam

Tradução de L. H. Marques Segundo
Universidade Federal de Santa Catarina

Uma formiga passa por cima de um pequeno monte de areia. Na medida em que vai passando, traça nela uma linha. Por puro acaso, a linha faz curvas e se cruza de tal modo que claramente acaba parecendo com uma caricatura de Winston Churchill. Terá a formiga traçado uma imagem de Winston Churchill, uma imagem que *retrate* Churchill? Muitos, à primeira vista, diriam que não. Afinal, a formiga nunca vira Churchill, ou sequer uma imagem de Churchill, e não tivera a intenção de retratar Churchill. Ela simplesmente traçou uma linha (não intencionalmente), uma linha que *nós* conseguimos “ver como” uma imagem de Churchill.

Podemos expressar isso dizendo que a linha “em si” não é uma representação² de

* “Brains in a Vat”, in *Reason, Truth, and History*, pp. 1-21.

2 - Neste [artigo] os termos “representação” e “referência” vão se referir sempre à relação entre uma palavra (ou outro tipo de signo, símbolo, ou representação) e algo que efetivamente existe (*i.e.* não apenas a “objeto do pensamento”). Há um sentido de “referir” no qual posso “referir” a aquilo que não existe; não é esse o sentido usado aqui. Uma palavra mais antiga para aquilo que chamo “representação” ou “referência” é *denotação*.

Em segundo lugar, sigo o costume dos lógicos contemporâneos e uso “existe” com o significado de “existe no passado, no presente ou no futuro”. Assim, Winston Churchill “existe”, e podemos nos “referir a” ou “representar” Winston Churchill

qualquer coisa que seja. Do mesmo modo (no que diz respeito a várias coisas complicadas) as características de Winston Churchill não são suficientes para fazer com que algo represente ou refira Winston Churchill. E nem é necessário: na nossa comunidade, a forma impressa “Winston Churchill”, as palavras proferidas “Winston Churchill”, e muitas outras coisas são usadas para representar Churchill (ainda que de maneira não pictórica), embora não tenham o tipo de similaridade com Churchill – mesmo a linha traçada – que um retrato tem. Se a *similaridade* não é nem necessária nem suficiente para fazer com que algo represente outra coisa, como *alguma coisa* poderia ser necessária ou suficiente a esse propósito? De que modo, afinal de contas, uma coisa pode representar (ou “corresponder a”, etc.) outra coisa diferente?

A resposta parece fácil. Suponha que a formiga tivesse visto Winston Churchill, e suponha que ela tivesse inteligência e habilidade para desenhar um retrato dele. Suponha que ela fez a caricatura *intencionalmente*. A linha, nesse caso, representaria Churchill.

Por outro lado, suponha que a linha tivesse a forma WINSTON CHURCHILL. E suponha que isso tenha sido apenas um acidente (ignorando a improbabilidade envolvida). Assim, a “forma impressa” WINSTON CHURCHILL *não* teria representado Churchill, muito embora a forma impressa hoje represente Churchill quase que em todos os livros.

Portanto, pode parecer que o que é necessário para a representação, ou o que é principalmente necessário para a representação, é a *intenção*.

Mas para que a intenção de que *algo*, ainda que de uma linguagem privada (em que as palavras “Winston Churchill” fossem ditas em minha mente e não em voz alta), *representasse* Churchill, eu teria de ter sido capaz de *pensar sobre* Churchill em primeiro lugar. Se linhas na areia, barulhos, etc., não puderem “por si mesmos” representar algo, então de que modo é que essas formas pensadas conseguem “em sim mesmas” representar algo, se é que conseguem? De que modo o pensamento atinge e “apreende” aquilo que é externo?

No passado, alguns filósofos saltaram desse tipo de consideração àquilo que consideraram uma prova de que a mente é, *em natureza, essencialmente não física*. O argumento é simples: aquilo que dissemos sobre a linha traçada pela formiga se aplica a qualquer objeto físico. Nenhum objeto físico pode, em si mesmo, referir a uma coisa ao invés de outra; não obstante, os *pensamentos em nossa mente* obviamente referem-se, de maneira bem sucedida, a uma coisa ao invés de outra. Assim, os pensamentos (e, por conseguinte, a mente) são de uma natureza essencialmente diferente da dos objetos físicos. Os pensamentos têm a característica da *intencionalidade* – eles podem se referir a coisas; nada físico tem “intencionalidade”, salvo quando a intencionalidade deriva de algum emprego dessa coisa física pela mente. Ou, pelo menos, se diz ser assim. Mas muito embora há já não esteja vivo.

isso é apressado; postular poderes misteriosos da mente nada resolve. O problema, contudo, é bastante real. Como são possíveis a intencionalidade e a referência?

As teorias mágicas da referência

Vimos que o “retrato” feito pela formiga não tem uma conexão necessária com Winston Churchill. O mero fato de esse “retrato” possuir uma “semelhança” com Winston Churchill não o torna uma imagem real, nem mesmo uma representação de Churchill. A menos que a formiga fosse inteligente (o que não é) e tivesse conhecimento de Churchill (o que não tem), a linha por ela traçada não é uma imagem e nem mesmo uma representação do que quer que seja. Alguns povos primitivos acreditam que algumas representações (*nomes*, em particular) têm uma conexão necessária com seus portadores; que saber o “verdadeiro nome” de alguém ou de algo lhe confere poder sobre eles. Esse poder advém da *conexão mágica* entre o nome e o portador do nome; mas uma vez que se percebe que um nome *apenas* tem uma conexão contextual, contingente e convencional com seu portador, fica difícil ver por que o conhecimento do nome deveria ter qualquer significado místico.

O que é importante perceber é que aquilo que vale para os retratos físicos também vale para as imagens mentais e também para as representações em geral; as representações mentais não têm uma conexão necessária com aquilo que representam, não mais do que as representação físicas o têm. A suposição contrária é uma volta ao pensamento mágico.

Talvez, no caso das *imagens* mentais, seja mais fácil de se compreender o que está em questão (talvez o primeiro filósofo a ter compreendido a enorme importância desse ponto, ainda que não o primeiro a efetivamente enfatizá-lo, foi Wittgenstein). Suponha que algures há um planeta onde os seres humanos tenham evoluído (ou sido deixados lá por alienígenas, ou a hipótese que mais lhe aprouver). Suponha que esses humanos, embora em muitos aspectos sejam como nós, nunca tenham visto *árvores*. Suponha que eles nunca tenham imaginado *árvores* (talvez a vida vegetal exista nesse planeta apenas em forma de bolor). Suponha que, um dia, um retrato de uma árvore caia acidentalmente de uma nave que por ali passava, sem qualquer contato com eles. Imagine-os intrigados com esse retrato. Que coisa no mundo é isso? Todo o tipo de especulação ocorre a eles: um edifício, um abrigo ou, até mesmo, alguma espécie de animal. Mas suponha que eles nunca tivessem se aproximado da verdade.

Para nós, o retrato é uma representação de uma árvore. Para esses humanos, o retrato representa apenas um estranho objeto de natureza e função desconhecidos. Suponha que

um deles, tendo visto o retrato, tenha uma imagem mental que seja exatamente como uma das minhas imagens mentais de uma árvore. A sua imagem mental não é um *representação de uma árvore*. É apenas uma representação de um objeto estranho (qualquer que seja) que o misterioso retrato representa.

Contudo, alguém poderia argumentar que a imagem mental seja, *de fato*, uma representação de uma árvore, pois, para início de conversa, o retrato que causou a imagem mental era, ele próprio, uma representação de uma árvore. Há uma cadeia causal das árvores reais à imagem mental, ainda que uma cadeia causal bastante estranha.

Poderíamos imaginar, contudo, a ausência dessa cadeia causal. Suponha que o “retrato da árvore” deixado pela nave espacial não era realmente um retrato de uma árvore, mas o resultado accidental de tinta derramada. Ainda que fosse exatamente como o retrato de uma árvore, não era, na verdade, um retrato de uma árvore, não mais do que a “caricatura” de Churchill feita pela formiga era um retrato de Churchill. Podemos mesmo imaginar que a espaçonave que deixou o “retrato” veio de um planeta que não conhecesse árvores. Assim, os humanos ainda teriam imagens mentais qualitativamente idênticas à minha imagem de uma árvore, mas que não seriam imagens que representassem árvores ou qualquer outra coisa mais.

O mesmo vale para as *palavras*. Um discurso no papel poderia parecer ser uma descrição perfeita de árvores, mas se tivesse sido produzido por macacos a bater aleatoriamente as teclas de uma máquina de escrever durante milhões de anos, então as palavras não se refeririam a qualquer coisa. Se houvesse uma pessoa que memorizasse essas palavras e as recitasse mentalmente sem entendê-las, então elas sequer se refeririam a algo quando pensadas na mente.

Imagine que a pessoa a dizer essas palavras tenha sido hipnotizada. Suponha que as palavras estejam em japonês, e que disseram a essa pessoa que ela entende japonês. Suponho que, ao pensar nessas palavras, ela tenha uma “sensação de compreensão” (muito embora, caso alguém interrompesse a sequência do seu pensamento e lhe perguntasse o que *significam* aquelas palavras, ela descobriria que não conseguiria responder). Talvez a ilusão fosse tão perfeita que a pessoa pudesse até tapear um telepata japonês! Mas se ela não conseguisse usar as palavras nos contextos corretos, responder às perguntas sobre o que ele estava “pensado”, etc., então ela não as teria entendido.

Combinando esses contos de ficção científica que acabei de contar, podemos inventar um caso no qual alguém pensa palavras que de fato são uma descrição de árvores em alguma linguagem e que simultaneamente tem imagens mentais apropriadas, mas que *nem* entende

as palavras e nem sabe o que seja uma árvore. Podemos imaginar que as imagens mentais tivessem sido causadas por respingos de tinta (embora a pessoa tenha sido hipnotizada para pensar que fossem imagens de algo apropriado a seu pensamento – algo que não seria capaz de dizer o que é, caso lhe fosse perguntado). E podemos imaginar que a linguagem na qual a pessoa está pensando é uma linguagem que nem o hipnotizador nem o hipnotizado tivessem ouvido falar – talvez fosse apenas uma coincidência que essas “frases sem sentido”, como supõe o hipnotizador, sejam uma descrição de árvores em japonês. Em suma, tudo o que se passa diante da mente da pessoa poderia ser qualitativamente idêntico ao que estivesse se passando na mente de um falante japonês que estivesse, *de fato*, pensando em árvores – mas não se referiria a árvores.

Tudo isso é realmente impossível, assim como é realmente impossível que macacos pudessem, por acaso, datilografar uma cópia de *Hamlet*. Isso significa que as probabilidades contra tal são tão altas, que nunca ocorrerá na realidade (pensamos nós). Mas não é logicamente impossível, e nem mesmo fisicamente impossível. *Poderia* acontecer (é compatível com as leis da física e, talvez, compatível com as condições efetivas do universo, caso houvesse seres inteligentes em outros planetas). E, se acontecesse, seria uma demonstração impressionante de uma verdade conceitual importante; que mesmo um sistema amplo e complexo de representações, tanto verbal quanto visual, não tem uma conexão mágica *intrínseca*, embutida, com aquilo que representa – uma conexão independente de como foi causado e de quais são as disposições do falante ou pensante. E isso é verdade, esteja o sistema de representações (palavras e imagens, no caso do exemplo) fisicamente realizado – as palavras estão escritas ou proferidas, e as imagens são imagens físicas –, esteja apenas realizado na mente. Palavras pensadas e retratos mentais não representam *intrinsecamente* aquilo a que se referem.

O caso dos cérebros numa cuba

Eis uma possibilidade de ficção científica que os filósofos discutem: imagine que um ser humano (você pode imaginar isso para si mesmo) tenha sido submetido a uma operação por um cientista maligno. O cérebro dessa pessoa (o seu cérebro) foi removido de seu corpo e colocado numa cuba com nutrientes que mantêm o cérebro vivo. As terminações nervosas foram conectadas a um supercomputador que causa na pessoa, cujo cérebro ela é, a ilusão de que tudo é perfeitamente normal. Parece haver pessoas, objetos, o céu, etc.; mas na verdade tudo o que as pessoas (você) estão experienciando é o resultado de impulsos elétricos

vijando do computador às terminações nervosas. O computador é tão engenhoso que se a pessoa tenta levantar a mão, um *feedback* logo o fará “ver” e “sentir” a mão sendo levantada. Ademais, o cientista maligno pode, variando o programa, fazer a vítima “experienciar” (ou alucinar) qualquer situação ou ambiente. Ele pode também obliterar a memória da operação cerebral, de modo que a vítima veja a si próprio com sempre tendo estado naquele ambiente. Pode até parecer à vítima que ela está sentada e lendo estas palavras sobre essa divertida, embora absurda, suposição de que há um cientista maligno que remove o cérebro das pessoas de seus corpos e os coloca numa cuba com nutrientes que mantém o cérebro vivo. As terminações nervosas, supostamente, estão conectadas num supercomputador que causa na pessoa, cujo cérebro ela é, a ilusão de que...

Esse tipo de possibilidade, quando mencionada numa aula de Teoria do Conhecimento, tem certamente o propósito de levantar o problema clássico do ceticismo sobre o mundo externo de maneira contemporânea (*como você sabe que não está nessa situação?*). Mas essa situação é também um artifício útil para se levantar questões sobre a relação mente/mundo.

Ao invés de ter apenas um cérebro numa cuba, poderíamos imaginar que todos os seres humanos (talvez todos os seres sencientes) são cérebros numa cuba (ou sistemas nervosos numa cuba, no caso de alguns seres com apenas um sistema nervoso mínimo contarem como “sencientes”). É claro que o cientista maligno teria de estar do lado de fora – ou onde estaria? Talvez não haja qualquer cientista maligno, talvez (embora absurdo) seja apenas o caso de o universo consistir de uma maquinaria automática que tende a uma cuba cheia de cérebros e sistemas nervosos.

Isso, por sua vez, nos leva a supor que a maquinaria automática foi programada para nos causar uma alucinação *coletiva*, ao invés de nos causar, separadamente, diversas alucinações não relacionadas. Assim, quando me parece que estou conversando com você, parece-te que você está ouvindo as minhas palavras. É claro que não é o caso que as minhas palavras chegam de fato aos seus ouvidos – pois você não tem ouvidos (reais), e nem eu tenho uma boca e uma língua reais. Antes, quando produzo minhas palavras, o que acontece é que os impulsos eferentes vão do meu cérebro até o computador, que faz com que eu “ouça” minha própria voz proferindo aquelas palavras e “sinta” a língua se movimentar, etc., e também faz com que você “ouça” as minhas palavras, “veja”-me falando, etc. Nesse caso, estamos, em certo sentido, nos comunicando de fato. Não estou errado sobre a sua existência real (apenas sobre a existência de seu corpo e do “mundo externo”, à parte os cérebros). De certo ponto de vista, não importa sequer que “todo o mundo” seja uma alucinação coletiva; pois, afinal de contas, você, de fato, houve as minhas palavras quando as digo a você, ainda que o mecanismo não seja aquele que supúnhamos ser (certamente, se estivéssemos a fazer

amor, e não apenas estivéssemos conversando, então a sugestão de que somos apenas dois cérebros numa cuba seria perturbadora).

Gostaria, agora, de fazer uma pergunta que parecerá um tanto tola e óbvia (pelo menos a alguns, incluindo alguns filósofos sofisticados), mas que rapidamente nos levará às profundezas filosóficas. Suponha que toda essa estória seja de fato verdadeira. Poderíamos, caso fôssemos cérebros numa cuba, *dizer* ou *pensar* que o somos?

Argumentarei que a resposta é “Não, não poderíamos”. Na verdade, argumentarei que a suposição de que somos, de fato, cérebros numa cuba, embora não viole qualquer lei da física e seja perfeitamente consistente com tudo o que experienciamos, não pode ser verdadeira. *Ela não pode ser verdadeira* pois é, de certo modo, autorrefutante.

O argumento que irei apresentar não é usual, e levei vários anos a me convencer de que ele estivesse realmente correto. Mas é um argumento correto. O que o torna tão estranho é que ele esteja conectado com algumas das mais profundas questões filosóficas (tal argumento me ocorreu pela primeira vez quando eu estava pensando sobre um teorema da lógica moderna, o “Teorema de Skolem-Löwenheim”, e, repentinamente, vi uma conexão entre esse teorema e alguns argumentos nas *Investigações Filosóficas* de Wittgenstein).

Uma “suposição autorrefutante” é aquela cuja verdade implica a sua própria falsidade. Por exemplo, considere a tese de que *todas as afirmações gerais são falsas*. Essa é uma afirmação geral. Portanto, se for verdadeira, tem de ser falsa. Por conseguinte, é falsa. Às vezes chama-se a uma tese “autorrefutante” se *a sua suposição, assentida ou enunciada*, implica a sua falsidade. Por exemplo, “Eu não existo” é autorrefutante se pensado por mim (para qualquer “mim”). Assim, alguém que pensa sobre si pode estar certo de que existe (como argumentou Descartes).

O que vou mostrar é que a suposição de que somos cérebros numa cuba tem exatamente essa propriedade. Se a pudermos considerar como verdadeira ou falsa, então ela não é verdadeira (mostrarei). Assim, ela não é verdadeira.

Antes de oferecer o argumento, consideremos por que parece tão estranho que tal argumento possa ser oferecido (pelo menos aos filósofos que subscrevem a concepção de verdade como “cópia”). Concedemos que é compatível com as leis da física que houvesse um mundo no qual todos os seres sencientes são cérebros numa cuba. Como dizem os filósofos, há um “mundo possível” no qual todos os seres sencientes são cérebros numa cuba (essa conversa sobre “mundo possível” soa como se houvesse um *lugar* onde qualquer proposição absurda fosse verdadeira, e é esse o motivo pelo qual isso possa ser tão enganador na filosofia). Os humanos, nesse mundo possível, têm exatamente a mesma experiência

que *nós* temos. Eles têm o mesmo pensamento que nós (pelo menos as mesmas palavras, imagens, formas mentais, etc. passam pelas suas mentes). Contudo, estou afirmando que há um argumento que possa ser oferecido que mostre que não somos cérebros numa cuba. Como? E por que não poderiam as pessoas do mundo possível que *são*, de fato, cérebros numa cuba oferecê-lo também?

A resposta será (basicamente) esta: muito embora as pessoas nesse mundo possível possam pensar e “dizer” quaisquer palavras que possamos pensar e dizer, elas não podem (defenderei) se *referir* ao que nós podemos. Em particular, elas não podem pensar ou dizer que são cérebros numa cuba (*mesmo ao pensar “nós somos cérebros numa cuba”*).

O teste de Turing

Suponha que alguém invente um computador que pode, de fato, ter uma conversa com alguém (com tantos sujeitos quanto uma pessoa inteligente poderia). Como se poderia decidir se tal computador é “consciente”?

O lógico britânico Alan Turing propôs o seguinte teste:³ faça com que alguém tenha uma conversa com o computador e com uma pessoa que ela não conheça. Se ela não conseguir dizer qual é o computador e qual é o humano, então (suponha que o teste possa ser repetido uma quantidade suficiente de vezes com diferentes interlocutores) o computador é consciente. Em suma, uma máquina de computar é consciente se puder passar no “Teste de Turing” (as conversas não são cara a cara, é claro, uma vez que o interlocutor não pode saber a aparência visual de ambos os seus parceiros. E nem se pode usar a voz, uma vez que a voz mecânica poderia soar diferente de uma voz humana. Imagine que o interlocutor digita as suas afirmações, perguntas, etc., e os dois parceiros – a máquina e a pessoa – respondem via um teclado eletrônico. Além disso, a máquina pode *mentir* – perguntada “Você é uma máquina”, ela poderia responder, “Não, sou um assistente daqui do laboratório”).

A ideia de que esse teste seja de fato um teste definitivo de consciência tem sido criticada por diversos autores (que não são por isso, a princípio, hostis à ideia de que uma máquina pudesse ser consciente). Mas esse não é o nosso tópico aqui. Gostaria de usar a ideia geral do teste de Turing, a ideia geral de um *teste dialógico de competência*, para um propósito diferente, o propósito de explorar a noção de *referência*.

3 - A. M. Turing, “Computing Machinery and Intelligence”, *Mind* (1950), reimpresso em A. R. Anderson (ed.), *Minds and Machines*.

Imagine uma situação na qual o problema não é o de determinar se o parceiro é de fato uma pessoa ou uma máquina, mas antes determinar se o parceiro usa as palavras para referir assim como nós. O teste óbvio é, novamente, começar uma conversa, e, caso nenhum problema surja, se o parceiro “passa” no sentido de ser indistinguível de alguém que fale certificadamente a mesma língua de maneira avançada, refere-se aos tipos usuais de objetos, etc., para concluir que o parceiro refere a objetos assim como nós. Quando o propósito do teste de Turing for como o descrito acima, isto é, determinar a existência de referência (compartilhada), vou me referir ao teste como *Teste de Turing para a Referência*. E, assim como os filósofos têm discutido a questão de se o teste de Turing original é um teste *definitivo* para a consciência, *i.e.* a questão de se uma máquina que “passa” no teste não apenas uma vez, mas regularmente, é *necessariamente* consciente, da mesma maneira quero discutir a questão se o Teste de Turing para a Referência anteriormente sugerido é um teste definitivo para a referência compartilhada.

A resposta será “Não”. O Teste de Turing para a Referência não é definitivo. Certamente que é um excelente teste na prática; mas não é logicamente impossível (embora seja altamente improvável) que alguém que pudesse passar no Teste de Turing para a Referência e não se referisse a qualquer coisa. Segue-se disso, como veremos, que podemos estender a nossa observação de que as palavras (e textos e discursos completos) não têm uma conexão necessária com os seus referentes. Ainda que consideremos não as palavras mas as regras que decidem quais palavras podem apropriadamente ser produzidas em certos contextos – ainda que consideremos, no jargão da computação, *programas para usar palavras* – a menos que os próprios programas *se refiram a algo extralinguístico*, não haverá, contudo, referência determinada que essas palavras possuam. Esse será um passo crucial no processo de atingir a conclusão de que os Habitantes do Mundo do Cérebro numa Cuba não podem se referir a qualquer coisa externa (e, por conseguinte, não podem dizer *que* são Habitantes do Mundo do Cérebro numa Cuba).

Suponha, por exemplo, que estou numa situação de Turing (jogando o “Jogo da Imitação”, na terminologia de Turing) e meu parceiro seja, de fato, uma máquina. Suponha que essa máquina seja programada para vencer o jogo (“passe” no teste). Imagine que a máquina foi programada para dar belas respostas em português às afirmações, perguntas, observações, etc., mas que não tem órgãos sensoriais (nada além das conexões com o meu teclado eletrônico) e nem órgãos motores (nada além do teclado eletrônico) (tanto quanto posso entender, Turing não supõe que a posse de órgãos dos sentidos ou de órgão motores seja necessário para consciência ou inteligência). Suponha que não apenas a máquina careça de olhos e ouvidos eletrônicos, etc., mas que também não há recursos no programa da máquina, o programa para jogar o Jogo da Imitação, para incorporar *inputs* de

tais órgãos, ou para controlar um corpo. O que dizer de tal máquina?

A mim parece evidente que não podemos e não deveríamos atribuir referência a tal dispositivo. É verdade que a máquina pode discursar belamente sobre, digamos, a vista na Nova Inglaterra. Mas não conseguiria reconhecer uma macieira ou uma maçã, uma montanha ou uma vaca, um campo ou um campanário, caso estivesse na frente de algum deles.

O que temos é um dispositivo para produzir frases em respostas a outras frases. Mas nenhuma dessas frases está de todo conectada ao mundo real. *Se ligássemos duas dessas máquinas e as deixássemos jogar o Jogo da Imitação uma com a outra, elas continuariam a “enrolar” uma a outra para sempre, ainda que o resto do mundo desaparecesse!* Não há mais razão para considerar a conversa da máquina sobre maçãs como se referindo a maçãs do mundo real do que há para considerar os “traços” da formiga com se referindo a Winston Churchill.

O que produz aqui a ilusão de referência, significado, inteligência, etc., é o fato de que há uma convenção de representação da qual *nós* aceitamos de que o discurso da máquina se refere a maçãs, campanários, a Nova Inglaterra, etc. Similarmente, há a *ilusão* de que a formiga fez uma caricatura de Churchill pela mesma razão. Mas nós somos capazes de perceber, manipular, lidar com maçãs e campos. A nossa conversa sobre maçãs e campos está intimamente conectada com as nossas transições *não-verbais* com as maçãs e os campos. Há “regras de entrada da linguagem” que nos levam de experiências de maçãs a elocuições como “Vejo uma maçã”, e “regras de saída da linguagem” que nos levam de decisões expressas na forma linguística (“Estou indo comprar maçãs”) a outras ações além da fala. Na ausência de regras de entrada ou saída da linguagem, não há razão para considerar a conversa da máquina (ou das duas máquinas, no caso em que pensamos nas duas máquinas jogando do Jogo da Imitação uma com outra) algo mais do que um jogo sintático. Um jogo sintático que *se assemelha* ao discurso inteligente, na verdade; mas apenas o tanto quanto (e não mais que isso) os traços feitos pela formiga se assemelham a uma sarcástica caricatura.

No caso da formiga, poderíamos ter argumentado que ela teria traçado a mesma linha, ainda que Winston Churchill nunca tivesse existido. No caso da máquina, não podemos usar exatamente o mesmo argumento; se as maçãs, as árvores, os campanários e campos não tivessem existido, então, presumivelmente, os programadores não teriam produzido esse mesmo programa. Embora a máquina não *perceba* maçãs, campos ou campanários, os seus criadores perceberam. Há uma conexão causal entre a máquina e as maçãs do mundo real, etc., por meio da experiência e conhecimento perceptual dos criadores. Mas tal conexão fraca dificilmente pode ser suficiente para a referência. Não apenas é logicamente possível, embora fantasticamente improvável, que a mesma máquina *pudesse* ter existido ainda que as maçãs, os campos e os campanários não; mais importante, a máquina é completamente

insensível à existência continuada das maçãs, dos campos, dos campanários, etc. Ainda que todas essas coisas *deixassem* de existir, a máquina ainda discursaria alegremente do mesmo jeito. É por isso que a máquina não pode ser considerada com se referindo a algo.

O ponto relevante para a nossa discussão é o de que nada há no Teste de Turing que exclua uma máquina programada para fazer nada *mais* do que jogar o Jogo da Imitação, e que exclua que tal máquina *claramente* não se refira a qualquer coisa, não mais do que um gravador se refere.

Cérebros numa cuba (novamente)

Comparemos os hipotéticos “cérebros numa cuba” com as máquinas que acabamos de descrever. Há, obviamente, diferenças importantes. Os cérebros numa cuba não têm órgãos sensoriais, mas têm *suporte* para tais órgãos; isto é, há terminações nervosas aferentes, há *inputs* dessas terminações nervosas aferentes, e esses *inputs* figuram no “programa” dos cérebros na cuba assim como figuram no programa dos nossos cérebros. Os cérebros numa cuba são *cérebros*; ademais, eles são cérebros *em funcionamento*, e funcionam por meio das mesmas regras que os cérebros do mundo efetivo funcionam. Por essas razões, pareceria absurdo negar consciência ou inteligência a eles. Mas o fato de que são conscientes e inteligentes não quer dizer que as suas palavras se refiram àquilo que as nossas palavras se referem. A questão que nos interessa é esta: as suas verbalizações contendo, digamos, a palavra “árvore” se referem realmente a *árvores*? De modo mais geral: eles podem se referir a objetos *externos* (em oposição, por exemplo, aos objetos na imagem produzida pela maquinaria automática)?

Para ajustar as nossas ideias, especifiquemos que a maquinaria automática veio a existir supostamente de algum tipo de acaso ou coincidência cósmica (ou, talvez, tenha sempre existido). Nesse mundo hipotético, a própria maquinaria automática supostamente não tem criadores inteligentes. Na verdade, como dissemos no início, podemos imaginar que todos os seres sencientes (ainda que minimamente sencientes) estão dentro da cuba.

Essa suposição não ajuda. Pois não há conexão entre a *palavra* “árvore” como usada por esses cérebros e as árvores reais. Eles continuariam a usar a palavra “árvore” como usam, teriam os pensamentos que têm, teriam as imagens que têm, ainda que não houvesse árvores reais. As suas imagens, palavras, etc. são qualitativamente idênticas às imagens, palavras, etc. que representam árvores no *nosso* mundo; mas já vimos (novamente a formiga!) que similaridade qualitativa a algo que representa um objeto (Winston Churchill ou uma

árvore) não faz de algo uma representação por si próprio. Em suma, os cérebros numa cuba não estão pensando sobre árvores reais quando pensam “há uma árvore à minha frente”, porque nada há em virtude do qual o seu pensamento “árvore” represente árvores reais.

Caso isso pareça apressado, reflita sobre o seguinte: vimos que as palavras não necessariamente se referem a árvores, ainda que estejam dispostas numa sequência que seja idêntica a um discurso que (estivesse a ocorrer em uma de nossas mentes) fosse inquestionavelmente *sobre árvores* no mundo efetivo. Nem o “programa”, no sentido das regras, práticas, disposições dos cérebros ao comportamento verbal, refere-se necessariamente a árvores ou produz a referência a árvores por meio das conexões que estabelece entre palavras e palavras, ou sugestões *linguísticas* e respostas *linguísticas*. Se esses cérebros pensam sobre, referem-se a, representam árvores (árvores reais, fora da cuba), então isso tem de ser por causa do modo como o “programa” conecta o sistema da linguagem a *inputs* e *outputs* não-verbais. Há, de fato, tais *inputs* e *outputs* não-verbais no Mundo do Cérebro numa Cuba (as terminações nervosas eferentes e aferentes novamente!), mas também vimos que os “dados dos sentidos” produzidos pela maquinaria automática não representam árvores (ou algo externo) mesmo quando eles se assemelham exatamente às nossas imagens de árvores. Assim como um salpico de tinta poderia se assemelhar a um retrato de uma árvore sem *ser* um retrato de uma árvore, do mesmo modo, vimos, um “dado do sentido” poderia ser qualitativamente idêntico a uma “imagem de uma árvore” sem ser uma imagem de uma árvore. Como é que pode o fato de, no caso dos cérebros numa cuba, a linguagem conectada pelo programa com os *inputs* sensoriais que não representam intrínseca ou extrinsecamente árvores (ou algo externo) faça com que todo o sistema de representações, a linguagem em uso, refira-se a ou represente árvores ou algo externo?

A resposta é que não pode. Todo o sistema de dados dos sentidos, de sinais motores às terminações eferentes, e de pensamento verbalmente ou conceitualmente mediado conectado pelas “regras de entrada da linguagem” aos dados dos sentidos (ou seja lá o que for) como *inputs* e pelas “regras de saída da linguagem” aos sinais motores como *outputs*, não tem mais conexão com *árvores* do que as linhas feitas pela formiga tem com Winston Churchill. Uma vez que percebemos que a *similaridade qualitativa* (denotando, caso queiras, identidade qualitativa) entre os pensamentos dos cérebros numa cuba e os pensamentos de alguém no mundo efetivo já não implica o compartilhamento de referência, não é difícil ver que não há bases de todo em todo para se considerar o cérebro numa cuba como se referindo a coisas externas.

A premissa do argumento

Ofereci o argumento prometido para mostrar que os cérebros numa cuba não podem pensar ou dizer que são cérebros numa cuba. Falta apenas torná-lo explícito e examinar a sua estrutura.

Por aquilo que foi dito, quando o cérebro numa cuba (no mundo onde todo ser senciente é e sempre foi uma cuba) pensa “Há uma árvore à minha frente”, o seu pensamento não se refere às árvores reais. De acordo com algumas teorias que discutiremos, poderia se referir às árvores na imagem, ou às características do programa que são responsáveis pelos impulsos elétricos. Essas teorias não são excluídas pelo que foi dito, pois há uma conexão causal próxima entre o uso da palavra “árvore” no português da cuba e a presença das árvores na imagem, a presença dos impulsos elétricos de certo tipo, e a presença de certas características no programa da máquina. De acordo com essas teorias, o cérebro está *correto*, e não *errado* em pensar “Há uma árvore à minha frente”. Dado ao que “árvore” se refere no português da cuba e ao que “em frente de” se refere, supondo que uma dessas teorias esteja correta, então as condições de verdade para “Há uma árvore à minha frente” quando ocorre no português da cuba são simplesmente as de que uma árvore na imagem esteja “em frente a” o “mim” em questão – na imagem – ou, talvez, que o tipo de impulso elétrico que normalmente produz essa experiência esteja vindo da maquinaria automática, ou, talvez, que a característica da maquinaria que supostamente produz a experiência da “árvore em minha frente” esteja operando. E essas condições de verdade são certamente cumpridas.

Pelo mesmo argumento, “cuba” se refere, no português da cuba, a cubas na imagem, ou a algo de algum modo relacionado (impulsos elétricos ou características de programa), mas não certamente a cubas reais, uma vez que o uso de “cuba” no português da cuba não tem conexão causal com as cubas reais (com exceção da conexão em que os cérebros numa cuba não seriam capazes de usar a palavra “cuba”, caso não fosse pela presença de uma cuba particular – a cuba em que eles estão; mas essa conexão é obtida entre o uso de *cada* palavra no português da cuba e essa cuba particular; não é uma conexão especial entre o uso da palavra *particular* “cuba” e as cubas). Similarmente, “fluido nutricional”, no português da cuba, refere-se a um líquido, ou a algo relacionado (impulsos elétricos ou características do programa). Segue-se disso que se tal “mundo possível” for de fato o mundo efetivo, e nós formos de fato cérebros numa cuba, então aquilo que queremos dizer por “somos cérebros numa cuba” é que *somos cérebros numa cuba na imagem* ou algo do tipo (se é que dizemos algo). Mas parte da hipótese de que somos cérebros numa cuba é que não somos cérebros numa cuba na imagem (*i.e.* aquilo que estamos a “alucinar” não é que somos cérebros numa cuba). Assim, se somos cérebros numa cuba, então a frase “Somos cérebros numa cuba” diz

algo falso (se é que o diz). Em suma, se somos cérebros numa cuba, então “Somos cérebros numa cuba” é falsa. Portanto, é (necessariamente) falsa.

A suposição de que tal possibilidade faz sentido surge da combinação de dois erros: (1) levar demasiadamente a sério a *possibilidade física*; e (2) usar inconscientemente uma teoria mágica da referência, uma teoria segundo a qual certas representações mentais referem-se necessariamente a coisas e tipos de coisas externos.

Há um “mundo fisicamente possível” no qual somos cérebros numa cuba – o que isso significa, exceto que há uma *descrição* de tal estado de coisas que é compatível com as leis da física? Assim como há uma tendência em nossa cultura (e tem sido assim desde o século dezessete) de considerar a *física* como a nossa metafísica, isto é, ver as ciências exatas como a tão estimada descrição da “verdadeira e última mobilha do universo”, há, como uma consequência imediata, uma tendência a considerar a “possibilidade física” com a própria pedra de toque daquilo que poderia realmente ser de fato o caso. De acordo com tal perspectiva, a verdade é a verdade física; a possibilidade, a possibilidade física; e a necessidade, a necessidade física. Mas acabamos de ver, mesmo que apenas no caso de um exemplo bastante artificial, que tal perspectiva está errada. A existência de um “mundo fisicamente possível” no qual somos cérebro numa cuba (e sempre fomos e sempre seremos) não significa que realmente, efetivamente, pudéssemos *ser* cérebros numa cuba. O que exclui essa possibilidade não é a física, mas a *filosofia*.

Alguns filósofos, ávidos em asseverar e minimizar as pretensões de sua profissão (a típica mentalidade da filosofia anglo-saxônica no século vinte), diriam: “Certo. Você mostrou que algumas coisas que parecem ser possibilidades físicas são impossibilidades *conceituais*. O que há de tão surpreendente nisso?”

Ora, para ser sincero, o meu argumento pode ser descrito como um argumento “conceitual”. Mas descrever a atividade filosófica como a busca por verdades “conceituais” faz tudo soar como se fosse uma *investigação sobre o significado das palavras*. E isso não é de todo o que estivemos a fazer.

Aquilo que estivemos a fazer foi considerar as *precondições* para o *pensamento acerca de algo*, para a *representação*, para a *referência*, etc. Investigamos essas precondições não por meio da investigação do significado dessas palavras e expressões (como um linguista faria, por exemplo), mas pelo *raciocínio a priori*. Não no velho sentido “absoluto” (uma vez que não afirmamos que as teorias mágicas da referência estão *a priori* erradas), mas no sentido de investigar aquilo que é *razoavelmente* possível *se aceitarmos* certas premissas gerais, ou fazer suposições teóricas bastante amplas. Tal procedimento não é nem “empírico” nem

completamente *a priori*, mas tem elementos de ambos os modos de investigação. A despeito da falibilidade do meu procedimento, e sua dependência de suposições que poderiam ser descritas como “empíricas” (e.g. a suposição de que a mente não tem acesso a propriedades ou coisas externas, a menos que fornecido pelos sentidos), ele tem uma relação íntima com aquilo que Kant chamou de investigação “transcendental”; pois é uma investigação, repito, das *precondições* da referência e, por conseguinte, do pensamento – precondições embutidas na natureza das nossas próprias mentes, embora não completamente (como Kant esperava) independente de suposições empíricas.

Uma das premissas do argumento é óbvia: as teorias mágicas da referência estão erradas, e não apenas erradas para as representações físicas, mas também para as mentais. A outra premissa é que não podemos referir a certos tipos de coisas, e.g. *árvores*, se não tivermos qualquer interação causal com elas,⁴ ou com as coisas em termos das quais elas podem ser descritas. Mas por que deveríamos aceitar essas premissas? Uma vez que constituem o enquadramento amplo no qual estou argumentando, é hora de examiná-las mais de perto.

As razões para rejeitar as conexões necessárias entre as representações e seus referentes

Mencionei anteriormente que alguns filósofos (o mais famoso deles, Brentano) atribuíram um poder à mente, a “intencionalidade”, que é precisamente aquilo que nos permite *referir*. Evidentemente, rejeitei isso como solução. O que, contudo, me dá esse direito? Talvez eu não tenha sido apressado demais?

Esses filósofos não afirmam que podemos pensar sobre coisas ou propriedades externas sem usar representações de todo em todo. E o argumento que ofereci acima compara os dados dos sentidos visuais com o “retrato” feito pela formiga (o argumento no conto de ficção científica sobre o “retrato” de uma árvore, surgido de um respingo de tinta, e que produziu os dados dos sentidos qualitativamente similares às nossas “imagens visuais de árvores”, mas que vêm desacompanhados de qualquer *conceito* de árvore) teria sido aceito como mostrando que essas *imagens* não necessariamente referem. Se há representações mentais que necessariamente referem (a coisas externas) elas têm de ser da mesma natureza dos *conceitos* e não da natureza de imagens. Mas o que são *conceitos*?

⁴ - Se os cérebros numa cuba terão conexão causal com árvores, digamos, *no futuro*, então, talvez, eles possam se referir agora a árvores por meio da descrição “as coisas às quais irei me referir como ‘árvores’ em tal e tal tempo futuro”. Mas estamos a imaginar um caso no qual os cérebros numa cuba nunca saíram da cuba e, por isso, *nunca* tiveram conexão causal com árvores, etc.

Ao introspectarmos, não percebemos “conceitos” fluindo em nossas mentes como tais. Interrompa o fluxo do pensamento quando ou onde quiser e tudo o que teremos serão palavras, imagens, sensações, sentimentos. Quando penso em voz alta, não penso os meus pensamentos duas vezes. Escuto as minhas palavras assim como você. Para dizer a verdade, é diferente para mim quando profiro palavras que acredito e quando profiro palavras que não acredito (mas, às vezes, quando estou nervoso, ou frente a uma audiência hostil, sinto como se eu estivesse a mentir quando sei que estou dizendo a verdade); e é diferente quando profiro palavras que entendo e quando profiro palavras que não entendo. Mas posso imaginar sem dificuldade alguém a pensar exatamente essas palavras (no sentido de dizê-las em sua mente) e a ter exatamente a sensação de entendimento, de asserção, etc. que tenho, mas que perceba, logo em seguida (ou ao ser despertado por um hipnotizador), que não entendeu aquilo que acabou de passar em sua mente, que nem sequer entendeu a língua na qual essas palavras estão. Não digo que isso seja muito provável; digo simplesmente que nada há de ininteligível com isso. E o que isso mostra não é que conceitos *são* palavras (ou imagens, ou sensações, etc.), mas que atribuir um “conceito” ou um “pensamento” a alguém é completamente diferente de atribuir uma “apresentação” mental, alguma entidade ou evento introspectível, a essa pessoa. Conceitos não são apresentações mentais que intrinsecamente referem a objetos externos pela mesma razão decisiva que não são sequer representações mentais. Conceitos são signos usados de certas formas; os signos podem ser públicos ou privados, entidades mentais ou físicas, mas mesmo quando são “mentais” e “privados”, o próprio signo, à parte de seu uso, não é um conceito. E signos em si mesmos não referem intrinsecamente.

Podemos ver isso por meio de um experimento mental bastante simples. Suponha que você é como a mim mesmo e não consegue distinguir um olmo de uma faia. Dizemos, contudo, que a referência de “olmo” em minha fala é a mesma referência de “olmo” na fala do restante das pessoas, *viz.* olmos, e que o conjunto de todas as faias é a extensão de “faia” (*i.e.* o conjunto das coisas dais quais “faia” é verdadeiramente predicada) tanto na sua fala quanto na minha. É realmente crível que a diferença entre aquilo a que “olmo” se refere e aquilo a que “faia” se refere seja produzida por uma diferença nos nossos *conceitos*? O meu conceito de olmo é exatamente o mesmo que meu conceito de faia (tenho de confessar) (isso mostra que a determinação da referência é social e não individual, a propósito; você e eu diferimos dos especialistas que *podem* distinguir olmos de faias). Se alguém heroicamente tentar manter que a diferença entre as referências de “olmo” e “faia” na *minha* fala é explicada pela diferença em meu estado psicológico, então façamos com que ele imagine uma Terra Gêmea onde as palavras são trocadas. A Terra Gêmea é muito parecida com a Terra; na verdade, com a exceção de que “olmo” e “faia” são intercambiáveis, o leitor pode supor que

a Terra Gêmea é exatamente como a Terra. Suponha que eu tenha um *Doppelganger* na Terra Gêmea que é, molécula a molécula, idêntico a mim (no sentido em que duas gravatas podem ser “iguais”). Se você é um dualista, então suponha que o meu *Doppelganger* tem os mesmos pensamentos verbalizados que eu, que tem os mesmo dados dos sentidos, as mesmas disposições, etc. É absurdo pensar que o seu estado psicológico seja um pouco diferente do meu: contudo, a sua palavra “olmo” representa *faias*, e a minha palavra “olmo” representa olmos. (Similarmente, se a “água” na Terra Gêmea for um líquido diferente – digamos, XYZ e não H₂O – então “água” representa um líquido diferente quando usada na Terra Gêmea e quando usada na Terra, etc.). Contrário a uma doutrina que tem nos acompanhado desde o século dezessete, *os significados simplesmente não estão na cabeça*.

Vimos que possuir um conceito não é uma questão de possuir imagens (digamos, de árvores – ou mesmo imagens, “visuais” ou “acústicas”, de frases, ou de discursos completos), uma vez que podemos possuir qualquer sistema de imagens e não possuir a *habilidade* de usar as frases em situações apropriadas (considere tanto fatores linguísticos – aquilo que foi dito antes – quanto fatores não linguísticos como aquilo que determina o que é “situacionalmente apropriado”). Um homem pode ter todas as imagens que você quiser e, ainda assim, ficar completamente perdido se alguém lhe disser “aponte-me uma árvore”, ainda que haja muitas árvores ao redor. Ele pode até ter a imagem daquilo que tem de fazer e, ainda assim, não saber o que tem de fazer. Pois a imagem, se não for acompanhada pela habilidade de agir de certo modo, é apenas uma *imagem*, e agir de acordo com uma imagem é em si uma habilidade que alguém pode ter ou não. (O homem poderia imaginar-se a apontar para uma árvore, mas apenas a fim de contemplar algo logicamente possível; ele próprio apontando para a árvore depois que alguém produziu a sequência de sons – para ele destituída de sentido – “aponte-me uma árvore, por favor.”). Ele ainda não saberia que tinha de apontar para uma árvore, e ainda não *entenderia* “aponte-me uma árvore”.

Considerarei a habilidade de usar certas frases como sendo o critério para a posse de um conceito bem desenvolvido, mas isso pode ser facilmente liberalizado. Poderíamos permitir que o simbolismo consistisse de elementos que não são palavras na linguagem natural, por exemplo, e poderíamos permitir tais fenômenos mentais como imagens e outros tipos de eventos internos. O que é essencial é que tais coisas tivessem a mesma complexidade, a habilidade de serem combinados uns com os outros, etc., que as palavras de uma linguagem natural têm. Pois, embora uma apresentação particular – de um *flash* azul, digamos – pudesse servir a um matemático particular como uma expressão interna de uma prova completa do Teorema dos Números Primos, ainda não haveria a tentação de dizer isso (e seria falso dizê-lo), caso o matemático não pudesse apresentar o seu “flash azul” em passos separados e conexões lógicas. Mas não importa que tipo de fenômenos internos

aceitamos como possíveis *expressões* de pensamento, argumentos exatamente similares ao anterior mostrarão que não são os fenômenos em si mesmos que constituem o significado, mas, antes, a habilidade do sujeito em *empregar* esses fenômenos, de produzir o fenômeno correto nas circunstâncias corretas.

Isso é uma versão bastante abreviada do argumento de Wittgenstein nas *Investigações Filosóficas*. Se estiver correto, então a tentativa de entender o pensamento por meio daquilo a que se chama investigação “fenomenológica” está fundamentalmente errado; pois o que os fenomenólogos não conseguem ver é que aquilo que eles descrevem é a *expressão* interna do pensamento, mas que o *entendimento* da expressão – o entendimento que alguém tem de seus próprios pensamentos – não é uma *ocorrência*, mas, sim, uma *habilidade*. O nosso exemplo de um homem fingindo pensar em japonês (e enganando o telepata japonês) já mostrou a futilidade de uma abordagem fenomenológica ao problema do *entendimento*. Pois, ainda que haja alguma qualidade introspectível que se apresente quando, e apenas quando, alguém *de fato* entenda (isso parece falso na introspecção, na verdade), ainda assim essa qualidade *se relaciona* apenas ao entendimento, e é ainda possível que o homem tapeando o japonês tenha também essa qualidade e, contudo, não entenda uma palavras de japonês.

Por outro lado, considere a hipótese, perfeitamente possível, de um homem que não tem qualquer “monólogo interior”. Ele fala um inglês perfeitamente bom, e, caso peçam a sua opinião sobre um dado assunto, ele a emitirá precisamente. Mas ele nunca pensa (em palavras, imagens, etc.) quando não está falando em voz alta; nem algo “passa pela sua cabeça”, exceto (claro) que ele ouve a própria voz ao falar, e tem as impressões comuns de seus arredores, e mais uma “sensação de entendimento” geral (talvez ele tenha o hábito de falar consigo próprio). Quando ele escreve uma carta ou vai ao mercado, etc., ele não está tendo um “fluxo de pensamento” interno; mas as suas ações são inteligentes e dotadas de propósito, e se alguém vai até ele e pergunta “O que você está fazendo?” ele dará uma resposta perfeitamente coerente.

Esse homem parece perfeitamente imaginável. Ninguém hesitaria em dizer que ele fosse consciente, que não curtisse rock’n’roll (caso expressasse com frequência uma forte aversão ao rock’n’roll, etc., só porque ele não tem pensamentos conscientes, exceto quando fala em voz alta.

O que se segue disso é que (a) nenhum conjunto de eventos mentais – imagens ou acontecimentos e qualidades mentais mais “abstratos” – *constitui* o entendimento; e (b) nenhum conjunto de eventos mentais é *necessário* para o entendimento. Em particular, *conceitos não podem ser idênticos a objetos mentais de qualquer tipo*. Pois, supondo que por um objeto mental estamos designando algo introspectível, já vimos que, seja ele o que for, pode estar ausente num homem que entende a palavras apropriadas (e, por isso, tem o conceito

bem desenvolvido), e presente num homem que não tem conceito de todo em todo.

Voltando agora à nossa crítica das teorias mágicas da referência (um tópico que também interessava a Wittgenstein), vimos que, por um lado, aqueles “objetos mentais” que *podemos* detectar introspectivamente – palavras, imagens, sensações, etc. – não referem intrinsecamente, não mais do que o retrato feito pela formiga (e pelas mesmas razões), ao passo que as tentativas de postular objetos mentais, “conceitos”, que têm uma conexão necessária com os seus referentes, e que apenas fenomenólogos treinados podem detectar, cometem uma asneira *lógica*; pois conceitos são (pelo menos em parte) *habilidades*, e não ocorrências. A doutrina de que há apresentações mentais que necessariamente referem a coisas externas não é apenas má ciência natural; é também má fenomenologia e confusão conceitual.

Resenha

Schopenhauer e os Filósofos, de Hélio Lopes da Silva

Review: Hélio Lopes da Silva's "Schopenhauer and the Philosophers"

*José Luiz Furtado
Universidade Federal de Ouro Preto*

Qual o valor de uma obra de filosofia? Mapear sua importância, aferindo citações e influências neste ou naquele filósofo, pode ser uma via a ser seguida. Mas importaria distinguir, aqui, importância e valor, porque a primeira é relativa, e o segundo não. O valor de uma obra de filosofia, enquanto invólucro de um pensamento, nada tem a ver com sua importância histórica, por uma razão, ela própria, filosófica. “É sábio ouvir não a mim, mas ao logos”, afirmou Heráclito. O discurso humano revela sua essência racional na medida em que se deixa atravessar pela verdade, fazendo dela seu veículo no mundo. A verdade é, pois, a única medida da efetiva importância que devemos conferir a qualquer texto de filosofia, e o único fundamento filosófico da sua apreciação valorativa.

Nesta medida, *Schopenhauer e os filósofos*, de Hélio Lopes da Silva, cumpre a função essencial de fazer sobressair à força do texto seminal contra as apropriações posteriores, confrontando o filósofo alemão com aqueles que, de uma de outra forma, apoiaram-se em seu pensamento. Os confrontos já estavam ali, postos e dispostos pela história, disseminados por meio de vários pensadores e filósofos influenciados pelo mestre de *Mundo e Representação*. Hélio Lopes da Silva apenas os trouxe de novo à tona, reunindo-os em um só livro. Pelo jogo dialético das oposições e das nem sempre possíveis sínteses, dos diálogos às vezes cordiais, outras vezes tensos, vemos sobressair a fisionomia de uma obra à qual nem sempre

se reconhece o caráter de expressar um pensamento propriamente filosófico. “Maladroite” (p. 28) (desajeitado, inconveniente), Schopenhauer seria este filósofo que, acima de tudo, não compreendeu a si mesmo e à sua própria filosofia. “É preciso separar, na consideração de sua filosofia, entre sua intuição original, que é uma contribuição significativa para o pensamento filosófico, e a expressão desastrada desta intuição, a sua acomodação difícil, e no fundo incoerente, a um arranjo conceitual que lhe é inóspito” (p. 29). Porque, segundo o autor, Schopenhauer criticava seus interlocutores principais, Hegel, Kant e Shelling, como se fossem seus competidores em mesmo campo conceitual, quando, na verdade, abria um horizonte original. O que dizer da compreensão histórica da obra de um autor que sequer compreendeu a si mesmo: “maladroite”!

Entre os parceiros solicitados à cena do debate pelo livro de Hélio Lopes da Silva encontramos, no capítulo 4, a figura de Nietzsche. O autor assume, ali, a personalidade de Schopenhauer, para dialogar com o filósofo da vontade de potência e, ficticiamente, responder a possíveis e indevidas apropriações da sua obra. Porque se, mais do que tudo, Schopenhauer equivocou-se sobre a dimensão dos conceitos por ele próprio desenvolvidos, o que dirá dos que vieram depois se arvorando em seus intérpretes e herdeiros. A estratégia retórica dá gosto ao texto que, em nenhum momento, perde a concisão conceitual por render-se ao sabor do estilo. Um pouco de ironia cai bem nesta hora, se é verdade que Nietzsche, com sua leitura intempestiva, nada mais fez do que comparar-se pessoalmente a Schopenhauer: “entre Nietzsche e Schopenhauer trata-se (...) não da confrontação entre teses opostas candidatas à verdade, mas sim da aceitação e valoração moral de uma atitude comparativamente a outra atitude” (p. 52).

Mas, quanto mais nos distanciamos na história da recepção da obra de Schopenhauer, mais encontramos distorções que dificultam sua assimilação. Mesmo nas referências de Freud a Kant, que ele leu, seguramente, no tempo em que estudava com Brentano em Viena¹, podemos encontrar “a influência da inflexão psico-antropológica” através da qual Schopenhauer teria desvirtuado o fundamento transcendental da filosofia de Kant” (p. 75). Mas Lopes da Silva não se limita a meramente constatar, no caso, a afirmações de COHEN, segundo as quais Schopenhauer teria distorcido o pensamento de Kant. Com erudição e domínio da obra teórica deste último, o autor mostra que, na verdade, foi Cohen quem compreendeu mal “aquilo que está sendo contestado por Schopenhauer” na obra de Kant (p. 88). A querela, no entanto, vai além da obra de Schopenhauer, valendo como ocasião para Lopes revisar os conceitos mais fundamentais da metodologia crítica, de modo que o capítulo vale também para os estudiosos do kantismo que pretendem abordar o pensamento de Kant pela vida do confronto com seus comentadores, que abrange, inclusive, a Heidegger.

1 - Provavelmente na mesma época em que Husserl.

É assim que a leitura schopenhauriana de Kant, e a defesa da compreensão do primeiro da obra do segundo, conduz Lopes da Silva até o famoso Kantbuch de Heidegger. Nessa obra, o autor de *Ser e Tempo* tenta levar a cabo uma leitura fenomenológica da primeira crítica, privilegiando a imaginação transcendental e seu papel de destaque, notadamente na edição A, na arquitetura do pensamento kantiano, interpretado enquanto pensamento da finitude. Pois Lopes da Silva nos traz justamente a estranheza do fato de Heidegger jamais mencionar ter sido Schopenhauer o primeiro leitor a denunciar em Kant este “recuo” diante das consequências metafísicas do seu conceito de imaginação na *Crítica da Razão Pura* (p. 98). “Esta proximidade entre a crítica de Schopenhauer e a proposta do Kant-Bush de Heidegger torna-se ainda mais evidente quando consideramos que aquele ser racional finito, mas não sensível, aquele ser moral em função do qual Kant teria recuado diante da Imaginação, encontra por parte de Schopenhauer uma inequívoca e peremptória rejeição” (p. 98).

A partir daí, Lopes da Silva parte para uma análise extremamente oportuna, porque recenseia o estágio em que se encontram as pesquisas, hoje, em torno da temática - a meu ver a mais importante da primeira Crítica - da “apercepção transcendental”, do “eu penso que deve poder acompanhar toda representação”. A novidade do expediente teórico adotado está em fazê-lo tendo sempre como pano de fundo não apenas a obra de Schopenhauer, mas sua recepção contemporânea. O que importa, como mostra Lopes da Silva, não é de forma de nenhuma o que Schopenhauer teria realmente dito. Segundo uma forma típica de querela filosófica fadada, já de antemão, ao fracasso, na medida em que as intenções dos autores sempre já se perderam desde que foram escritas. Diante do seu texto, cada autor é nada mais que mais um intérprete. E nem sempre o mais abalizado, como o mostra a “maladroitness” de Schopenhauer, a propósito das suas próprias ideias. Sendo assim, justifica-se que o livro de Lopes da Silva possa, em longos trechos, abandonar a obra do autor de *O mundo como vontade e representação*, entregando-se totalmente a discussões que tangem o cerne próprio ao pensamento filosófico como tal, interessando inclusive aos leitores de Husserl e Merleau-Ponty.

De fato, a discussão em torno da teoria kantiana da finitude do conhecimento humano enquanto exigência de que os objetos nos sejam dados através da experiência, intuitivamente, sob a forma de fenômenos, antes que possam ser julgados mediante conceitos e assim conhecidos, termina conduzindo a uma reflexão sobre o estatuto mesmo dos fenômenos e sua doação. Não haveria “antes da análise do que é dado na sensibilidade”, pergunta o autor através de uma citação de Longuenesse, comentador de Kant, uma “síntese que precisa ter ocorrido antes da análise” e “exigida pela representação sensível dos *x*'s que podem ser refletidos sob conceitos”? (p. 102). Mediante essa reflexão, acompanhamos as dificuldades da passagem de dados perceptivos originariamente caóticos “as sensações puras” ao estatuto

de conteúdo material da intuição dos objetos do conhecimento. Diante desse dilema, o kantismo nada mais faz do que multiplicar, em uma ordem ascendente de formalizações, as diversas sínteses que, ao final, culminarão na arquitetônica das ideias da razão. A prova de que Schopenhauer está mesmo à frente das questões da sua época, no sentido de levá-las mais adiante e de elevá-las ao um patamar de interrogação mais exato, está na forma como supera essa indecisão ou, pode ser, lacuna do sistema kantiano. No entanto, como que envergonhado de exibir suas próprias reflexões, o autor fala através de uma fictícia voz tomada a Schopenhauer ele mesmo. “A réplica que Schopenhauer, no entanto, *realmente faria (...)*” (p. 108, grifado por mim) é na verdade a que Lopes efetivamente faz por si mesmo, embora exponha seu pensamento próprio, na ânsia de dar um suporte credível ao seu subterfúgio, como se fosse uma citação. “Entender o objeto (Gegenstand) real da intuição empírica como produto de uma auto-consciência inexoravelmente conceitual, mesmo que nestas outras acepções da expressão conceito, além de retirar dos animais a capacidade de tal intuição, faz da inteligência humana uma atividade essencialmente teórica, ao passo que de fato, ela é uma atividade essencialmente prática ou pragmática” (p. 108).

Não sabemos se, com isso, restituímos aos animais a capacidade que lhes fora furtada, pelo kantismo, de intuir objetos, na medida que retiramos da percepção sensível qualquer atividade conceitual mesmo coadjuvante, mas, seguramente, plasmamos a reflexão em um âmbito agora não mais exclusivamente teórico ou lógico. Anteriormente, Lopes havia citado a Nietzsche quando este afirma que a moral kantiana possuía mãos limpas - na medida em que não implica o comprometimento com as consequências das nossas ações no mundo, restringindo-se apenas à determinação interior dos princípios da vontade - porque não tinha mãos. O mesmo podemos dizer do “sujeito transcendental”, que nada mais é do que a representação do conjunto das condições aprióricas de representação do objeto do conhecimento. Mas agora - e aqui o elogio de Nietzsche ao seu mestre torna-se compreensível - os objetos não são mais imediatamente destinados a serem elaborados conceitualmente em conceitos e conhecidos teoricamente. O homem, antes de ser inteligência teórica, é vontade e sentimento, e assim que se relaciona originariamente ao mundo. Esse mundo “mais velho do que todo pensamento”, como diria Merleau-Ponty. Esse mundo e esse “corpo” que Schopenhauer introduz de chofre na reflexão sobre o sujeito em uma época em que esse nome designava a matéria” em movimento no espaço. Ao contrário, afirma Lopes, a doação originária do objeto ocorre como “vontade objetivada no corpo do sujeito individual que o conhece” (p. 109). E tudo isso se dá porque a atitude do homem comum que a filosofia tem por missão combater, mas da qual ela tem o dever de partir e por objetivo compreender, julga ou conceitua antes de ver. De modo que, embora Kant tenha feito da contemplação desinteressada o cerne da sua teoria estética, deixa escapar no

entanto que esse desinteresse desde que considerado no plano teórico. Deve ser um ponto de chegada e não de partida, como, aliás, Heidegger tentará fazer em *Ser e tempo* com as categorias de *Vondeheit* e *Zunderheit* (ver). Então não é, de forma nenhuma, que as críticas de Schopenhauer ainda “aguardem respostas” (p. 111). As respostas estão aí nas filosofias fenomenológicas florescentes, desde Husserl. Mas esquecidas deste que foi um antecedente e um inaugurador de questões que soaram, no entanto, como uma música para a qual sua época ainda não tinha desenvolvido ouvidos suficientemente capazes de entendê-la, soando, por isso, como melodia dissonante. Porque, talvez, as múltiplas leituras da obra de Schopenhauer sobre as quais Lopes se debruça como grande erudição, concisão de estilo e inteligência refinadíssima, tenham feito a propósito como o “homem comum” ao qual o filósofo se refere numa das últimas passagens citadas no livro, como aquele “preguiçoso” que procura uma cadeira para sentar-se deixando-a logo em seguida ao seu uso por não mais interessá-lo. Como esse homem comum, incapaz de “demorar-se na intuição”, assim também procedem os diversos leitores apressados em enquadrar Schopenhauer em um conceito - ou preconceito - que ele próprio no entanto se encarrega de sugerir: “maladrite”.