Revista de Pesquisa em Filosofia

FUNDAMENTO

Ouro Preto – Minas Gerais – Brasil

Fundamento Revista de Pesquisa em Filosofia / Universidade Federal de Ouro Preto / Grupo de Pesquisa em Filosofia Contemporânea. v. 1 n. 3, (maio – ago. 2011) – Ouro Preto:

Ed. UFOP, 2011

Quadrimestral – Tiragem: 200 exemplares.

ISSN: 2177–6563

1. Filosofia - Pediódicos. 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

SUMÁRIO

CONVITE À PESQUISA EM FILOSOFIA

ARTIGOS INÉDITOS

RESENHA

TRADUÇÃO

Convite à Pesquisa em Filosofia

O artigo "Sobre o sentido e a referência" de Frege

Sérgio R. N. Miranda
Universidade Federal de Ouro Preto

O matemático e filósofo alemão Gottlob Frege nasceu em 1848 em Wismar, uma pequena cidade costeira situada no norte da Alemanha, estudou nas universidades de Goettingen e Jena, onde lecionou até aposentar-se em 1918, vindo a falecer em 1925. Durante a sua vida, publicou quatro livros: *Conceitografia* (1879), *Fundamentos da aritmética* (1884), e dois volumes (inicialmente seriam três) das *Leis fundamentais da aritmética* (1893 e 1902). Publicou ainda uma série de artigos científicos, entre os quais os clássicos "Sobre o sentido e a referência" (1892) e "O pensamento" (1918). Somado isso à sua correspondência com alguns filósofos e matemáticos da sua época, além dos esboços de artigos que não chegou a publicar, o *corpus* fregiano é relativamente modesto, bem como são limitados os âmbitos da principal questãoque procurou responder ao longo da sua carreira e do seu projeto intelectual: qual é abase do conhecimento aritmético? E o seu projeto, conhecido como logicismo, seria a resposta: as nossas crenças nas proposições da aritmética seriam justificáveis a partir exclusivamente de leis e princípios lógicos, sendo, pois, a capacidade de pensar logicamente a base do conhecimento aritmético.

Apesar disso, a obra de Frege é fundamental para o desenvolvimento da filosofia no século XX, podendo ser considerada um marco inicial da filosofia analítica. Frege exerceu enorme influência sobre o pensamento de autores importantes dessa tradição, como, por exemplo, admitem explicitamente Wittgenstein e Carnap. Ele é também o responsável por inovações técnicas e conceituais que permitiram o grande desenvolvimento da lógica no século passado, desenvolvimento esse que é indissociável da história da tradição analítica. Além disso, a sua obra introduziu as questões e inaugurou o modo contemporâneo de fazer filosofia em diversas áreas, como, por exemplo, filosofia da lógica, filosofia da matemática e filosofia da linguagem, além de ter enriquecido sensivelmente o debate filosófico em áreas centrais como a epistemologia e a metafísica. Por tudo isso, estudar e compreender o pensamento de Frege, além de ser indispensável para uma boa formação

em filosofia, é indispensável para o pesquisador em alguma dessas áreas mencionadas.

O “Sobre o sentido e a referência” é um clássico da filosofia da lógica e da linguagem e, sem dúvida, é o texto mais famoso de Frege. Trata-se de uma reflexão sobre a linguagem intimamente relacionada a problemas que encontramos em obras anteriores, particularmente na *Conceitografia*. De fato, o artigo começa com a exposição de um enigma sobre a relação de igualdade (identidade), apresenta a solução desenvolvida naquela obra e passa então a criticá-la. Somente depois disso apresenta a sua tese da distinção entre o sentido e a referência como solução do enigma.

O outro modo no qual Frege apresenta a distinção entre o sentido e a referência, especificamente no Prefácio das *Leis fundamentais da aritmética*, também faz menção à sua primeira obra: a distinção entre o sentido e a referência é resultado de uma divisão do que ele chamara na *Conceitografia* de conteúdo conceitual. Entender o que seja esse conteúdo, bem como a primeira solução fregiana para o enigma da identidade, pode então ajudar-nos no começo dos estudos do “Sobre o sentido e a referência”.

A *Conceitografia* nasce da intenção do autor de construir provas para noções e princípios elementares da aritmética a partir de noções e princípios elementares da lógica. Portanto, não é surpreendente a sua insistência de que a sua conceitografia ou escrita conceitual deva captar somente o que for relevante para a construção dessas provas. A isso que as frases da *Conceitografia* devem captar ele deu o nome de conteúdo conceitual.

Um exemplo bastante simples serve para ilustrar o que Frege tem em mente quando restringe a capacidade representativa da conceitografia ao conteúdo conceitual. A voz passiva difere da voz ativa, a frase do português “Brutus assassinou César” difere de “César foi assassinado por Brutus”, mas a escrita conceitual dessas frases não capta essa diferença, uma vez que o que podemos inferir de uma também podemos inferir de outra, podendo o seu conteúdoser encarado como *o mesmo*. Outras distinções são ignoradas na *Conceitografia*, que tiveram consequências importantes para o desenvolvimento da lógica: as

distinções entre sujeito e objeto, e entre tipos de juízos, particularmente a distinção entre juízos assertóricos e apodícticos.

Nessa breve exposição, podemos ver que as consequências das frases podem ser usadas como critério de identidade do conteúdo conceitual. De fato, esse critério á apresentado por Frege assim:

> *(...) os conteúdos de dois juízos podem ser diferentes de maneira dupla: primeiramente, se as consequências, que se podem tirar de um deles quando concatenados com certos outros, também podem sempre ser tiradas do segundo quando concatenados com esses mesmos outros juízos; em segundo lugar, se esse não for o caso* (Conceitografia, § 3).

É razoável entender que a diferença do primeiro tipo é justamente aquela que há, por exemplo, entre o conteúdo de "Brutus assassinou César" e "César foi assassinato por Brutus". Ela seria apenas gramatical e não uma diferença lógica, não sendo, pois, relevante para os propósitos da conceitografia. Frege assume o segundo tipo de diferença como fundamental, estabelecendo que a igualdade do potencial inferencial de dois juízos deve ser visto como uma condição necessária para a identidade dos seus conteúdos, uma vez que se for falso que eles tenham as mesmas consequências, também é falso que tenham o mesmo conteúdo. Não há no texto indicações explícitas sobre se a igualdade do potencial inferencial seria também suficiente para a identidade do conteúdo.

Isso pode gerar uma boa discussão, mas ela não afetaria diretamente o ponto principal que queremos desenvolver, e podemos, pois, deixá-la de lado. Creio que podemos então formular o critério de identidade do conteúdo proposto por Frege em termos de condições necessárias e suficientes, mesmo que depois se tenha de acrescentar algo mais a fim de propiciar um conjunto de características que formem uma condição também suficiente. O critério agora seria este: os juízos A e B têm o mesmo conteúdo conceitual se, e somente se, para um conjunto determinado de juízos S (que pode ser vazio) e uma conclusão C, caso A e S acarretem a conclusão C, B e S também acarretam C.

Logo no início da *Conceitografia*, ao apresentar o simbolismo que introduz na sua obra, Frege sugere que o usuário desse simbolismo pode

escrever que reconhece ou não a verdade do que é representado nesse simbolismo. Quando reconhece, trata-se de um juízo, quando não reconhece, apenas considera uma ligação de representações (*Vorstellungsverbindung*), preferencialmente com o intuito de tirar daí consequências. Frege denomina essa ligação de conteúdo judicatório, que logo depois receberá o nome de conteúdo conceitual. Devemos então entender o conteúdo conceitual de um juízo como "Brutus assassinou César" como a representação de cada um do assassinato de César por Brutus?

Algum tempo depois, especificamente nos *Fundamentos da aritmética*, Frege afirma haver uma ambiguidade em relação à expressão "Representação" ("*Vorstellung*"), que pode ser entendida tanto em um sentido objetivo quanto subjetivo. Na sua opinião, a representação objetiva seria a mesma para diferentes pessoas, enquanto a subjetiva não. A representação objetiva poderia ser divida em conceito e objeto, enquanto a subjetiva não (ou seja, a representação objetiva envolveria conceitos como *ser assassino*, que geraria uma verdade se isso for dito do objeto [no caso, a própria pessoa] Brutus, enquanto a representação subjetiva não admitiria essa distinção). Fundamentalmente, a representação objetiva interessa ao lógico, enquanto a subjetiva seria de interesse tão somente para a psicologia. (FA §27, n. 47) Esse parece-me um forte indício de que devemos entender a expressão "ligação de representações" na *Conceitografia* preferencialmente no sentido objetivo esclarecido depois nos *Fundamentos da aritmética.*

Uma pessoa mais cautelosa poderia julgar que essa interpretação seria precipitada, pois a nota dos *Fundamentos da aritmética* poderia estabelecer, no máximo, que a expressão "ligação de representação" (*Vorstellungsverbindung*) foi usada na *Conceitografia* ambiguamente. Mas na própria *Conceitografia* há elementos que sugerem a correção da interpretação proposta.

Frege sugere que o conteúdo judicatório ou conceitual pode ser reescrito por meio da expressão "A circunstância que..." ("*Der Umstand, dass*..."). A mesma expressão ocorrerá na discussão sobre a condicional material. Nessa discussão, ele afirma que uma condicional "Se B, então A" deve ser afirmada (reconhecida como verdadeira) sempre que A tiver de ser reconhecida como verdadeira, por exemplo, se A for uma verdade matemática como 3 × 7 = 21; a verdade ou falsidade do conteúdo de B,

por exemplo, a circunstância de que o Sol brilha, seria então irrelevante. Do mesmo modo, a condicional deverá ser afirmada (reconhecida como verdadeira) se a falsidade de B tiver de ser reconhecida, por exemplo, caso o conteúdo de B seja a circunstância de que haveria uma máquina de movimento perpétuo; e a verdade ou a falsidade do conteúdo de A, por exemplo, que o Universo é infinito, seria então irrelevante. A sugestão de que certos conteúdos *devem* ser afirmados ou reconhecidos como verdadeiros ou falsos não seria pertinente se Frege estivesse a encarar esses conteúdos em termos meramente subjetivos. Além disso, considere que, se os conteúdos devessem ser encarados subjetivamente, eles não seriam apresentados como, por exemplo, a circunstância de que o Sol brilha, mas sim a circunstância na qual alguém representa ou concebe o brilho do Sol.

Se nos fiarmos ainda nas observações dos *Fundamentos da aritmética* na nota citada anteriormente, podemos concluir que o conteúdo conceitual, além de objetivo, pode ser dividido em conceito e objeto. Isso é bastante polêmico, dado que Frege afirma na *Conceitografia* que a distinção análoga entre função e argumento diz respeito tão somente às expressões para o conteúdo, mas não ao próprio conteúdo conceitual (*Conceitografia*, § 9). Porém, penso que M. Textor tem razão ao ressaltar, na sua leitura desta obra de Frege,a pertinência da aplicação da distinção função e argumento ao próprio conteúdo, pelo menos nos casos em que a articulação entre função e argumento for relevante para propósitos lógicos, como quando fazemos inferências que envolvam generalidades.[1] De fato, dizer que "Todo filósofo é sábio" não é só dizer que este ou aquele filósofo é sábio, mas que seja o que for que cair sob a função-conceito "ser um filósofo", também cai sob a outra "ser um sábio". E estamos aqui a falar justamente dessas funções-conceitos e de argumento-objetos, e não só de palavras do português.

Podemos agora formular o problema que Frege tenta resolver no *Conceitografia* acerca da relação de identidade. Lembremos que o critério de identidade de conteúdos afirmava que A e B teriam o mesmo conteúdo se tivessem as mesmas consequências quando associados ao conjunto S de juízos. A partir da análise do conteúdo em função-conceito e argumento-objeto, parece que um conteúdo expresso, por exemplo, por

[1] Para detalhes dessa interpretação, remeto o leitor ao capítulo 3 do livro de Textor *Frege on sense and reference*, Londres: Routledge, 2011, pp. 74-102.

"Sócrates é sábio", deveria envolver um objeto particular, a pessoa de Sócrates, e a função-conceito de ser sábio. Chegamos a uma contradição quando aplicamos essa compreensão do conteúdo a juízos que envolvem igualdade (identidade).

Considere o juízo:

(1) Mohammad Ali é Cassius Clay Jr.

De acordo com a proposta de análise do conteúdo em termos de conceito-função e objeto-argumento, o conteúdo desse juízo envolve o objeto particular denotado por "Mohammad Ali", *i.e.,* um determinado boxeador, uma função-conceito, especificamente a relação "ser idêntico a", e o objeto particular denotado por "Cassius Clay Jr.", *i.e.,* exatamente aquele mesmo boxeador. Ora, o conteúdo desse juízo não deveria então diferir de

(2) Mohammad Ali é Mohammad Ali,

que também envolve um determinado boxeador, a relação de "ser idêntico a", e esse mesmo boxeador. Nos dois casos, ficamos sabendo que a pessoa Mohammad Ali é idêntica a si mesma.

O critério de identidade do conteúdo nos dá um resultado diferente. Considere:

(1) Mohammad Ali é Cassius Clay Jr.

(3) Mohammad Ali foi o maior boxeador de todos os tempos

---

(4) Logo, Cassius Clay Jr. foi o maior boxeador de todos os tempos.

A conclusão segue-se das premissas. O mesmo não ocorre se substituímos (1) por (2). Obteremos então:

(2) Mohammad Ali é Mohammad Ali

(3) Mohammad Ali foi o maior boxeador de todos os tempos

---

(4) Logo, Cassius Clay Jr. foi o maior boxeador de todos os tempos.

A inferência é inválida, a não ser, como é o caso, que Mohammad Ali seja realmente Cassius Clay Jr. Mas então não é o caso que (1) e (2) têm o mesmo potencial inferencial, visto que concluímos (4) a partir de (1) acrescentando (3), mas só concluímos (4) a partir de (2) se, além de (3), acrescentarmos agora (1). Segundo o critério de identidade de conteúdo descrito, (1) e (2) não teriam, pois, o mesmo conteúdo judicatório ou conceitual.

Essa tensão entre a análise do conteúdo e o critério de identidade será na *Conceitografia* resolvida nos seguintes termos:

> *A identidade do conteúdo diferencia-se da condicional e da negação porque dizem respeito aos nomes, e não aos conteúdos. Se em geral os sinais são apenas representantes de seus conteúdos, de tal modo que em cada cadeia na qual entram exprimem apenas a relação entre os seus conteúdos, de repente voltam-se sobre si mesmos tão-logo são ligados através do sinal de identidade de conteúdo; pois nesse caso indica-se a circunstância que dois nomes têm omesmo conteúdo* (Conceitografia, § 8).

Basicamente, a proposta de Frege é que

(1) Mohammad Ali é Cassius Clay Jr.

não deve ser analisada como envolvendo a pessoa do maior boxeador de todos os tempos, mais a relação de identidade e novamente esse mesmo boxeador. Isso é assim porque, nesse caso, estamos a falar das expressões "Mohammad Ali" e "Cassius Clay Jr.", indicando então a circunstância que esses dois nomes têm o mesmo conteúdo. A solução aparentemente resolve o problema da *Conceitografia*, uma vez que não haveria mais o conflito entre a análise do conteúdo em termos de função-conceito e argumento-objeto com o critério de identidade entre conteúdos concebido nos termos de potencial inferencial.

Mas solução não é perfeita. Sem dúvida, ela gera uma ambiguidade de uso e menção, como na regra $c \equiv d \rightarrow (\mathrm{F}(c) \rightarrow \mathrm{F}(d))$ (essa é uma lei básica introduzida na *Conceitografia*, que autoriza-nos a substituir c por d caso eles designem a mesma coisa), na qual os símbolos c e d ora são usados para designar objetos, ora representam a si mesmos. Em "Sobre a

justificação científica de uma conceitografia", escrito algum tempo depois, Frege é enfático quanto ao prejuízo que expressões ambíguas podem causar. Outro forte motivo para o abandono de Frege da sua solução parece ser a constatação nos *Fundamentos da aritmética* de que o contexto apropriado para respondermos à pergunta sobre o que são números são enunciados de identidade que envolvem expressões numéricas, e que esses enunciados envolvem, fundamentalmente, os objetos denotados por essas expressões(FA, § 57)[2].A solução definitiva que ele encontra para o enigma encontra-se em "*Sobre o sentido e a referência"*, totalmente em conformidade com a exigência de entendermos os enunciados envolvendo identidade entre expressões numéricas como relacionando objetos, e não simplesmente como afirmando algo acerca desses símbolos.

É difícil acompanhar e explicar a trajetória que levou Frege à distinção entre o sentido e a referência, como também é trabalhoso conhecer a fundo as implicações para a filosofia da linguagem que os seus textos comportam, bem como a enorme literatura a favor e contra as suas posições. Mas a leitura do texto do "Sobre o sentido e a referência" é relativamente simples. A sua estrutura é a seguinte:

> (1) apresentação do problema e do argumento da diferença do valor cognitivo entre $a = a$ e $a = b$, marcado pela diferença entre o caráter *a priori* do nosso conhecimento de $a = a$ e *a posteriori* de $a = b$ (§ 1);
>
> (2) apresentação da distinção entre sentido e referência aplicada aos nomes próprios, incluindo aqui a distinção entre a referência direta e a indireta (§§ 2-6);
>
> (3) distinção entre as noções de sentido e de referência da noção de representação (§§ 7-12);
>
> (4) resposta à objeção cética sobre a pressuposição da referência de nomes (§13);

[2]Para uma discussão detalhada, conferir o artigo de Kremer "Sense and reference: the origins and development of the distinction", In: Potter, M. *The Cambridge Companion to Frege*, Cambridge: Cambridge University Press, 2010, p. 220-293.

(5) apresentação da tese da referência de frases como o seu valor de verdade (§ 14-18);

(6) apresentação do critério leibniziano de substitutibilidade *salva veritate* (§19);

(7) avaliação dos contraexemplos à tese de Frege por meio da discussão dos casos em que o critério de substutibilidade *salva veritate* não funciona*, i.e.,* os casos em que a substituição em frases em compostas de determinadas frases por outras com o mesmo valor de verdade levaria à alteração do valor de verdade do todo (§§ 20-53);

(8) resumo (§§ 54-57);

(9) consideração final (§ 58).

O texto se desenvolve de maneira clara e objetiva, os argumentos apresentados são precisos, e a leitura muito agradável. Deixo então ao leitor só mais algumas sugestões de leituras que realmente podem ajudar no início, bem como o convite à pesquisa nas áreas relacionados com esse e outros textos de Frege.

Referências

AZAMBUJA, A. *Frege, fazedores-de-verdade e o argumento da funda*. PUC-RJ, Rio de Janeiro, março de 2007, 223 pp. Disponível em: http://criticanarede.com/teses/tese_abilio.pdf.

KLEMENT, K. *Frege*, In: *The Internet Encyclopedia of Philosophy* http://www.iep.utm.edu/frege/

KREMER, M. *Sense and reference: the origins and development of the distinction*. In: POTTER, M. *The Cambridge Companion to Frege*, Cambridge: Cambridge University Press, 2010, p. 220-293

MILLER, A. *Filosofia da linguagem*, São Paulo: Loyola, 2010

SAINSBURY, R. M. *Frege e Russell*. In: BUNNIN&TSUI-JAMES (orgs), *Compêndio de filosofia*, São Paulo: Loyola, 2007.

TEXTOR, M. *Frege on sense and reference*, London: Routledge, 2011.

WEINER, J. *Frege explained*, Open Court, 2004.

WETTSTEIN, H. *The magic prism*, Oxford: Oxford University Press, 2006 disponível em: http://criticanarede.com/lin_magicprism.html

Sobre o sentido e a referência

Frege

Tradução de Sérgio R. N. Miranda
Universidade Federal de Ouro Preto

A igualdade[1] desafia a reflexão com questões a seu respeito que não são fáceis de responder. Ela é uma relação? Uma relação entre objetos? Ou entre nomes ou símbolos de objetos? Havia admitido essa última alternativa na *Conceitografia* (*Begriffsschrift*). As razões que parecem depor a seu favor são as seguintes: $a = a$ e $a = b$ são nitidamente frases com valores cognitivos diferentes: $a = a$ vale *a priori* e, conforme Kant, deve ser denominada "analítica", enquanto frases da forma $a = b$ geralmente contêm ampliações valiosas do nosso conhecimento e nem sempre podem ser justificadas *a priori*. A descoberta que não é um novo Sol que se levanta a cada manhã, mas sempre o mesmo, foi, sem dúvida, uma das mais extraordinárias na astronomia. Ainda hoje o reconhecimento de um pequeno planeta ou de um cometa nem sempre é algo óbvio. Ora, se quiséssemos ver a identidade como uma relação entre aquilo a que se referem os nomes "*a*" e "*b*", pareceria que $a = b$ não poderia ser diferenciado de $a = a$, caso $a = b$ seja verdadeiro. Por esse meio seria exprimida uma relação de uma coisa consigo mesma, e, na verdade, uma relação que cada coisa mantém consigo mesma, mas que nenhuma mantém com outra. O que se quer dizer com $a = b$ parece ser que os símbolos ou nomes "*a*" e "*b*" se referem à mesma coisa, e assim estaríamos a falar desses símbolos; uma relação entre eles seria afirmada. Mas essa relação só existiria entre os nomes ou símbolos à medida que eles nomeassem ou designassem algo. Ela seria uma relação estabelecida pela associação de cada um dos dois símbolos com a mesma coisa designada. Mas essa associação é arbitrária. Ninguém pode ser proibido de tomar como símbolo de algo um processo ou objeto qualquer que possa ser gerado arbitrariamente. Nesse caso, a frase $a = b$ não diria mais respeito à coisa mesma, mas só ao nosso modo de designação; não exprimiríamos com ela nenhum conhecimento real. Mas em muitos casos é justamente isso o que queremos. Se o símbolo "*a*" distingue-se do símbolo "*b*" só como objeto (aqui através da sua forma), e não como símbolo – quer dizer: não no modo como designa algo –,

[1] Uso essa palavra no sentido de identidade e entendo "$a = b$" como "a é o mesmo que b" ou "a e b coincidem".

então o valor cognitivo de *a = a* seria basicamente o mesmo que o de *a = b*, caso *a = b* seja verdadeiro. Uma diferença só pode tornar-se efetiva caso a diferença dos símbolos corresponda a uma diferença no modo como é apresentado aquilo que é designado. Sejam *a*, *b* e *c* retas que unam o ângulo de um triângulo com o meio do seu lado oposto. O ponto de intersecção de *a* e *b* é então o mesmo que o ponto de intersecção de *b* e *c*. Temos, pois, diferentes designações para o mesmo ponto, e ao mesmo tempo esses nomes ("ponto de intersecção de *a* e *b*", "ponto de intersecção de *b* e *c*") indicam o modo de apresentação, e por isso há na frase um conhecimento real.

Parece então evidente que se pode pensar como associado a um símbolo (nomes, combinação de palavras, caracteres), além daquilo que designa, que se pode chamar de "referência do símbolo", também o que gostaria de chamar de "sentido do símbolo", no qual está contido o modo de apresentação. Assim, a referência das expressões no nosso exemplo "o ponto de intersecção de *a* e *b*" e "o ponto de intersecção de *b* e *c*" seria a mesma, mas não o seu sentido. A referência de "a estrela da manhã" e "a estrela da tarde" seria a mesma, mas não o sentido.

O contexto deixa claro que por "símbolo" e "nome" compreendi qualquer designação que tenha a função de nome próprio, que, portanto, tenha como referência um determinado objeto (esta palavra tomada no seu alcance mais amplo), mas não um conceito ou relação, que deverão ser tratados por mim em outro artigo. A designação de um objeto singular pode também consistir em muitas palavras ou em outros símbolos. Por brevidade, tal designação será denominada "nome próprio".

O sentido de um nome próprio será apreendido por qualquer um que conheça suficientemente a linguagem ou o conjunto das designações à qual ele pertença[2]; nesse caso, contudo, a referência, caso ela exista, sempre será focada apenas parcialmente. O conhecimento completo da

[2] Certamente, as opiniões podem divergir em relação ao sentido de um nome próprio como "Aristóteles". Poder-se-ia, por exemplo, tomar como seu sentido: o aluno de Platão e preceptor de Alexandre Magno. Quem assim procede, relacionará à frase "Aristóteles nasceu em Estagira" um sentido diferente daquele de alguém que tomasse como sentido deste nome: o preceptor de Alexandre Magno nascido em Estagira. Enquanto a referência permanecer a mesma, pode-se tolerar essa variação do sentido, muito embora ela deva ser evitada no corpo doutrinal de uma ciência demonstrativa e não deva ocorrer em uma linguagem perfeita.

referência exigiria que pudéssemos dizer imediatamente se um dado sentido a ela pertence. Jamais chegamos a esse ponto.

A conexão regular entre o símbolo, seu sentido e a sua referência é tal que ao símbolo corresponde um sentido determinado, que por sua vez corresponde a uma referência determinada, enquanto à referência (a um objeto) não é só um símbolo que lhe corresponde. O mesmo sentido tem diferentes expressões em linguagens diferentes, até na mesma linguagem. Certamente, existem exceções para esse comportamento regular. É certo que em um conjunto perfeito de símbolos cada expressão deveria corresponder a um sentido; mas a linguagem comum não satisfaz muitas vezes essa exigência, e devemos já ficar satisfeitos se no mesmo contexto a mesma palavra tiver sempre o mesmo sentido. Talvez se possa conceder que uma expressão, que seja gramaticalmente bem construída e que tenha a função nome próprio, tenha sempre o mesmo sentido. Entretanto, por esse meio não fica estabelecido que ao sentido corresponda também uma referência. A expressão "o corpo celeste mais afastado da Terra" tem um sentido; mas é bastante duvidoso se ela tem também uma referência. A expressão "a série menos convergente" tem um sentido; mas prova-se que ela não tem referência, à medida que se pode encontrar para cada série convergente uma outra que seja ainda convergente. Portanto, mesmo que se apreenda o sentido, não se tem ainda com segurança uma referência.

Se usarmos as palavras de maneira usual, aquilo sobre o que queremos falar é a sua referência. Mas pode acontecer que se queira falar sobre as próprias palavras ou sobre os seus sentidos. Isso ocorre, por exemplo, quando as palavras de terceiros são introduzidas no discurso direto. Então as próprias palavras se referem primariamente às palavras dos outros, e apenas estas têm a referência usual. Temos então símbolos para símbolos. Na escrita, as palavras são nesse caso introduzidas entre aspas. Uma palavra entre aspas não deve então ser tomada como tendo a sua referência usual.

Quando se quer falar do sentido de uma expressão "A", pode-se simplesmente lançar mão da frase "o sentido da expressão "A"". No discurso indireto, falamos do sentido, por exemplo, daquilo que diz um terceiro. Isso torna claro que também nesse caso as palavras não têm a sua referência usual, mas referem-se ao que é o seu sentido usual. Para dispormos de uma expressão sucinta, podemos dizer: as palavras são usadas no discurso indireto *indiretamente*, ou elas têm uma referência

indireta. Distinguimos, pois, a referência *usual* de uma palavra da sua referência *indireta* e o seu sentido *usual* do seu sentido *indireto*. A referência indireta de uma palavra é assim o seu sentido usual. Devemos sempre levar em conta tal exceção se queremos apreender corretamente como se dá a conexão entre símbolos, sentido e referência em casos particulares.

A representação associada a um símbolo deve ser diferenciada da sua referência e do seu sentido. Se a referência de um símbolo é um objeto empírico passível de ser percebido, a representação que tenho dele é uma imagem interna[3] decorrente da memória de impressões sensíveis que tive e de ações, tanto internas quanto externas, executadas por mim. Essa imagem é geralmente embebida em sentimentos; a clareza das suas partes separadas é variada e oscilante. Nem sempre a mesma representação é associada ao mesmo sentido, nem mesmo para a mesma pessoa. A representação é subjetiva: a representação de uma pessoa não é a mesma que a de outra. Apenas isso já deixa claro que há uma multiplicidade de representações associadas ao mesmo sentido. Um pintor, um cavaleiro e um zoólogo irão provavelmente associar ao nome "Bucéfalo" representações bem diferentes. Assim, a representação se diferencia fundamentalmente do sentido de um símbolo, que pode ser um bem compartilhado por muitos e, desse modo, não é uma parte ou um modo da mente individual; pois dificilmente se poderá negar que a humanidade tenha um tesouro comum de pensamentos que são transmitidos de uma geração para outra.[4]

Assim, enquanto não hesitamos em falar simplesmente do sentido, para sermos exatos em relação à representação é preciso acrescentar a quem ela pertence e quando. Talvez pudéssemos dizer: assim como a uma mesma palavra uma pessoa pode associar uma representação e outra uma representação diferente, também uma pessoa pode associar a ela um sentido e outra um sentido diferente. Entretanto, a diferença então reside só no modo dessa conexão. Isso não impede que ambos apreendam o mesmo sentido; seja como for, eles não podem ter a mesma

[3] Podemos acrescentar às representações as intuições, casos em que as próprias impressões sensíveis e ações tomam o lugar dos traços que estas deixam na mente. A diferença não é importante para o nosso propósito, especialmente porque, ao lado dos sentimentos e ações, sempre ocorrem lembranças de tais coisas que auxiliam a completar a imagem da intuição. Porém, pode-se também entender a intuição como envolvendo um objeto, à medida que este for perceptível pelos sentidos ou for espacial.

[4] Por isso é um despropósito designar coisas tão diferentes com a palavra "representação".

representação. *Si duo idem faciunt, non est idem*. Se duas pessoas representassem o mesmo, cada uma teria ainda a sua própria representação. Algumas vezes é possível constatar a diferença entre as representações, e até mesmo a diferença dos sentimentos, de pessoas diferentes; mas uma comparação mais exata não é possível, porque não podemos ter essas representações juntas em uma mesma consciência.

A referência de um nome próprio é o próprio objeto que designamos com ele; a representação que então temos é totalmente subjetiva; entre os dois reside o sentido, que não é subjetivo como a representação, mas por certo não é o próprio objeto. A seguinte analogia talvez seja apropriada para ilustrar essas relações. Alguém observa a Lua através de um telescópio. Comparo a própria Lua com a referência; ela é o objeto da observação, que é veiculado pela imagem real construída no interior do telescópio pela lente objetiva e pela imagem na retina do observador. Aquela comparo com o sentido, esta com a representação ou a intuição. A imagem no telescópio é apenas parcial; ela é dependente do lugar; mas ela é por certo objetiva, uma vez que vários observadores podem fazer uso dela. Pode-se mesmo orientar várias pessoas a fazer uso dela ao mesmo tempo. Mas em relação à imagem na retina cada um teria a sua própria. Mesmo uma congruência geométrica quase não pode ser alcançada em função da diferença no formato dos olhos, e um ajuste perfeito seria impossível. Talvez a analogia possa ser levada adiante se supomos que a imagem na retina de A pode tornar-se visível para B; ou que o próprio A possa ver a imagem da sua retina no espelho. Poder-se-ia aqui mostrar como a própria representação pode ser tomada como objeto, mas nesse caso ela não seria para o observador o que é imediatamente para quem se representa algo. Porém, seguir esta via seria um desvio muito grande.

Podemos então reconhecer três níveis de diferenciação para palavras, expressões e frases completas. Ou a diferença diz respeito no máximo às representações, ou ao sentido, mas não à referência, ou finalmente também à referência. Em relação ao primeiro nível deve-se notar que, em função da ligação incerta das representações com as palavras, para uma pessoa pode haver uma diferença que uma outra não reconheça. A diferença entre a tradução e o texto original não deve realmente ir além desse primeiro nível. O tom e a atmosfera, que a poesia e a eloquência procuram dar ao sentido, pertencem a possíveis diferenças ainda nesse nível. Esse tom e atmosfera não são objetivos; pelo contrário, cada ouvinte e leitor precisa forjá-los por conta própria

seguindo as indicações do poeta ou orador. Certamente, sem uma afinidade das representações humanas, a arte não seria possível; entretanto, jamais se pode estabelecer de modo exato em que medida a intenção do poeta é correspondida.

Na sequência não se falará mais das representações e intuições; elas foram aqui mencionadas só para que a representação que uma palavra desperta no ouvinte não seja confundida com o seu sentido ou com a sua referência.

Para propiciar uma maneira sucinta e precisa de se exprimir, sejam as seguintes locuções estabelecidas:

> *Um nome próprio (palavra, símbolo, combinação de símbolos, expressão) exprime o seu sentido, refere-se a ou designa a sua referência. Exprimimos com um símbolo o seu sentido e designamos com ele a sua referência.*

Talvez o cético e o idealista já tenham objetado há muito tempo: "Você fala aqui da Lua como um objeto sem oferecer maiores explicações; mas como você sabe que o nome "a Lua" tem realmente uma referência, como você sabe que alguma coisa tem realmente uma referência?" Respondo que não é a nossa intenção falar da nossa representação da Lua, e que também não nos contentamos com o sentido quando dizemos "a Lua"; pressupomos, antes, uma referência. Se quiséssemos tomar a frase "a Lua é menor do que a Terra" como dizendo algo a respeito da representação da Lua, estaríamos realmente a distorcer o seu sentido. Caso o falante queira dizer tal coisa, ele usaria como recurso a expressão "a minha representação da Lua". Certamente, podemos errar nessa pressuposição, e tais erros já ocorreram. Mas aqui pode ficar sem resposta a questão se talvez aí sempre erramos; inicialmente é suficiente indicar a nossa intenção ao falar ou pensar para justificar a nossa preleção sobre a referência de um símbolo, mesmo que com a reserva: caso exista essa referência.

Até aqui foram considerados o sentido e a referência apenas daquelas expressões, palavras ou símbolos que chamamos de "nomes próprios". Perguntamos agora sobre o sentido e a referência de uma frase declarativa completa. Tal frase contém um pensamento.[5] Devemos

[5] Entendo por pensamento não o ato subjetivo de pensar, mas o seu conteúdo objetivo, que é passível de ser uma propriedade comum de várias pessoas.

encarar esse pensamento como o seu sentido ou como a sua referência? Admitamos uma vez que a frase tenha uma referência! Se nela substituímos uma palavra por outra que tenha a mesma referência, mas um sentido diferente, isso não pode ter qualquer influência sobre a referência da frase. Mas notamos agora que o pensamento modifica-se em uma situação dessas; por exemplo, o pensamento da frase "a estrela a manhã é um corpo iluminado pelo sol" é diferente daquele da frase "a estrela da tarde é um corpo iluminado pelo sol". Alguém que não saiba que a estrela da manhã é a estrela da tarde poderia tomar um pensamento por verdadeiro e o outro por falso. O pensamento não pode ser então a referência da frase e deveremos antes concebê-lo como o seu sentido. O que dizer agora da referência? Podemos afinal perguntar sobre tal coisa? Talvez a frase como um todo só tenha sentido, mas não referência? Em todo caso, pode-se esperar que frases assim existam, do mesmo modo que existem componentes de frases que até mesmo possuem um sentido, mas não têm referência. E as frases que contêm nomes próprios sem referência serão frases desse tipo. A frase "Ulisses desembarcou em Ítaca dormindo profundamente" tem claramente um sentido. Porém, visto que é duvidoso se o nome que nela ocorre, "Ulisses", tem uma referência, é também duvidoso se a frase completa tem uma referência. Em todo caso, é certo que uma pessoa, ao tomar seriamente a frase por verdadeira ou por falsa, também atribuirá uma referência ao nome "Ulisses", e não apenas um sentido; pois é à referência desse nome que o predicado será atribuído ou negado. Quem não reconhece uma referência, não pode atribuir ou negar a ela um predicado. Ora, se queremos parar no pensamento, este avanço até a referência do nome seria desnecessário; poderíamos ficar satisfeitos com o sentido. Se o interesse fosse só o sentido da frase, ou seja, o pensamento, então não seria preciso inquietar-se com a referência de uma parte da frase; pois para o sentido de uma frase é relevante apenas o sentido, e não a referência dessa parte. O pensamento permanece o mesmo, caso o nome "Ulisses" tenha ou não uma referência. Que realmente nos preocupemos com a referência de uma parte da frase, é um indício de que em geral também reconhecemos e demandamos uma referência para a própria frase. O pensamento perde em valor para nós tão logo reconheçamos que uma das suas partes carece de referência. Estamos assim justificados a não ficar satisfeitos só com o sentido de uma frase mas também a perguntar por sua referência. Por que então queremos ter para cada nome próprio não só um sentido, mas também uma referência? Por que só o pensamento não é o suficiente? Porque e à

medida que o seu valor de verdade nos interessa. Isso nem sempre é o caso. Ao ouvir uma epopeia, por exemplo, além da melodia da linguagem, somos cativados apenas pelo sentido das frases e pelas representações e sentimentos que então são despertados. Com a questão sobre a verdade abdicaríamos do prazer estético e nos voltaríamos para uma consideração científica. Portanto, à medida que o poema é tomado como uma obra de arte, para nós é indiferente se o nome "Ulisses", por exemplo, tem referência.[6] Acima de tudo, é a busca pela verdade o que nos impulsiona a avançar do sentido para a referência.

Vimos então que se deve sempre procurar uma referência para uma frase caso a referência das suas partes seja relevante; e isso acontece no caso, e somente no caso, em que perguntamos pelo valor de verdade.

Somos assim impelidos a aceitar o *valor de verdade* de uma frase como a sua referência. Por "valor de verdade de uma frase" entendo o fato de uma frase ser verdadeira ou ser falsa. Não existem outros valores de verdade. Por brevidade denomino um deles como "o Verdadeiro" e o outro como "o Falso". Cada frase declarativa, em relação à qual a referência das palavras é relevante, deve assim ser compreendida como um nome próprio, e a sua referência, caso ela exista, é o Verdadeiro ou o Falso. Esses dois objetos são reconhecidos, ainda que implicitamente, por cada um que realmente faça um juízo, que tome algo por verdadeiro, portanto também pelo cético. A designação dos valores de verdade como objetos pode parecer aqui uma ideia arbitrária e talvez como um mero jogo com palavras, a partir do qual não se deveria tirar qualquer consequência profunda. O que eu chamo de objeto pode ser explicado com maior detalhe somente em relação às noções de conceito e relação. Desejo reservar isso para outro artigo. Mas gostaria de deixar aqui esclarecido pelo menos que em cada juízo[7] – não importa quão óbvio ele seja – o passo do nível do pensamento para o nível da referência (o objetivo) já foi dado.

Poder-se-ia ficar tentado a encarar a relação do pensamento com o Verdadeiro não como aquela entre o sentido e a referência, mas como aquela entre o sujeito e o predicado. Pois realmente podemos dizer: "O

[6] Seria desejável ter uma expressão especial para símbolos que devem ter apenas sentido. Se os chamarmos, por exemplo, de "imagem", as palavras de um ator no palco seriam imagens, até mesmo o próprio ator seria uma imagem.
[7] Um juízo não é só a apreensão de um pensamento, mas o reconhecimento da sua verdade.

pensamento de que 5 é um número primo é verdadeiro". Entretanto, se observarmos melhor, notaremos que com isso não é realmente dito nada mais do que a frase simples "5 é um número primo". A afirmação da verdade encontra-se nos dois casos na forma da frase declarativa, e nos casos em que ela não tem a sua força habitual, por exemplo, quando pronunciada por um ator no palco, a frase "o pensamento de que 5 é um número primo é verdadeiro" só contém um pensamento, que é justamente o mesmo que o da frase simples "5 é um número primo". Pode-se inferir daí que a relação do pensamento com o Verdadeiro não deve ser comparada com aquela entre o sujeito e o predicado. Decerto, o sujeito e o predicado, compreendidos no sentido lógico, são partes do pensamento; para o entendimento eles encontram-se no mesmo nível. Por meio da junção do sujeito com o predicado sempre se chega somente a um pensamento, jamais se passa de um sentido para a referência, jamais do pensamento para o seu valor de verdade. Nós nos movimentamos no mesmo nível, mas não passamos de um nível para o próximo. Um valor de verdade não pode ser parte de um pensamento, tampouco como, por exemplo, o Sol, pois ele não é um sentido, mas um objeto.

Se for correta a nossa suposição de que a referência de uma frase é o seu valor de verdade, este deve permanecer inalterado quando substituímos uma das partes da frase por outra expressão com a mesma referência, mas com outro sentido. E isto é realmente o que acontece. Leibniz explica aqui: "*Eadem sunt, quae sibi, mutuo substitui possunt, salva veritate*". Se não for o valor de verdade, o que mais pode ser encontrado que em geral pertença a cada frase, em relação a que a referência das partes componentes seja relevante, e que permaneça inalterado em uma substituição do tipo proposto?

Se o valor de verdade de uma frase é a sua referência, todas as frases verdadeiras, de um lado, têm a mesma referência, e todas as falsas, de outro, têm a mesma referência. Donde se vê que na referência de uma frase são apagadas todas as particularidades. Assim, não é só a referência de uma frase o que nos importa; mas o pensamento sozinho também não constitui um conhecimento, antes o pensamento junto com a sua referência, quer dizer, o seu valor de verdade. Julgar pode ser entendido como o avanço de um pensamento para o seu valor de verdade. Por certo isso não é uma definição. O julgar é algo bem peculiar e incomparável. Pode-se também dizer que julgar seja a diferenciação de partes no interior do valor de verdade. Essa diferenciação ocorre quando se volta

para o pensamento. Cada sentido, que pertence a um valor de verdade, corresponderia a um modo próprio da análise. Porém, utilizei aqui a palavra "parte" de um modo especial. Transferi a relação do todo e da parte das frases para a sua referência à medida que denominei a referência de uma palavra, quando esta for parte da frase, como parte da referência da frase. Esse modo de exprimir é certamente criticável, porque o todo não determina uma referência e a parte alguma outra, e porque a palavra parte já tem outro sentido em relação a corpos. Uma expressão própria deveria ser aqui inventada.

A suposição de que o valor de verdade de uma frase é a sua referência deve continuar a ser avaliada. Descobrimos que o valor de verdade de uma frase permanece intacto quando substituímos nela uma expressão por outra com a mesma referência: mas ainda não consideramos o caso no qual a expressão a ser substituída é ela mesma uma frase. Se a nossa perspectiva é correta, o valor de verdade de uma frase, que contém uma outra como parte, deve permanecer o mesmo se introduzirmos no lugar dessa parte uma outra cujo valor de verdade seja o mesmo. Exceções devem ser esperadas se o todo ou a parte da frase encontra-se no discurso direto ou indireto; pois, como vimos, a referência de uma expressão não é então a usual. Uma frase refere-se no discurso direto a uma frase e no discurso indireto refere-se a um pensamento.

Somos assim levados à consideração das orações subordinadas. Elas são complementos de uma parte da frase que, do ponto de vista lógico, aparece também como oração, especificamente como oração principal. Mas aqui estamos frente à questão se em relação às orações subordinadas vale igualmente que a sua referência é um valor de verdade. Do discurso indireto já sabemos o contrário. Os gramáticos encaram as orações subordinadas como substitutos de partes da frase, e as dividem então em orações substantivas, adjetivas e adverbiais. Isso pode levar à suposição de que a referência de uma oração subordinada não seja um valor de verdade, mas seja análoga àquela de um substantivo ou adjetivo ou advérbio, em suma, de uma parte da frase que não tem como sentido um pensamento, mas só uma parte desse pensamento. Apenas uma investigação acurada pode lançar luz sobre isso. Nela não seguiremos à risca as distinções da gramática, mas agruparemos o que for similar do ponto de vista lógico. Selecionemos inicialmente os casos em que o sentido de uma oração subordinada, tal como a pouco supomos, não seja um pensamento independente.

O caso das orações subordinadas substantivas objetivas introduzidas por "que" inclui também o discurso indireto, no qual, como vimos, as palavras têm referência indireta, que corresponde ao seu sentido usual. Aqui a oração subordinada tem como referência um pensamento, não um valor de verdade; como sentido não um pensamento, mas o sentido das palavras "o pensamento que ...", que é só uma parte do pensamento da frase composta. Isso ocorre após "dizer", "ouvir", "acreditar", "estar convencido", "concluir", e palavras semelhantes.[8] A situação é diferente e mais complexa após palavras como "reconhecer", "saber", "iludir-se", que serão posteriormente consideradas.

Que nesses casos a referência de uma oração subordinada seja realmente um pensamento, pode-se ver também no fato de ser indiferente para a verdade do todo se esse pensamento é verdadeiro ou falso. Como exemplo podemos comparar as duas frases: "Copérnico acreditou que as órbitas dos planetas eram circulares" e "Copérnico acreditou que o movimento aparente do Sol era causado pelo movimento real da Terra". Pode-se aqui empregar uma oração subordinada no lugar da outra sem prejuízo para a verdade. A oração principal junto com a oração subordinada tem como sentido só um único pensamento, e a verdade do todo não implica a verdade nem a falsidade da oração subordinada. Nesses casos não é permitido substituir na oração subordinada uma expressão por outra que tenha a mesma referência usual, mas somente por outra que tenha a mesma referência indireta, ou seja, o mesmo sentido usual. Se alguém quisesse concluir: a referência de uma frase não é o seu valor de verdade, "pois então ela sempre poderia ser substituída por outra com o mesmo valor de verdade", ele provaria demais; poder-se-ia igualmente afirmar que a referência da palavra "estrela da manhã" não seria o planeta Vênus; pois nem sempre se poderia dizer "Vênus" no lugar de "estrela da manhã". Aqui se pode apenas tirar a consequência de que a referência de uma frase nem sempre é o seu valor de verdade, e que "estrela da manhã" nem sempre se refere ao planeta Vênus, a saber, quando a palavra tem a sua referência indireta. Tal exceção ocorre em relação às orações subordinadas que consideramos há pouco, quais sejam, as orações subordinadas que têm como referência um pensamento.

[8] Em "A mentiu que ele viu B", a oração subordinada se refere a um pensamento, do qual é dito, primeiramente, que A o afirmou como verdadeiro, e, em segundo lugar, que A estava convencido da sua falsidade.

Se dissermos "parece que...", queremos dizer com isso "parece-me que..." ou "acredito que ...". Encontramos novamente o caso mencionado. A situação é similar em relação às expressões "ficar feliz", "lamentar", "aprovar", "criticar", "esperar", "temer". Se, ao final da batalha de Waterloo, Wellington ficou feliz com o fato de que os prussianos se aproximavam, a razão para tanto foi uma convicção. Ele não teria ficado menos feliz se tivesse se enganado, pelo menos enquanto durasse a sua ilusão, e antes que tivesse a convicção que os prussianos se aproximavam, ele não poderia ficar feliz com isso, mesmo que eles de fato se aproximassem.

Assim como a convicção ou uma crença é a razão para um sentimento, ela pode ser também a razão para uma convicção, como acontece em inferências. Na frase "Colombo inferiu a partir da forma esférica da Terra que ele poderia atingir a Índia viajando para o oeste", temos dois pensamentos como referências das partes: que a Terra é redonda e que Colombo pode chegar à Índia viajando para o oeste. Novamente, aqui só importa que Colombo estava convencido de um e de outro, e que uma convicção foi a razão da outra. Se a Terra realmente é redonda e se Colombo, assim como ele pensou, pudesse realmente atingir a Índia viajando para o oeste, é irrelevante para a verdade da nossa frase; mas não é irrelevante se no lugar de "Terra" introduzirmos "o planeta que tem uma lua cujo diâmetro é maior do que a quarta parte do seu próprio tamanho". Também aqui temos a referência indireta da expressão.

Orações adverbiais finais introduzidas por "para que" e "a fim de que" também pertencem a essa classe; pois o fim é claramente um pensamento; portanto: referência indireta da palavra, modo subjuntivo.

A oração subordinada introduzida por "que" após "ordenar", "pedir", "proibir" teria no discurso direto a forma do imperativo. Essa oração não tem referência, mas só um sentido. É certo que uma ordem e um pedido não são pensamentos, mas eles estão no mesmo nível que os pensamentos. Portanto, as palavras têm as suas referências indiretas em orações subordinadas dependentes de "ordenar", "pedir", etc. As referências dessas orações não são assim os seus valores de verdade, mas uma ordem, um pedido, e similares.

O caso é parecido com o das orações que complementam locuções como "duvidar se", "não saber que". Aqui também é fácil ver que as

palavras devam ser tomadas em sua referência indireta. Muitas vezes as orações interrogativas complementares introduzidas por "quem", "que", "onde", "quando", "como", "pelo que", etc. parecem muito com orações adverbiais, nas quais as palavras têm a referência usual. Esses dois casos se diferenciam linguisticamente pelo modo verbal. Temos as orações interrogativas complementares no subjuntivo e palavras com referência indireta, de tal modo que em geral um nome próprio não pode ser substituído por outro nome próprio do mesmo objeto.

Nos casos até aqui considerados as palavras, em orações subordinadas, tinham a sua referência indireta, e a partir disso foi esclarecido que a referência de uma oração subordinada também é uma referência indireta; quer dizer, não um valor de verdade, mas um pensamento, uma ordem, um pedido, uma interrogação. A oração subordinada poderia ser concebida como um substantivo, e poder-se-ia mesmo dizer: como um nome próprio daquele pensamento, daquela ordem, etc., e como tal entraria no contexto da frase composta.

Chegamos agora às orações subordinadas nas quais as expressões têm a sua referência usual sem que o pensamento figure como o sentido e o valor de verdade como a referência. Como isso é possível ficará claro por meio de exemplos.

> "Quem descobriu a forma elíptica da órbita dos planetas morreu na miséria"

Se aqui a oração subordinada tivesse como sentido um pensamento, deveria ser possível exprimi-lo também em uma oração principal. Mas isso não é possível, pois o sujeito gramatical "quem" não tem um sentido independente, introduzindo antes uma conexão com a oração subsequente "morreu na miséria". Portanto, o sentido da oração subordinada também não é um pensamento completo e a sua referência não é um valor de verdade, mas Kepler. Poder-se-ia objetar que o sentido do todo compreende em parte um pensamento, qual seja, o pensamento de que houve alguém que primeiramente reconheceu a órbita elíptica dos planetas; pois quem tomou o todo por verdadeiro não poderia negar essa parte. Não há dúvida quanto a isso; mas só porque, de outro modo, a oração subordinada "quem descobriu a ordem elíptica dos planetas" não teria uma referência. Se algo é afirmado, é sempre óbvia a pressuposição de que os nomes próprios simples ou complexos usados na frase têm uma referência. Se afirmarmos "Kepler morreu na miséria",

pressupomos que o nome "Kepler" designa algo; apesar disso, o pensamento de que o nome "Kepler" designa algo não está contido no sentido de "Kepler morreu na miséria". Se esse fosse o caso, a negação não deveria ser

> "Kepler não morreu na miséria",

mas

> "Kepler não morreu na miséria ou o nome 'Kepler' não tem referência".

Que o nome "Kepler" designa algo, é antes uma pressuposição tanto da afirmação

> "Kepler morreu na miséria"

como da sua contrária. Ora, as linguagens têm o defeito que nelas são possíveis expressões que, em conformidade com a sua forma gramatical, parecem destinadas a designar um objeto, mas em casos excepcionais não realizam esse fim porque isso depende da verdade de uma frase. Assim, depende da verdade da frase

> "existe alguém que descobriu a forma elíptica da órbita dos planetas"

se a oração subordinada

> "quem descobriu a forma elíptica da órbita dos planetas"

realmente designa um objeto ou só desperta essa impressão, quando na verdade não tem referência. E assim pode parecer que a nossa oração subordinada possui como parte do seu sentido o pensamento de que houve alguém que descobriu a órbita elíptica dos planetas. Se isso fosse correto, a negação deveria ser:

> "quem descobriu a forma elíptica da órbita dos planetas não morreu na miséria, ou não houve alguém que descobriu a forma elíptica da órbita dos planetas."

Isso é devido a uma imperfeição da linguagem que, diga-se de passagem, nem a linguagem simbólica da análise é totalmente livre; também nela podem ocorrer ligações de símbolos que para nós parecem se referir a algo, mas que, pelo menos até o momento, ainda são

desprovidas de referência, por exemplo, séries infinitas divergentes. Pode-se evitar isso, por exemplo, pela estipulação especial de que séries infinitas divergentes devem se referir ao número 0. Deve-se exigir de uma linguagem logicamente perfeita (*Begriffsschrift*) que cada expressão, construída gramaticalmente de modo correto como nome próprio a partir de símbolos já introduzidos, também designe realmente um objeto, e que nenhum novo símbolo seja introduzido como nome próprio sem que uma referência lhe seja assegurada. Nas lógicas se previne contra a ambiguidade das expressões como fonte de erros lógicos. Assumo como sendo igualmente pertinente a prevenção contra nomes próprios aparentes que não têm referência. A história da matemática pode nos relatar erros que daí surgem. Essa situação também propicia o abuso demagógico, talvez mais do que no caso de palavras ambíguas. "A vontade do povo" pode servir como exemplo; pois pelo menos pode ser facilmente constatado que não há uma referência geralmente aceita para essa expressão. Assim, definitivamente não é sem importância estancar de uma vez por todas a fonte de tais erros, pelo menos na ciência. As objeções que discutimos há pouco serão então impossíveis, pois jamais pode depender da verdade de um pensamento que um nome próprio tenha uma referência.

Na consideração dessas orações substantivas podemos acrescentar um tipo de orações adjetivas e adverbiais que, de um ponto de vista lógico, têm com elas algum parentesco.

Orações adjetivas também servem para construir nomes próprios complexos, mesmo que para isso elas não sejam sozinhas o bastante, como é o caso das orações substantivas. Essas orações adjetivas devem ser apreciadas como adjetivos. Ao invés de "a raiz quadrada de 4 que é menor do que 0", pode-se dizer "a raiz quadrada negativa de 4". Temos aqui o caso em que um nome próprio complexo é construído a partir de uma expressão para um conceito com a ajuda de um artigo definido no singular, o que em geral é permitido se um objeto, e somente um único, cair sob o conceito.[9] Assim, as expressões para conceitos podem ser construídas de tal modo que certas características são oferecidas pela oração adjetiva, como em nosso exemplo pela oração "que é menor do

[9] De acordo com o que foi estabelecido anteriormente, para uma expressão assim ficaria na verdade sempre assegurada uma referência por meio de uma estipulação especial, por exemplo, pela determinação de que o número 0 vale como a sua referência quando nenhum objeto ou mais do que um cair sob o conceito.

que 0". É esclarecedor que, tampouco como as orações substantivas, uma oração adjetiva possa não ter um pensamento como sentido e um valor de verdade como a sua referência, mas ter como sentido só uma parte do pensamento, que em muitos casos pode ser também exprimido por um único adjetivo. Tanto nesse caso como naquele das orações substantivas falta um sujeito independente, e por isso também fica excluída a possibilidade de o sentido da oração subordinada ser reproduzido em uma oração principal independente.

Lugares, instantes, porções de tempo, são, logicamente considerados, objetos; desse modo, deve-se compreender a designação linguística de um lugar específico, de um momento específico ou de uma porção de tempo como sendo um nome próprio. Orações adverbiais locativas e temporais podem então ser usadas para construir esses nomes próprios de maneira similar como vimos em relação às orações substantivas e adjetivas. Do mesmo modo, podem ser construídas expressões para conceitos que compreendam lugares etc. Deve-se aqui também notar que o sentido dessas orações subordinadas não pode ser reproduzido em uma oração principal, pois falta um componente essencial, qual seja a determinação do lugar ou do tempo, que somente é indicado por meio de um pronome relativo ou por uma conjunção.[10]

Também nas orações condicionais, como já vimos em relação às substantivas, adjetivas e adverbiais, deve-se em geral reconhecer um componente indicador indeterminado que na oração consequente a ele

[10] Vale notar que diferentes concepções são possíveis para frases desse tipo. Podemos reproduzir o sentido da frase "depois que Schleswig-Holstein foi separado da Dinamarca, a Prússia e a Áustria se desuniram" também na forma "depois da separação de Schleswig-Holstein da Dinamarca, a Prússia e a Áustria se desuniram". Nessa forma é claro que o pensamento que em algum momento Schleswig-Holstein foi separada da Dinamarca não deve ser concebido como parte do seu sentido, mas antes como a pressuposição necessária para que a expressão "depois da separação de Schleswig-Holstein da Dinamarca" tenha afinal uma referência. Nossa frase deixa-se certamente ser concebida de tal modo que com ela deva ser dito que em algum momento Schleswig-Holstein foi separada da Dinamarca. Então encontramos um caso que será posteriormente considerado. Para reconhecer mais claramente essa diferença, coloquemo-nos na pele de um Chinês que, por seu parco conhecimento da história europeia, toma por falso que em algum momento Schleswig-Holstein tenha sido separado da Dinamarca. Ele não tomaria a nossa frase, concebida na primeira forma, nem como verdadeira, nem como falsa, mas irá negar a ela qualquer referência, pois negaria que a oração subordinada tenha referência. Apenas aparentemente ela daria uma determinação temporal. Pelo contrário, se ele concebe a nossa frase do segundo modo, achará que nela um pensamento é exprimido, pensamento que tomaria por falso, ao lado do qual achará uma parte, que para ele seria sem referência.

corresponda. À medida que os dois apontam um para o outro, eles unem as duas orações em um todo, que de regra exprime apenas um pensamento. Na frase

> "se um número é menor do que 1 e maior do que 0, o seu quadrado é também menor do que 1 e maior do que 0"

este componente é "um número" na oração condicional e "seu" na consequente. É por meio dessa indeterminação que o sentido recebe a generalidade que se espera de uma lei. Mas disso também decorre que a oração condicional isolada não tem um pensamento completo como o seu sentido, e que ela junto com a oração consequente possa exprimir um e exatamente um pensamento, parte do qual não é mais um pensamento. Em geral é incorreto dizer que em um juízo hipotético dois juízos sejam colocados em uma relação de reciprocidade. Quando se diz isso, ou então algo parecido, a palavra "juízo" é usada no mesmo sentido que associei à palavra "pensamento", de tal modo que diria: "Em um pensamento hipotético, dois pensamentos são colocados em uma relação de reciprocidade". Isso só poderia ser verdade se não ocorresse um componente indicador indeterminado[11]; mas aí também não haveria generalidade.

Se um instante deve ser indicado como indeterminado nas orações condicional e consequente, não raramente isso ocorre apenas por meio do uso do *Tempus praesens* do verbo, que então não diz respeito ao momento atual. Essa forma gramatical é então o componente indeterminado de indicação na oração principal e na subordinada. "Se o Sol encontra-se no Trópico de Câncer, ocorre na parte norte da Terra o mais longo dos dias", é um exemplo disso. Também aqui é impossível exprimir o sentido da oração subordinada em uma oração principal, porque esse sentido não é um pensamento completo; pois se disséssemos: "o Sol encontra-se no Trópico de Câncer", relacionaríamos isso com o momento presente e assim o sentido seria diferente. Muito menos é o sentido da oração principal um pensamento; somente o todo da oração principal e subordinada contém um pensamento. Vale notar ainda que várias partes componentes, comuns às orações condicional e consequente, podem ser também indicadas de maneira indeterminada.

[11] Algumas vezes falta um indicador exprimido linguisticamente e este deve ser retirado do contexto global.

É claro que as orações substantivas introduzidas por "quem", "que" e as orações adverbiais introduzidas por "onde", "quando", "onde quer que", "sempre que" podem ser em muitos casos compreendidas, do ponto de vista do sentido, como orações condicionais, por exemplo, "quem toca a desgraça, lambuza-se".

Orações condicionais podem também representar orações adjetivas. Assim, podemos exprimir o sentido da nossa frase anteriormente introduzida também na forma "o quadrado de um número, que é menor do que 1 e maior do que 0, é menor do que 1 e maior do que 0".

A situação muda se o componente comum da oração principal e da subordinada for marcado por um nome próprio.

Na frase

> "O próprio Napoleão, que reconheceu o perigo para o seu flanco direito, conduziu os seus soldados ao encontro da posição inimiga"

são dois pensamentos exprimidos:

1. Napoleão reconheceu o perigo para o seu flanco direito;
2. O próprio Napoleão conduziu os seus soldados ao encontro da posição inimiga.

É só no contexto que se pode reconhecer quando e onde isso ocorreu, porém, por esse meio, isso tem de ser visto como determinado. Quando pronunciamos a nossa frase inteira como uma afirmação, afirmamos com ela os dois componentes da frase ao mesmo tempo. Se um desses componentes é falso, o todo é então falso. Aqui encontramos o caso no qual a oração subordinada contém em si mesma um pensamento completo como o seu sentido (se nós a completamos com advérbios de tempo e lugar). Assim, a referência da oração subordinada é um valor de verdade. Podemos então esperar que ela se deixe substituir por uma outra oração com o mesmo valor de verdade sem que haja prejuízo para a verdade do todo. Isso também ocorre; mas deve-se notar que o seu sujeito tem de ser "Napoleão" apenas em função de uma razão puramente gramatical, pois somente assim ela pode ser colocada na forma de uma oração adjetiva que diz respeito a "Napoleão". Porém,

quando se abandona a exigência de vê-la nessa forma, e se permite a coordenação com "e", essa limitação desaparece.

Também em frases subordinadas introduzidas por "embora" são exprimidos pensamentos completos. Essa conjunção não tem realmente um sentido e também não muda o sentido da frase, mas só lhe confere um tom bastante peculiar.[12] Ora, podemos substituir a oração concessiva por outra com o mesmo valor de verdade sem prejuízo para verdade do todo; mas o tom pareceria um pouco inadequado, como se uma pessoa quisesse cantar de uma maneira jocosa uma canção com um conteúdo triste.

Nos últimos casos a verdade do todo compreende a verdade das partes. A situação é diferente se a oração condicional exprimir um pensamento completo, à medida que, em vez de conter só um componente indicador, possui um nome próprio ou algo que deve ser visto como similar.

Na frase

"se neste momento o Sol já nasceu, o céu está muito nublado"

o tempo é o presente, portanto determinado. Também o lugar deve ser entendido como determinado. Pode-se dizer aqui que é estabelecida uma relação entre os valores de verdade das frases condicional e consequente, qual seja, que não há um caso no qual o antecedente se refere ao Verdadeiro e o consequente ao Falso. Por isso a nossa frase é verdadeira se o Sol ainda não nasceu, esteja o céu muito nublado ou não, assim como é verdadeira se o Sol já nasceu e o céu está muito nublado. Visto que aqui só interessa o valor de verdade, pode-se substituir cada uma das partes componentes por outra com o mesmo valor de verdade sem que o valor de verdade do todo mude. Certamente, na maioria das vezes o tom seria também inadequado: o pensamento pareceria ser um pouco de mau gosto; mas isso não teria nada a ver com o seu valor de verdade. Deve-se aqui sempre estar atento para o fato de que um pensamento concomitante é evocado, mas não é realmente exprimido, e

[12] O mesmo temos em relação a "mas" e "porém".

por isso não deve ser incluído no sentido da frase, não tendo então influência sobre o seu valor de verdade.[13]

Com isso foram discutidos os casos mais simples. Façamos agora uma síntese do que aprendemos!

A maioria das frases subordinadas não tem como sentido um pensamento, mas só uma parte disso e, consequentemente, não tem como referência um valor de verdade. Isso é assim ou porque na oração subordinada as palavras têm a sua referência indireta, de tal modo que a referência, não o sentido da oração subordinada, é um pensamento, ou porque a oração subordinada é incompleta pelo fato de que uma das suas partes componentes que serve de indicador é indeterminada, de tal modo que ela exprime um pensamento só quando associada à oração principal. Existem também casos nos quais o sentido da oração subordinada é um pensamento completo, e assim ela pode ser substituída por outra com o mesmo valor de verdade sem prejuízo para a verdade do todo, pelo menos enquanto não existirem dificuldades gramaticais.

Se considerarmos todas as orações subordinadas que se podem encontrar, rapidamente iremos nos deparar com algumas que não se deixam acomodar nessas categorias. Até onde vejo, a razão disso está no fato de que essas orações subordinadas não têm um sentido tão simples. Tudo indica que elas associam ao pensamento principal, que proferimos, pensamentos concomitantes que, embora não sejam proferidos, o ouvinte, em conformidade com leis psicológicas, também associa às nossas palavras. E porque parecem naturalmente associar-se às nossas palavras, quase como o próprio pensamento principal, então parece que também queremos exprimir esses pensamentos concomitantes. Com isso o sentido da frase torna-se mais rico, e pode mesmo acontecer que tenhamos mais pensamentos simples do que frases. Em alguns casos, a frase deve ser assim entendida, em outros pode ser duvidoso se o pensamento concomitante pertence ao sentido da frase ou apenas o acompanha.[14] Assim, talvez se possa achar que na frase

[13] Poder-se-ia exprimir o pensamento da nossa frase assim: "ou agora o Sol não nasceu ainda, ou o céu está muito nublado", pelo que vemos como é preciso conceber o modo de concatenação da frase.

[14] Isso pode tornar-se importante para a questão se uma afirmação é uma mentira, um juramento ou um perjúrio.

> "O próprio Napoleão, que reconheceu o perigo para o seu flanco direito, conduziu os seus soldados ao encontro da posição inimiga"

não seriam exprimidos apenas os dois pensamentos acima mencionados, mas também aquele de que o conhecimento do perigo foi uma razão pela qual ele conduziu os soldados ao encontro da posição inimiga. Pode-se de fato ficar em dúvida se esse pensamento só é sugerido ou se ele é realmente exprimido. Coloca-se a questão se a nossa frase seria falsa se a decisão de Napoleão já tivesse sido tomada antes da percepção do perigo. Se, apesar disso, a nossa frase pudesse ser verdadeira, nosso pensamento concomitante não deveria ser apreendido como parte do sentido da nossa frase. Provavelmente decidiremos a favor disso. Em outros casos, a situação seria deveras complicada: teríamos então mais pensamentos simples do que frases. Se agora também substituímos a frase

> "Napoleão reconheceu o perigo para o seu flanco direito"

por outra com o mesmo valor de verdade, como por exemplo, pela frase

> "Napoleão tinha então mais do que 45 anos de idade",

isso mudaria não apenas o nosso primeiro pensamento, mas também o nosso terceiro pensamento, e assim também o seu valor de verdade mudaria – a saber, se a idade dele não foi uma razão para a sua decisão de conduzir os soldados contra a posição inimiga. Podemos ver aqui por que em tais casos nem sempre frases com o mesmo valor de verdade podem ser substituídas entre si. Em função da sua ligação com outra frase, uma frase exprime mais do que exprime por si mesma quando tomada isoladamente.

Consideremos agora os casos nos quais isso ocorre regularmente. Na frase

> "Bebel tem a ilusão de que o desejo de vingança da França pode ser aplacado pela devolução da Alsácia-Lorena"

são exprimidos dois pensamentos, que, no entanto, não pertencem respectivamente à oração principal e à oração subordinada, quais sejam,

> 1. Bebel acredita que o desejo de vingança da França pode ser aplacado pela devolução da Alsácia-Lorena;

2. O desejo de vingança da França não pode ser aplacado pela devolução da Alsácia-Lorena.

Na expressão do primeiro pensamento, as palavras da oração subordinada têm referências indiretas, enquanto as mesmas palavras na expressão do segundo pensamento têm referências usuais. Vemos então que a oração subordinada em nossa frase complexa original deve ser realmente tomada como dupla, possuindo em cada caso uma referência distinta, que em um caso é o pensamento e no outro é um valor de verdade. Porque o valor de verdade não é a referência total da oração subordinada, não podemos simplesmente substituí-la por uma outra com o mesmo valor de verdade. Isso vale também para expressões como "saber", "reconhecer", "estar familiarizado".

Por meio de uma oração subordinada causal, juntamente com a sua oração principal, exprimimos vários pensamentos que não correspondem às orações tomadas individualmente.

Na frase

"Porque o gelo é mais leve do que a água, ele boia na água"

temos

1. o gelo é mais leve do que a água;
2. se algo é mais leve do que a água, ele boia na água;
3. o gelo boia na água.

A rigor, o terceiro pensamento não precisa ser expressamente estabelecido como estando contido nos dois primeiros. Pelo contrário, nem o primeiro e o terceiro, nem o segundo e o terceiro, quando juntos, constituiriam o sentido da nossa frase. Vemos assim que em nossa oração subordinada

"porque o gelo é mais leve do que a água"

é exprimido tanto o nosso primeiro pensamento como uma parte do nosso segundo. Disso decorre que não podemos simplesmente substituir a nossa oração subordinada por outra com o mesmo valor de verdade; pois então o nosso segundo pensamento também se modificaria, e com isso também o seu valor de verdade poderia ser atingido.

O caso é similar em relação à frase

> "se o ferro fosse mais leve do que a água, ele boiaria na água".

Temos aqui dois pensamentos, que o ferro não é mais leve do que a água e que algo boia na água se é mais leve do que a água. A oração subordinada exprime novamente um pensamento e uma parte de outro pensamento. Se concebermos a frase anteriormente considerada

> "depois que Schleswig-Holstein foi separada da Dinamarca, a Prússia e a Áustria se desuniram"

de tal modo que nela é exprimido o pensamento de que Schleswig-Holstein em algum momento foi separado da Dinamarca, então temos primeiramente esse pensamento, e, em segundo lugar, o pensamento de que em um momento, que fica determinado pela oração subordinada, a Prússia e a Dinamarca se desuniram. Também nesse caso a oração subordinada não exprime só um pensamento, mas também uma parte de outro. Por isso não se deve geralmente substituí-la por outra com o mesmo valor de verdade.

É difícil esgotar todas as possibilidades que são dadas na linguagem; mas espero principalmente ter encontrado as razões por que nem sempre podemos colocar no lugar de uma oração subordinada uma outra com o mesmo valor de verdade sem prejuízo para a verdade do todo da frase complexa. Elas são:

1. que a oração subordinada não se refere a um valor de verdade e só exprime apenas uma parte do pensamento;

2. que a oração subordinada se refere a um valor de verdade, mas não se limita a tanto, à medida que seu sentido compreende, além de um pensamento, ainda uma parte de outro pensamento.

O primeiro caso ocorre

a) em relação à referência indireta das expressões,

b) se uma parte da frase indica só indeterminadamente, em vez de ser um nome próprio.

No segundo caso, a oração subordinada pode ser tomada duplamente, ou seja, uma vez como tendo uma referência usual, outra

vez como tendo uma referência indireta; ou o sentido de uma parte da oração subordinada pode ser ao mesmo tempo parte de outro pensamento que, juntamente com o pensamento que é imediatamente exprimido na frase subordinada, forma o sentido total da oração principal e subordinada.

Podemos ver aqui que a emergência de casos nos quais uma oração subordinada não é substituível por outra com o mesmo valor de verdade nada prova contra a nossa opinião de que o valor de verdade seja a referência da frase, cujo sentido é um pensamento.

Voltemos ao nosso ponto de partida!

Se em geral achamos uma diferença no valor cognitivo de "*a = a*" e "*a = b*", isso se explica pelo fato de que, a respeito do valor cognitivo, o sentido de uma frase, a saber, o pensamento que nela é exprimido, não menos deva ser levado em conta que a sua referência, que é o seu valor de verdade. Se *a = b*, a referência de "*b*" é certamente a mesma que aquela de "*a*", e assim também o valor de verdade de "*a = b*" é o mesmo que o de "*a = a*". Apesar disso, o sentido de "*b*" pode ser diferente do sentido de "*a*", e com isso também o pensamento exprimido por "*a = b*" pode ser diferente daquele que é exprimido em "*a = a*"; assim as duas frases não têm o mesmo valor cognitivo. Se, como fizemos anteriormente, entendemos sob "juízo" o avanço do pensamento para o seu valor de verdade, também diremos que os juízos são diferentes.

Artigos Inéditos

Um falso contraexemplo à indiscernibilidade de idênticos

Luís Filipe Estevinha Lourenço Rodrigues
Centro de Filosofia da Universidade de Lisboa

Resumo

Averiguarei, neste ensaio, se, de um determinado contexto de atribuição de crença, é possível extrair um contraexemplo eficaz ao Princípio da Indiscernibilidade de Idênticos (PII), falsificando-o. O propósito final do trabalho é exibir um espécime desse contraexemplo e mostrar que, mesmo quando assume uma das suas configurações mais agressivas, não falsifica realmente o PII.

Palavras-chave: identidade, princípio da indiscernibilidade, propriedade

O plano de trabalho é o seguinte: Expor, na primeira seção, o PII e aquela que parece ser a sua correlata linguística: a Lei da Substituição de Idênticos (LSI). A segunda seção terá dois momentos: 1) introduzir um contraexemplo que falsifica a LSI e 2) ver como o PII resiste a esse tipo de contraexemplos. Na terceira seção, irá modificar-se o putativo contraexemplo para que, pelo menos como resultado de uma primeira análise, se constitua como uma tentativa bem-sucedida de falsificação do referido princípio. Por último, na quarta seção, vão-se apontar as principais razões pelas quais essa tentativa de falsificar o PII também falha.

I

O PII é um princípio metafísico e declara que se x é y – significando aqui a cópula "é" identidade estrita, quer dizer, identidade numérica –, então x e y têm exatamente as mesmas propriedades. Por exemplo, se Clark Kent e o Super-Homem são idênticos, ou seja, se são numericamente um só objeto, então têm exatamente as mesmas propriedades (JUBIEN, 1998, p.66).

A LSI é um princípio linguístico e declara que a substituição de expressões correferenciais, as quais se referem ao mesmo objeto, preserva o valor de verdade quando da passagem da frase original na

qual ocorrem essas expressões para outra(s) frase(s) que se obtenha(m) por intermédio dessa substituição. Isso quer basicamente dizer que expressões correferenciais são intersubstituíveis *salva veritate*. Por exemplo, dada a frase verdadeira (A) "Clark Kent é o Super-Homem" (supondo que a história do Super-Homem é verdadeira, algo que se irá admitir de aqui em diante), e sendo os nomes "Clark Kent" e "Kal-el" correferenciais, pode então substituir-se a ocorrência de "Clark Kent" em (A) por "Kal-el", obtendo-se dessa forma a frase (B), igualmente verdadeira, "Kal-el é o Super-Homem".

II

1) Considere-se agora a frase verdadeira (C) "Lois Lane acredita que o Super-Homem é o Super-Homem". Recuperando o processo autorizado pela LSI de substituição de idênticos, pode tentar-se substituir em (C) a primeira ocorrência da expressão "Super-Homem" pela expressão correferencial "Clark Kent". Constrói-se dessa forma a frase (D) "Lois Lane acredita que Clark Kent é o Super-Homem". Verifica-se então que, ao contrário de (C), que é verdadeira, (D) é falsa, uma vez que Lois não acredita de todo que Clark Kent seja o Super-Homem. O exemplo revela que nem sempre a substituição de expressões correferenciais que ocorrem em proposições submetidas ao âmbito de operadores epistêmicos preserva o valor de verdade. Por conseguinte, há contraexemplos à LSI que a falsificam a partir de contextos de atribuição de crença (contextos epistêmicos, portanto).

2) Pode agora experimentar-se uma tentativa semelhante de falsificação do PII. Neste novo registro, quer-se mostrar que o PII falha se o seguinte for o caso: x é y (x e y são idênticos) tal que x tem pelo menos uma propriedade, digamos, P, que y não tem. O objetivo do exercício é investigar se há pelo menos um caso de idênticos discerníveis, algo que a ocorrer seria suficiente para falsificar o PII.

Considere-se pois o seguinte:

> (E) Lois Lane acredita que o Super-Homem é vulnerável à *kriptonite*.

Contudo, há também bons indícios de que o seguinte é igualmente o caso:

(F) Lois Lane acredita que Clark Kent não é vulnerável à *kriptonite*.

Parece assim que, a confiar em (E) e (F), o Super-homem tem pelo menos uma propriedade que Clark Kent não tem, isso apesar de o Super-homem e Clark Kent serem numericamente a mesma entidade. Essa propriedade é a seguinte:

(P) Ser pensado por Lois Lane como sendo vulnerável à *kriptonite*.

Ora, se é o caso do Super-Homem ter P e Clark Kent não ter P, e se o Super-Homem é Clark Kent, ou seja, se são um só indivíduo (objeto, entidade, etc), então o PII está falsificado – pois há pelo menos um caso de idênticos que não têm as mesmas propriedades, *i.e.*, há pelo menos um caso de idênticos discerníveis.

Uma réplica habitual a esse suposto contraexemplo à indiscernibilidade de idênticos assenta na distinção entre crenças *de dicto* e crenças *de re*.[1] A réplica parece ter duas vertentes. Na primeira, faz-se notar a diferença dos alvos das crenças *de dicto* relativamente aos alvos das crenças *de re*: crenças *de dicto* incidem sobre frases e os seus respectivos conteúdos proposicionais; crenças *de re* incidem sobre um objeto, uma coisa, algo extralinguístico e extramental.[2] Argumenta-se

[1] Formalmente, parece ser a posição dos operadores existenciais relativamente aos operadores epistêmicos, e conversamente, que determina se uma crença é *de dicto* ou se é *de re*. No primeiro caso, o operador existencial está subordinado ao âmbito do operador epistêmico: x **acredita** que [**∃y** tal que y é vulnerável à *kriptonite*]. No segundo caso, acontece o inverso: é o operador epistémico que se encontra subordinado ao operador existencial. Em símbolos: **∃y** tal que [x **acredita** de y que é vulnerável à *kriptonite*]. O primeiro caso revela plausivelmente uma atitude de crença para com uma proposição. Já o segundo salienta um objeto sobre o qual incide uma atitude de crença. Existem, no entanto, várias complicações adjacentes a esta distinção *de dicto/de re*. Essas complicações surgem em larga medida devido às várias vertentes que a distinção pode admitir: a vertente epistêmica, a metafísica, a semântica e a sintática. O principal problema parece ser que a distinção assume diferentes figuras ou interpretações consoantes a vertente a partir da qual é analisada. Nem sempre se aceita, por exemplo, que o que é *de dicto* numa das vertentes o seja também noutra. Por não ser esse o assunto primário que me ocupa no corrente ensaio, assumo, sem grande discussão, e mais ou menos passivamente, que a explicação tradicional da distinção que ofereci nesta nota está suficientemente correta para apoiar-me nela. Sobre esse tópico consultar, MCKAY, T; NELSON, M, (2005).

[2] Claro que proposições, frases e outros itens linguísticos são plausivelmente objetos (linguísticos), podendo assim ser os alvos de crenças *de re* como portadores desse estatuto, quer dizer, *qua* objetos.

depois que (E) e (F) explicitam crenças *de dicto*, caso em que a atitude crença incide sobre uma proposição. Continua-se a réplica insinuando que para que o Super-Homem tenha a propriedade P é necessário que o seguinte seja o caso:

> (G) Lois Lane pense (acredite) o seguinte acerca do Super-Homem: é vulnerável à *kriptonite*.

Dada essa nova formulação de (E), prossegue-se a réplica afirmando que (G) explicita uma crença *de re*. Quer dizer, em (G) o alvo da crença não é uma proposição, como o é em (E) e (F), mas sim um objecto, um indivíduo: o Super-Homem. O ponto que se quer vindicar ao expor-se tal distinção é que somente no caso de a crença ser *de re* é que se torna possível extrair uma propriedade de um indivíduo tendo por base uma atitude de crença, visto que nesse caso o alvo da crença é esse indivíduo e não uma qualquer proposição na qual ele seja aludido.

Colocada essa formulação e o que ela implica, passa-se então à segunda vertente da réplica. Agora o objetivo é sugerir que – independentemente do modo de apresentação usado para o identificar, seja como Super-Homem, Kal-el ou Clark Kent –, o indivíduo que é alvo da crença *de re* explicitada em (G) tem a propriedade P. Dito de outro modo, quando o visado pela crença é o indivíduo chamado Super-Homem, Kal-el ou Clark Kent, esse indivíduo não pode deixar de ter a propriedade de ser pensado por Lois Lane como sendo vulnerável à *kriptonite*. Não é então o caso de, por um lado, o Super-Homem ter P, e de, por outro lado, Clark Kent não ter P – uma vez que o indivíduo Super-Homem, Kal-el ou Clark Kent tem sempre P. Mas se não é esse o caso, argumenta-se, falha o contraexemplo ao PII já apresentado acima, pois não se está perante um caso de idênticos discerníveis.

III

Apesar do que foi dito, pode ainda desenvolver-se um esforço adicional de falsificação do PII. Esse esforço irá também assentar num contexto de atribuição de crença, tal como anteriormente. Nesse sentido, importa alterar o contraexemplo apresentado na segunda secção. A ideia é que o novo contraexemplo nascido dessa tentativa de alteração acomode a réplica feita ao seu antecessor e consiga, além disso, ser imune a outras objeções que se lhe possam levantar. Só talvez assim esse

novo contraexemplo se poderá qualificar como minimamente plausível ao PII.

Considerem-se então as seguintes crenças *de re* de Lois:

(I) Acerca do Super-Homem, (Lois) pensa o seguinte: é o Super-Homem.

(L) Acerca de Clark Kent, (Lois) pensa o seguinte: não é o Super-Homem.

Segue-se que, por (I), o Super-Homem tem a seguinte propriedade:

(P´) Ser pensado por Lois Lane como sendo o Super-Homem.

Mas por (L) Clark Kent não tem (P´). Uma vez que Clark Kent é o Super-Homem, tudo indica, por conseguinte, que o indivíduo Super-Homem tem (P´) e não tem (P´). Conclui-se, de novo, que por se estar perante um caso de idênticos discerníveis – um caso em que idênticos não têm exatamente as mesmas propriedades – o PII é falso.

O que deve agora ser questionado de imediato é se esse novo candidato a contraexemplo à indiscernibilidade de idênticos é, primeiro, diferente, e, segundo, mais eficaz do que o contraexemplo apresentado na segunda seção.

Em resposta à primeira parte da questão, pode conceder-se provisoriamente que os contraexemplos são distintos. O que motiva essa resposta – condicional e sujeita à revisão – é que Lois não pode racionalmente acreditar acerca do indivíduo Super-Homem que não é vulnerável à *kriptonite*, mas pode racionalmente acreditar acerca do indivíduo Super-Homem que não é o Super-Homem. Parece haver de fato certo sentido em que Lois acredita corretamente acerca do Super-Homem que não é o Super-Homem: quando acredita acerca do Super-Homem que é Clark Kent, pois sem dúvida que o Super-Homem também é Clark Kent. Em todo o caso, há que ter bastante prudência ao abordar o problema. Mais será dito sobre ele na última seção.

Para se conseguir uma resposta afirmativa e suficientemente correta à segunda parte da questão, a da eficácia do novo contraexemplo, há que retorquir a algumas das possíveis objeções que se podem levantar ao argumento que lhe está na base. É o que se segue.

A primeira tentativa de refutação do novo contraexemplo passa por sugerir que as premissas do argumento que o sustenta, um argumento que parece válido, contêm ambiguidades. Mas não se vê como se possa sustentar essa pretensão. Verifique-se, por exemplo, se a habitual acusação de falácia de equívoco é aplicável. Onde está essa falácia nesse caso? Não há, note-se, qualquer ocorrência de nome ou descrição no argumento que não referencie sempre o mesmo indivíduo: o Super-Homem. Note-se também que o argumento não faz nenhum uso de expressões em uso anafórico, como por exemplo, "dele" ou "ele", ou quaisquer outras expressões que, ao não ocorrerem com um significado constante ao longo do raciocínio, pudessem de alguma forma gerar interpretações equívocas ou ambiguidades.

Dado que o argumento não sofre aparentemente de qualquer vício de forma ou ambiguidade, importa agora vistoriar uma possível segunda objeção. Esta objeção é similar àquela que foi usada na segunda secção para repudiar o contraexemplo aí apresentado. Nessa objeção, recorde-se, fazia-se uso da distinção *de dicto*/*de re* e reclamava-se que não se podiam extrair certas propriedades de crenças *de dicto*. Essa objeção não se afigura, porém, aplicável a esse novo caso. É que a propriedade (P´), uma propriedade que o Super-Homem certamente tem mas Clark Kent parece não ter, é extraída da crença (I) de Lois, uma crença que incide sobre o indivíduo Super-Homem e não sobre conteúdos proposicionais de frases em que ele é mencionado. Não parece, portanto, nem existir qualquer ilegitimidade na extração (P´), nem existirem transições ilícitas de crenças *de dicto* para crenças *de re* que pudessem de alguma forma bloquear a extração dessa mesma propriedade. Se assim for, o novo contraexemplo acomoda razoavelmente bem a réplica que foi dirigida ao contraexemplo da segunda secção.

Segue-se agora para outra falha porventura imputável a esse novo contraexemplo. O destaque agora é que é Lois que tem a propriedade de acreditar no que acredita a propósito do Super-Homem, e não que é o Super-Homem que tem a propriedade (P´) em virtude de Lois ter a crença que tem sobre ele. Mas parece bem claro que essa objeção não colhe. Por exemplo, afigura-se correto afirmar acerca de José Sócrates que tem a propriedade de ser pensado por alguns portugueses como sendo um bom primeiro-ministro. É óbvio que também é o caso que alguns portugueses têm a propriedade de pensar em José Sócrates como sendo um bom primeiro-ministro. Mas esta última propriedade é uma propriedade bem diferente da primeira, embora se mostre necessário

que algumas pessoas (pelo menos duas) a tenham para que Sócrates possa ter a outra.[3] Algo de semelhante parece acontecer com respeito ao novo contraexemplo: o Super-Homem tem provavelmente a propriedade (P´) porque Lois tem a propriedade de pensar acerca dele o que pensa – e, estranhamente, parece também não ter (P´) pelas mesmas razões. Se o que foi dito estiver correto, não se vislumbra como é que o fato de (P´) ter origem num pensamento de Lois pode inviabilizar que (P´) seja uma propriedade atribuível ao Super-Homem.

Pode, no entanto, insistir-se nesse tipo de objeção, argumentando que (P´) é, na melhor das hipóteses, uma propriedade inócua do Super-Homem, *i.e*., uma propriedade que ele de alguma forma possui apenas em virtude da existência ou da ação de outros, mas que em nada altera a sua condição – uma propriedade cuja posse ou não posse não implica qualquer diferença para quem a possui. Essas propriedades são habitualmente conhecidas por propriedades Cambridge. Porém, (P´) não se configura como uma propriedade desse tipo, uma vez que é certamente relevante para condição do Super-Homem ser pensado por Lois como sendo o Super-Homem. Ter tal propriedade permite, por exemplo, ao Super-Homem fazer com que Lois confie na sua capacidade de voar e, assim, possa convencê-la a cruzar os céus com ele. Veja-se que o desajeitado Clark Kent teria muita dificuldade em convencer Lois a voar com ele até ao momento em que Lois o pensasse como sendo o Super-Homem, o que se revela um bom indicador de que possuir a propriedade afeta a condição do Super-Homem. Tome-se também em consideração que só pelo fato de ter (P´) é que o Super-Homem é alvo da paixão de Lois, e que não tendo (P´) não o é (por certo que Clark Kent não é alvo da paixão de Lois). Portanto, dada a definição do que é uma propriedade Cambridge, vê-se que (P´) não pode ser uma dessas propriedades, uma vez que as suas características não caem nessa definição.

Em função do que foi sugerido anteriormente, apetece dizer que o novo contraexemplo se qualifica como uma tentativa bem-sucedida de falsificação do PII. Com efeito, nenhuma das objeções movidas até esta altura a esse contraexemplo parece ser eficaz na tarefa de refutá-lo. Nenhuma parece conseguir explicar cabalmente como é que o Super-Homem instancia (P´) e Clark Kent não instancia (P´). E se alguém

[3] Ambas as propriedades são plausivelmente propriedades relacionais e extrínsecas.

porventura sugerir que ambos instanciam (P´), já que são uma única entidade que tem quanto muito dois diferentes modos de apresentação (não apenas linguísticos), fica com o ônus de demonstrar como é que Clark Kent, sendo o Super-Homem, pode instanciar a propriedade (P´), a propriedade de ser pensado por Lois Lane como sendo o Super-Homem, uma vez que Lois não pensa de todo de Clark Kent – quando o Super-Homem lhe surge nesse modo de apresentação – que é o Super-Homem. Importa então, por certo, ver em que moldes Lois consegue alimentar crenças aparentemente contraditórias sobre o mesmo indivíduo, uma situação que parece comprometer-nos com a ideia de que esse indivíduo pode instanciar e não instanciar, em simultâneo, uma determinada propriedade – falsificando dessa forma o PII.

IV

Antes de continuar esta investigação, importa estabelecer uma premissa fundamental. Essa premissa vai auxiliar a circunscrever o seguinte problema: as atitudes de crença de Lois incidem diretamente sobre o indivíduo Kal-el (doravante assim designado para evitar confusões) e não sobre conteúdos proposicionais em que esse indivíduo seja aludido. Aceitando essa premissa, que não se afigura nada pacífica, diga-se, será por certo viável contornar grande parte da polêmica em redor do problema de se saber qual a teoria originária da Província da Semântica que melhor acomoda as crenças aparentemente contraditórias de Lois sobre Kal-el. A ideia é tentar uma resposta alternativa vinda da Província da Epistemologia.

Assente este ponto decisivo, há que armar as opções disponíveis. O objetivo é agora ver em que moldes o problema pode ser colocado quando visto pelo prisma da análise epistemológica. O que interessa então questionar é o seguinte: o que seria necessário de modo a que o indivíduo Kal-el pudesse instanciar e não instanciar, simultaneamente, (P´). A resposta é relativamente simples. Seria necessário que Lois pudesse ter também em simultâneo as seguintes crenças de re:

(I*) Acerca de Kal-el, (Lois) pensa o seguinte: é o Super-Homem.

(L*) Acerca de Kal-el, (Lois) pensa o seguinte: não é o Super-Homem.

Ora, devido a razões que serão adiante clarificadas, dificilmente Lois pode adotar (I*) e (L*) em simultâneo. Portanto, também é duvidoso que Kal-el possa instanciar e não instanciar simultaneamente (P´).

Assim visto, o problema parece ter uma solução relativamente simples. Fica, contudo, ainda por esclarecer por que razão Lois pensa de Kal-el que é e não é o Super-Homem. Há realmente que reconhecer (como se sugeriu no quarto parágrafo da terceira seção) que existe um sentido no qual Kal-el é pensado por Lois como não sendo o Super-Homem. Lois alimenta essa crença, explicitada em (L), de onde se extrai a seguinte propriedade:

(P*) Ser pensado por Lois Lane como *não* sendo o Super-Homem.

Essa é uma propriedade que é instanciada por Kal-el, uma vez que ele aparece por vezes a Lois como Clark Kent. É justamente este modo de apresentação *Clark Kent*, um modo não apenas linguístico, que condiciona a atitude da crença (L) de Lois e faz com que essa crença pareça equívoca, mal direcionada ou ambígua. Kal-el ilude Lois, fazendo-a pensar que ele, Kal-el, não é o indivíduo que ela, Lois, conhece sob o modo de apresentação *Super-Homem*. Até este instante não existe qualquer dificuldade de interpretação. Esta é, aliás, a intuição mais habitual de quem conhece a história; uma intuição provavelmente correta. Lois tem dois estados de crença – aparentemente contraditórios – para com um mesmo alvo dessas suas crenças.[4] Quando se encontra no estado de crença induzido pelo fato de Kal-el lhe surgir no modo de apresentação *Super-Homem*, Lois acredita acerca de Kal-el que é o

[4] Para ver como pode isso ser plausível, há que tomar em consideração a seguinte possibilidade: "Perry (1977) argumenta que assim que aceitamos distinção entre estados de crença e conteúdos de crença conseguiremos ver que agentes racionais podem acreditar numa proposição e na sua negação, desde que o façam por estarem em 'diferentes' estados de crença. Ou seja, Lois acredita simultaneamente que o Super-Homem é forte e o Super-Homem não é forte. Isto não coloca a sua racionalidade em questão, uma vez que ela acredita na primeira proposição por estar num determinado estado de crença devidamente relacionado com 'o Super-Homem é forte" e acredita na segunda por estar num determinado estado de crença devidamente relacionado com 'o Super-Homem não é forte'". Cf. MCKAY, T; NELSON, M, "Propositional Attitude Reports", The Stanford Encyclopedia of Philosophy (Winter, 2005 Edition).
Aceitando-se essa ideia de que um agente consegue desenvolver diferentes estados de crença sobre um mesmo conteúdo proposicional – ou, adaptando ao caso que nos ocupa presentemente, sobre um mesmo indivíduo –, pode também aceitar-se que Lois mantém crenças contraditórias sobre Kal-el por este ser o alvo de diferentes estados de crença dela, sem que ela possa ser considerada irracional por isso.

Super-Homem. Quando se encontra no estado de crença induzido pelo fato de Kal-el lhe surgir no modo de apresentação *Clark Kent*, Lois acredita acerca de Kal-el que não é o Super-Homem.

Os filósofos da linguagem mais puristas (e neorrusselianos) poderão talvez querer objetar a esta análise dizendo que (P*) não é de todo atribuível a Kal-el, uma vez que Lois não pode acreditar racionalmente acerca do indivíduo Kal-el que não é o Super-Homem, mesmo quando este lhe aparece no modo *Clark Kent*, visto que Kal-el é o Super-Homem. Mas suponha-se que Lois vem a descobrir que Clark Kent é o Super-Homem. Não é verdade que Lois deixaria de pensar, nem que fosse por breves instantes, no indivíduo Kal-el como sendo o Super-Homem para pensá-lo como sendo Clark Kent?

Em face desses dados, não se consegue evitar a impressão de que o problema não reside em Kal-el ter e não ter a propriedade (P´) em simultâneo, mas sim em Kal-el possuir (P´) e (P*). Se se conseguir provar que este último é o caso e que o primeiro não se segue deste, não restarão grandes dúvidas de que o contraexemplo ao PII apresentado na seção anterior falha redondamente. Isto na medida que esse contraexemplo não mostraria então que há idênticos discerníveis, por terem e não terem simultaneamente determinadas propriedades, mas sim que há indivíduos que instanciam propriedades contrárias (mas não contraditórias), algo de substancialmente diferente e que de nenhuma forma falsifica o PII.

Importa então visitar novamente às crenças de Lois, visto que são elas que estão na origem das propriedades de Kal-el discutidas neste trabalho. Agora que se desfizerem as ambiguidades semânticas contidas em (I) e (L), tendo resultado dessa desambiguação as crenças explicitadas em (I*) e (L*), há que verificar como é que o fato de Lois aceitar essas crenças pode contribuir para Kal-el instanciar (P´) e (P*), sem que isso arraste consigo qualquer esboço de contradição. Há, para esse efeito, que recuperar as conhecidas distinções entre, por um lado, crenças ocorrentes e crenças disposicionais, e, por outro, acreditar implicitamente ou explicitamente em algo.

No que diz respeito à primeira distinção, diz-se habitualmente que um agente tem uma crença ocorrente quando essa crença se apresenta num dado instante e de modo consciente na mente desse agente. Diz-se que um agente tem uma crença disposicional quando um agente possui

de alguma forma essa crença arquivada, embora disponível para recuperação, mas ela não está presente à sua consciência. Assim, por um lado, pode dizer-se que pelo fato de Lois estar a pensar conscientemente num determinado momento que o Super-Homem é o Super-Homem, Lois tem uma crença ocorrente acerca da identidade do Super-Homem. Por outro lado, no caso de Lois ter a crença referida neste preciso instante, no modo ocorrente, portanto, terá também a crença disposicional de que Clark Kent é seu colega no *Daily Planet* – uma vez que Lois acredita que Kent é seu colega no jornal onde trabalha, mas o seu pensamento não está nesse instante conscientemente direcionado para essa sua crença.

A segunda distinção é a seguinte: Diz-se que um agente acredita explicitamente em algo no caso de a sua mente possuir uma representação imediata e explícita desse algo que é por ele acreditado. Diz-se que um agente acredita implicitamente em algo no caso de acreditar potencialmente em algo que não está imediatamente configurado na sua mente, mas pode seguir-se do que está configurado na sua mente. Assim, por exemplo, pode dizer-se que pelo fato de Lois possuir agora uma representação clara e imediata do uniforme do Super-Homem, ela acredita explicitamente que o super-Homem veste-se de azul, vermelho e amarelo; mas que, embora não tenha essa informação devidamente configurada na sua mente, ela acredita implicitamente que o Super-Homem usa terno e gravata noutras ocasiões (quando assume o seu *alter ego* no dia a dia de *Metrópolis*).

Agora há que passar as crenças de Lois pelo crivo destas definições e verificar como é que não obstante ela acreditar no que acredita, nos moldes em que acredita, Kal-el pode ter (P´) e (P*) sem que isso implique que tenha e não tenha, simultaneamente, ou (P´) ou (P*) (esta última disjunção é exclusiva), algo que faria de Kal-el um idêntico discernível e falsificaria o PII.

O primeiro ponto a esclarecer é que, à medida que é uma pessoa racional, Lois não acredita explicitamente e, logo, simultaneamente em (I*) e (L*). Isso quase que afasta em definitivo qualquer possibilidade de Kal-el ter e não ter simultaneamente ou (P´) ou (P*). De fato, se Lois não acredita explicitamente acerca de Kal-el que é e que não é o Super-Homem, não se vê como pode Kal-el ter e não ter em simultâneo qualquer uma dessas propriedades. Outro aspecto que vem reforçar essa ideia é que é bastante improvável que Lois possa ter as crenças (I*) e (L*) ocorrentemente. Se se adotar uma postura minuciosa em face do

problema e se inquirir se duas crenças se podem sobrepor no mesmo instante na mente de um agente, no modo ocorrente, portanto, chega-se por certo à conclusão que tal acontecimento ou é bastante improvável ou é mesmo uma impossibilidade.

Todavia, não devemos esquecer que as dúvidas quanto a Kal-el poder ter e não ter em simultâneo uma determinada propriedade (seja a da existência ou outra qualquer) derivam em grande parte do fato de Lois ter as crenças aparentemente contraditórias (I) e (L). Essas crenças parecem contraditórias porque, como se viu anteriormente, envolvem diferentes modos de apresentação (linguísticos e ontológicos) de Kal-el. Independentemente de essas crenças já terem sido desambiguadas e de se ter mostrado que assim não arrastam nenhuma contradição, importa ver com maior detalhe por que razão não se pode retirar da sua conjunção que Kal-el, sendo também o Super-Homem e Clark Kent, tem e não tem simultaneamente ou (P´) ou (P*).

O ponto é agora relativamente simples. Recorde-se que, por um lado, (I) Lois pensa o seguinte acerca do Super-Homem: é o Super-Homem; e que, por outro lado, (L) Lois pensa o seguinte acerca de Clark Kent: não é o Super-Homem. Ora, em primeiro lugar, é muito difícil admitir que Lois tem ocorrentemente essas duas crenças. Se assim for, Kal-el não pode, de maneira nenhuma, ter e não ter (P´) num dado instante. E, se assim for, o mais que pode acontecer é Kal-el ter conjuntamente (P´) e (P*), mas em diferentes ocasiões. Assim, em segundo lugar, é muito provável que se Lois tiver, por exemplo, a crença (I) ocorrentemente e explicitamente, só poderá ter a crença (L) ou disposicionalmente ou implicitamente. No caso de se admitir que é implicitamente, não existe nenhuma consequência, uma vez que, nesse caso, Lois não tem sequer uma crença ativa acerca de Kal-el, uma crença da qual se possa extrair uma propriedade. No caso de se admitir que é disposicionalmente, ocorre uma situação semelhante à anterior. Embora a crença esteja neste último caso de alguma forma presente na mente de Lois, essa crença não é plausivelmente acionada e considerada por Lois até o momento em que ocorre conscientemente na sua mente (caso em que essa ocorrência não pode ser simultânea com a ocorrência de (I)). Portanto, dificilmente uma propriedade é extraível de uma crença que não está, por assim dizer, atualizada. Quer dizer, a disposição – não atualizada – para Lois acreditar que Clark Kent não é o Super-Homem não parece ter força suficiente para que se possa daí extrair uma

eventual propriedade de Kal-el (Clark Kent ou Super-Homem) de ser pensado (por Lois) como não sendo o Super-Homem.

Se o que agora foi dito estiver correto, não há a mínima hipótese de idênticos como Clark Kent e o Super-Homem instanciarem simultaneamente diferentes propriedades. Quanto muito, o indivíduo Kal-el tem – não simultaneamente – as propriedades (P´) e (P*), uma situação da qual não se segue de forma alguma a anterior. Seja como for, de modo nenhum o PII é posto em causa a partir desse contexto específico de atribuição de crença, um contexto que se usou na terceira seção para erigir o que pode agora ser visto como um putativo contraexemplo ao Princípio de Indiscernibilidade de Idênticos.

Referências

BLACK, M. *The identity of indiscernibles*. In: *Mind*, v. 51, 1952, p. 53–64.

FRENCH, S. *Why the principle of the identity of indiscernibles is not contingently true either*. In: *Synthese*, 1989, v.78, p. 141-166.

HACKING, I. *The identity of indiscernibles*. In: *Journal of Philosophy*, 1975, v.72, pp. 249-256.

HAWTHORNE, J. *Identity*. In: *LOUX*; M.J. ZIMMERMAN, D.W. (eds.), *The Oxford Handbook of Metaphysics*. Oxford: Oxford University Press, 2003.

JUBIEN, M. *Contemporary metaphysics, an introduction*. Oxford: Blackwell, 1998.

KRIPKE, S. *Naming and necessity*. Oxford: Basil Blackwell, 1980.

LOWE, E. *Objects and criteria of identity*. In: HALE, B; WRIGHT, C. (eds.), *A companion to the philosophy of language*. Oxford: Blackwell, 1997.

McKAY, T.; NELSON, M. *Propositional attitude reports*. In: *The Stanford Encyclopedia of Philosophy*, 2010. Disponível em: http://plato.stanford.edu/archives/win2005/entries/prop-attitude-reports.

_____.*Propositional attitude reports. ignorance of identities, supplement*. In: *The Stanford Encyclopedia of Philosophy* 2010, Disponível em:

http://plato.stanford.edu/entries/prop-attitude-reports/ignorance.html.

_____. *Propositional attitude reports. The de re/de dicto distinction*. In: *The Stanford Encyclopedia of Philosophy*, 2010. Disponível em:

http://plato.stanford.edu/entries/prop-attitude-reports/dere.html.

NOONAN, H. *Relative identity*. In: HALE, B; WRIGHT, C. (eds.). *A Companion to the philosophy of language*. Oxford: Blackwell, 1997.

QUINE, W., *Word and object*. Cambridge, MIT Press, 1960.

SCHWIZGEBEL, E. *Belief*. In: *The Stanford Encyclopedia of Philosophy*, 2010. Disponível em: http://plato.stanford.edu/archives/fall2006/entries/belief.

O argumento modal da consequência

Pedro Merlussi
Universidade Federal de Santa Catarina

Resumo

Em *An essay on free will*, van Inwagen apresenta três argumentos formais a favor do incompatibilismo que são três versões de um mesmo argumento. Neste artigo, discuto um desses argumentos, designadamente, o argumento modal da consequência. Esse argumento usa um operador modal sentencial que van Inwagen define como se segue: "Np" abrevia "p e ninguém tem, nem nunca teve, qualquer escolha sobre se p". Além disso, van Inwagen nos diz que esse operador tem duas regras de inferência: Alfa (de □p, podemos inferir Np) e Beta (de Np e N (p → q), podemos inferir Nq). McKay e Johnson (1996) apresentaram um bom contraexemplo à regra Beta. O objetivo deste artigo é oferecer uma resposta bem-sucedida a McKay e Johnson.

Palavras-chave: argumento modal da consequência, incompatibilismo, Peter van Inwagen

Abstract

In An Essay on Free Will, van Inwagen presents three formal arguments for incompatibilism which are three versions of the same argument. In this paper, I discuss one of these arguments, namely, the modal consequence argument. This argument uses a modal sentencial operator which van Inwagen defines as follows: "Np" stands for "p and no one has, or ever had, any choice about whether p". In addition, van Inwagen tell us that this operator has two inference rules: Alpha (from □p, we may infer Np) and Beta (from Np and N(p → q), we may infer Nq). McKay and Johnson (1996) presented a good counterexample to the rule Beta. The aim of this paper is to provide a successful response to McKay and Johnson.

Keywords: *modal consequence argument, incompatibilism, Peter van Inwagen*

Introdução

O problema de saber se o determinismo e o livre-arbítrio são compatíveis é um dos mais controversos em filosofia. Por um lado, comumente se aceita que uma condição necessária para a responsabilidade moral é termos livre-arbítrio (cf. VIHVELIN, 2011, §1). No entanto, há certo apelo intuitivo de que a tese determinista (grosso modo, a tese de que tudo o que acontece é determinado por condições iniciais mais as leis da natureza) acarreta que não temos livre-arbítrio, e consequentemente que não há responsabilidade moral. É argumentável que a incompatibilidade entre o determinismo e o livre-arbítrio seja apenas aparente, e os defensores dessa tese são conhecidos como compatibilistas ou deterministas moderados. Já os proponentes da tese de que o determinismo e o livre-arbítrio são incompatíveis são conhecidos como incompatibilistas. Até há pouco tempo, a posição compatibilista era predominante nas discussões sobre o problema. Tipicamente se pensava que os argumentos a favor do incompatibilismo repousavam em alguma falácia modal óbvia (por exemplo, num argumento do gênero da falácia fatalista). Esse erro atribuído à tese incompatibilista, entretanto, não é mais comum. E isso se deve, em parte, à defesa de Peter van Inwagen do incompatibilismo. O trabalho de van Inwagen consistiu em tornar explícito um importante argumento incompatibilista – o argumento da consequência – oferecendo-lhe três formulações. A formulação de longe mais discutida é a terceira, conhecida como argumento modal da consequência. Muitas objeções lhe foram formuladas, mas é ainda muito controverso se o argumento modal da consequência é cogente.

Este artigo é uma introdução a um aspecto central da discussão contemporânea sobre o problema do livre-arbítrio; em particular, é uma introdução ao argumento modal da consequência. O objetivo é formular esse argumento e discutir uma de suas principais objeções.

O argumento modal da consequência

Peter van Inwagen apresentou em seu *An essay on free will* (VAN INWAGEN, 1983, p. 55-105) três argumentos formais a favor do incompatibilismo que, segundo o filósofo, são diferentes versões do mesmo argumento. Trata-se do argumento da consequência.

> *Se o determinismo for verdadeiro, então as nossas ações são consequências das leis da natureza e dos eventos no passado remoto. Mas não cabe a nós o que se passou antes de nascermos, e nem cabe a nós o que as leis da natureza são. Portanto, as consequências dessas coisas (incluindo as nossas ações presentes) não cabem a nós* (INWAGEN, 1983, p. 56).

Neste artigo, apresento apenas uma formulação do argumento da consequência, conhecida como argumento modal. A razão disso é que o argumento modal é muito mais influente nas discussões sobre o problema. Com esse argumento, Inwagen procura mostrar que se segue do determinismo que não temos escolha alguma. A formulação do argumento exige breves comentários sobre os conceitos de *(a)* leis da natureza e *(b)* estado total do mundo num passado distante.

É ainda uma questão em aberto saber o que é uma lei da natureza (a), mas o que se pressupõe neste artigo não nos compromete substancialmente com uma resposta ao problema. O objetivo aqui é apenas mostrar, para fins de argumentação, o que uma lei da natureza não é.

Em primeiro lugar, é tipicamente aceito pelos filósofos que uma lei da natureza não é uma generalização acidental. Assim, por exemplo, a proposição expressa pela frase "Todas as notas de cinco reais são azuis" não é uma lei da natureza; embora essa proposição seja uma generalização verdadeira, é simplesmente acidental. Em segundo lugar, o conceito de leis da natureza não é epistêmico. Uma proposição é uma lei da natureza independentemente de a conhecermos (ou de termos uma crença justificada sobre ela). Pode ser que as proposições que pensamos serem leis da natureza de fato não o sejam, e que também não venhamos a descobrir qualquer lei da natureza, ou a ter uma crença justificada na sua existência. Em terceiro lugar, para fins de argumentação, não fará parte da extensão do conceito de leis da natureza leis psicológicas que incluam leis sobre o comportamento voluntário dos agentes racionais (cf. VAN INWAGEN, 2004, p. 696-697 e 1983, p. 64). Essa restrição serve para impedir que o determinismo seja trivialmente incompatível com o livre-arbítrio: é que se houver leis sobre o comportamento dos agentes, então esses agentes têm de se comportar do modo como essas leis descrevem os seus comportamentos.

Embora aceite-se que uma lei da natureza não seja uma generalização acidental, ainda é controverso se as leis da natureza são necessariamente verdadeiras. Por exemplo, admitindo que a proposição expressa pela frase "Nenhum objeto viaja mais rápido que a luz" seja uma lei da natureza, é argumentável (cf. CARROLL, 2010, §8) que, por ser concebível que um objeto não viaje mais depressa do que a luz, essa proposição é apenas contingentemente verdadeira. Evidentemente muitos filósofos não ficam persuadidos por esse argumento, e o seu defensor terá a dificuldade adicional em compatibilizar a tese de que as leis da natureza são contingentes com a de que não são meras generalizações acidentais. Contudo, a despeito disso, é importante notar que este artigo não pressupõe que as leis da natureza sejam necessariamente verdadeiras de modo que não se pressupõe uma concepção "necessitarista" de leis da natureza.

Quanto ao conceito de *(b),* estado total do mundo num passado distante, ficará consideravelmente inexplicado ao longo do artigo, de modo que lançarei mão de uma compreensão intuitiva. A ideia é que há uma proposição que descreve o estado total do mundo num passado distante, antes da existência de quaisquer agentes, por exemplo, uma proposição que descreva o início do universo. Não é decisivo explicar o conceito de estado total do mundo porque o argumento de Inwagen é independente de seu conteúdo. O importante é ter em mente a seguinte restrição: o conceito de estado total do mundo tem de ser tal que, dado o que o mundo é num certo estado e em certo instante, nada se segue desse estado em qualquer outro instante. Por exemplo, o conceito de estado não pode permitir a cláusula "e, neste instante, o mundo é tal que a mão de alguém será levantada 10 segundos depois deste instante" (cf. VAN INWAGEN, 2004, p. 696 e 1983, p. 58). Essa restrição serve para não permitir que o conceito *estado do mundo* possa ser definido de tal modo que o determinismo seja trivialmente verdadeiro, pois poderíamos acrescentar uma cláusula contendo informações sobre o futuro que dirão sempre o que iremos escolher amanhã.

Finalmente, tenhamos em mente as seguintes definições para a formulação da tese determinista:

> "L" é a abreviação de uma frase que expressa uma proposição que é a conjunção de todas as leis da natureza.

> "H" é a abreviação de uma frase que expressa uma proposição verdadeira sobre o estado total do mundo em algum tempo, num passado distante, antes de quaisquer agentes existirem.
>
> "L" e "H" são entendidas como abreviações de frases que expressam proposições, e não como nomes, por uma razão simples. Se fossem entendidas como nomes, uma frase como "H ∧ L" seria agramatical, pelo mesmo motivo que a frase "Pedro e João" é agramatical. E, como se pode observar logo abaixo, "H ∧ L" é necessária para a formulação do determinismo:
>
> Determinismo: é a tese segundo a qual, necessariamente, se "H" e "L" são verdadeiras, então "P" é verdadeira (sendo "P" a abreviação de qualquer frase que exprima qualquer proposição).

Formalização da tese determinista: $\Box((H \wedge L) \rightarrow P)$

A tese determinista é uma tese sobre proposições. Trata-se da tese segundo a qual uma proposição verdadeira que descreve o estado total do mundo no passado, mais uma proposição que seja uma conjunção de todas as leis da natureza, acarretam quaisquer proposições posteriores a esse estado do mundo. Já o livre-arbítrio não é uma tese sobre proposições, e sim sobre agentes. E, como nota Inwagen, se quisermos analisar as relações conceituais entre a tese determinista e a tese de que temos livre-arbítrio, teremos de tomar o livre-arbítrio como uma tese sobre agentes e proposições (cf. VAN INWAGEN, 1983, p. 66). O modo como Peter van Inwagen propõe fazer isso é o seguinte: um agente tem escolha sobre se *P* se e somente se pode tornar *P* falsa. Considere, por exemplo, as seguintes proposições:

> (a) Nenhum objeto viaja mais depressa que a luz
>
> (b) Ninguém leu este artigo inteiro em voz alta

Presumivelmente ninguém tem escolha sobre a proposição (a) porque *prima facie* ninguém pode torná-la falsa. Contudo, pelos menos alguém tem escolha sobre a proposição (b); por exemplo, o leitor tem escolha sobre a proposição (b) porque poderia simplesmente ler todo este artigo em voz alta.

No que se segue, formalizarei a tese de que não temos livre-arbítrio. Para tanto, apresentarei o operador modal N, que é caracterizado como se segue:

> "NP" abrevia "P e ninguém tem, nem nunca teve, qualquer escolha sobre se P".

Por exemplo,

> N Os dinossauros foram extintos

é uma abreviação para

> Os dinossauros foram extintos e ninguém tem, nem nunca teve, qualquer escolha sobre se os dinossauros foram extintos.

Assim, a tese de que não temos livre-arbítrio é formalizada do seguinte modo: NP (e *P* é uma variável proposicional que pode ser substituída por qualquer proposição).

Finalmente, é preciso notar que o operador N tem duas regras de inferência (VAN INWAGEN, 1983, p. 94):

(α) $\Box P \vdash \mathrm{N}P$

(β) $\mathrm{N}(P \rightarrow Q), \mathrm{N}P \vdash \mathrm{N}Q$

A regra (α) diz que, no caso de *P* ser uma proposição necessariamente verdadeira, então podemos concluir que ninguém tem, nem nunca teve, qualquer escolha sobre se *P.* A regra (β) afirma o seguinte: no caso de não termos escolha sobre *P* acarretar *Q*, e não termos escolha sobre *P,* então podemos concluir que não temos escolha sobre *Q.* Agora o argumento.

Suponhamos que o determinismo seja verdadeiro, isto é, suponhamos que necessariamente, se "H" e "L" são verdadeiras, então *P* também o é:

$\Box((H \wedge L) \rightarrow P)$

Da primeira premissa, infere-se pela regra de exportação.

$\Box(H \rightarrow (L \rightarrow P))$

Aplicando a regra (α) a 2, temos

$\mathrm{N}(H \rightarrow (L \rightarrow P))$

Agora é preciso introduzir a premissa de que não temos escolha sobre o passado. A ideia subjacente a essa premissa é que não temos o poder de tornar falsa uma proposição sobre o estado total do mundo num passado distante.

$\mathrm{N}H$

e de 3 e 4 infere-se, pela regra (β) que

$\mathrm{N}(L \rightarrow P)$

Novamente acrescentamos mais uma premissa, a saber, a premissa segundo a qual não temos escolha sobre as leis da natureza. A ideia subjacente a essa premissa é que não temos o poder de tornar falsa uma lei da natureza.

6. $\mathrm{N}L$

Inferimos agora de 5 e 6, pela regra (β)

$\mathrm{N}P$

Logo, se 1 é verdadeira (isto é, se o determinismo é verdadeiro), então 7 é verdadeira. Como P é uma variável proposicional, e podemos substituí-la por qualquer proposição, conclui-se que, se o determinismo é verdadeiro, ninguém tem, nem nunca teve, qualquer escolha.

$\Box ((H \wedge L) \rightarrow P) \rightarrow \mathrm{N}P$

O argumento parece colocar o determinista moderado em dificuldades. Para um determinista moderado mostrar que o argumento não é cogente, terá de mostrar que pelo menos uma de nossas crenças nestas proposições não é plausível:

A regra de exportação para a lógica modal alética é válida;

A regra (α) é válida;

A regra (β) é válida;

NL;

NH.

Filósofos que têm alguma razão independente para rejeitar a lógica modal alética – como foi o caso de W. O. Quine – podem rejeitar 1. Entretanto, pelo fato de a lógica modal ser bem-sucedida, rejeitar 1 para refutar o argumento modal da consequência não parece um bom caminho. Isso ocorre porque teremos de usar 1 como premissa para pretender sustentar a conclusão de que o argumento modal da consequência é inválido. Ora, mas a premissa que nega 1 é menos plausível que a conclusão estabelecida pelo argumento modal da consequência. Portanto, esse tipo de argumento não é cogente (considerando um argumento cogente como um argumento sólido cujas premissas são menos disputáveis que a conclusão).

Já a crença em 2 parece muito plausível. A regra (α) é praticamente incontestável. Eis um argumento a favor de (α):

1. Se *P* é necessariamente verdadeira, então não pode ser falsa.

2. Se *P* não pode ser falsa, então ninguém pode torná-la falsa.

3. Se ninguém pode tornar *P* falsa, então ninguém tem escolha sobre se *P*.

4. Logo, se *P* é necessariamente verdadeira, então ninguém tem escolha sobre se *P.*

Portanto, contraexemplos a (α) parecem estar excluídos.

Para alguém mostrar que 4 é falsa, terá de encontrar um contraexemplo ao esquema "se *P* é uma lei da natureza, então ninguém pode torná-la falsa"; assim, terá de mostrar que alguém tem o poder de tornar uma lei da natureza falsa, o que é implausível. Já para mostrar que 5 é falsa, terá analogamente de encontrar um contraexemplo ao esquema "se *P* é uma proposição verdadeira sobre o estado total do mundo num passado distante (antes de agentes existirem), então ninguém pode torná-la falsa".

Há objeções a 4 e 5, mas neste artigo apresento apenas uma objeção a 3. Essa objeção, de longe a mais influente, ataca a validade da regra (β), e foi apresentada por McKay e Johnson (1996).

Breve nota sobre a regra beta

Antes de tudo, oferecerei uma breve nota sobre a regra (β). Pelo menos até o trabalho de Inwagen, tipicamente se pensava que os argumentos a favor do incompatibilismo repousavam em alguma falácia modal óbvia (como um argumento do gênero da falácia fatalista). Mas é importante mostrar que as coisas aqui são diferentes. É importante não confundir a regra (β) com a seguinte forma argumentativa inválida:

1. □ (P → Q)

2. P

3. □Q

A regra (β) do argumento de Inwagen estabelece a diferença entre o argumento modal da consequência e a falácia fatalista. Considere a seguinte forma argumentativa válida:

□(P → Q)

□ P

Logo, □Q

Recorrerei a uma árvore lógica para demonstrar que a forma argumentativa supracitada é válida:

1.	□ (P → Q)		0
2.	□ P		0
3.	¬□Q		0
4.	◊¬Q	(3)	0
5.	0⁓1	(4)	0
6.	¬ Q	(4,5)	1
7.	P	(2,5)	1
8.	P → Q	(1,5)	1

9. ¬P	(8)	1		10. Q	(8)	1
	x				x	

O operador modal da "não escolha" é, de fato, diferente do operador modal da necessidade. Portanto, do fato de a regra funcionar para o operador modal da necessidade não se segue que ela também funcione para o operador da não escolha. Não obstante, o operador da não escolha parece similar o bastante ao operador da necessidade para pensarmos que a regra (β) seja uma regra de inferência válida[1]. Porém, McKay e Johnson apresentaram um bom argumento para mostrar que a regra (β) é inválida.

A objeção de McKay e Johnson

Em "A reconsideration of an argument against *compatibilism*", McKay e Johnson argumentaram que a regra (β) implica o princípio de aglomeração, e apresentaram um contraexemplo para mostrar que o princípio de aglomeração é inválido. Portanto, se o princípio de aglomeração é inválido e se a regra (β) implica o princípio de aglomeração, então a proposição segundo a qual a regra (β) é válida é afinal de contas falsa.

Princípio de Aglomeração: $\mathbf{N}P \wedge \mathbf{N}Q \vdash \mathbf{N}(P \wedge Q)$

O que o princípio de aglomeração diz é o seguinte: uma vez que ninguém tem, nem nunca teve, qualquer escolha sobre se *P*, e que também ninguém tem, nem nunca teve, qualquer escolha sobre se *Q,* segue-se que ninguém tem, nem nunca teve, qualquer escolha sobre se *P* e *Q.* Assim, o contraexemplo de McKay e Johnson procura mostrar que é inválido concluir $\mathrm{N}(P \wedge Q)$ a partir da premissa $\mathrm{N}P \wedge \mathrm{N}Q$. Vejamos o contraexemplo.

Suponha que ninguém lançou certa moeda, mas que poderia lançá-la. É importante acrescentar ainda que a moeda tem apenas dois lados, não foi alterada e está em condições normais de uso.

[1] Michael Slote (1982) argumentou que a regra (β) não se assemelha ao operador da necessidade. Como McKay e Johnson apresentaram um argumento mais poderoso para mostrar que a regra (β) é inválida, irei desconsiderar neste ensaio a objeção de Slote.

P = a moeda não virou cara.

Q = a moeda não virou coroa.

Lembre-se que "ter escolha sobre se P" se define como "um dado agente tem escolha sobre se P se, e só se, esse agente pode tornar P falsa".

A premissa $NP \wedge NQ$ é verdadeira. Se alguém tem escolha sobre a moeda não virar cara, então pode tornar P falsa. No entanto, ninguém pode tornar P falsa, pois ninguém tem o poder de, ao lançar a moeda, fazê-la virar cara. Logo, ninguém tem escolha sobre se a moeda não virou cara. Analogamente, se alguém tem escolha sobre a moeda não virar coroa, então essa pessoa pode tornar Q falsa. Porém, ninguém pode escolher tornar Q falsa, pois ninguém tem o poder de, ao lançar a moeda, fazê-la virar coroa. Logo, ninguém tem escolha sobre a moeda não virar coroa.

Mas, surpreendentemente, a conclusão $N(P \wedge Q)$ é falsa. Isso porque uma pessoa poderia tornar $P \wedge Q$ falsa ao arremessar a moeda. Se uma pessoa arremessasse a moeda, esta viraria cara ou coroa. Ora, a negação da frase "a moeda não virou cara e não virou coroa", $P \wedge Q$, é justamente "a moeda virou cara ou coroa", $\neg P \vee \neg Q$. Se alguém pode tornar $P \wedge Q$ falsa, então tem escolha sobre se $P \wedge Q$. E como se pode tornar $P \wedge Q$ falsa ao arremessar a moeda, caso em que ela viraria cara ou coroa, $\neg P \vee \neg Q$, segue-se que alguém tem escolha sobre se $P \wedge Q$. Logo, é falso que ninguém tem, nem nunca teve, qualquer escolha sobre se $P \wedge Q$. Desse modo, o princípio de aglomeração é inválido. E o problema é que a regra (β) implica o princípio de aglomeração:

NP	premissa
NQ	premissa
$\Box (P \rightarrow (Q \rightarrow (P \wedge Q)))$	necessidade de uma verdade lógica
$N(P \rightarrow (Q \rightarrow (P \wedge Q)))$	de 3 e da regra (α)
$N(Q \rightarrow (P \wedge Q))$	de 1, 4 e da regra (β)
$N(P \wedge Q)$	de 2, 5 e da regra (β)

As premissas, como o leitor pode notar, são verdadeiras. NP é verdadeira porque ninguém tem o poder de, ao lançar a moeda, fazê-la virar cara. NQ é verdadeira porque ninguém tem o poder de, ao lançar a moeda, fazê-la virar coroa. 3 é uma tautologia. 4 se segue de 3, que à primeira vista é indisputável, e de 4, que, como vimos anteriormente, é muito plausível. Por meio da regra (β) chegamos facilmente a 6, que, como vimos, é falsa. Portanto, o que permitiu, no raciocínio anterior, chegar de premissas verdadeiras a uma conclusão falsa foi justamente a regra (β). Logo, a regra (β) é inválida.

Saber se esse é um argumento cogente, é algo que permanece em aberto. Há filósofos que pensam o contrário. Por exemplo, Finch e Warfield (1998) propuseram uma nova forma de tornar o argumento modal da consequência imune à objeção supracitada, substituindo a regra (β) pela regra (β)2.

(β)2: NP, $\Box (P \rightarrow Q) \vdash$ NQ

Essa substituição nos daria o seguinte argumento:

$\Box((H \wedge L) \rightarrow P)$	suposição
N$(H \wedge L)$	premissa
NP	1, 2 (β) 2
$\Box((H \wedge L) \rightarrow P) \rightarrow$ NP	1-3, I→

Ao substituir a regra (β) por (β) 2, observa-se que esta última não implica o princípio de aglomeração:

NP	premissa
NQ	premissa
$\Box (P \rightarrow (Q \rightarrow (P \wedge Q)))$	necessidade de uma verdade lógica
N $(Q \rightarrow (P \wedge Q))$	de 1 e 3 por (β) 2

A partir de 4, só seria possível derivar N $(P \wedge Q)$ caso admitíssemos a regra (β) original. No entanto, na reformulação de Finch e Warfield, o argumento modal da consequência não precisa da regra (β) original. E como a regra (β) 2 não implica a aglomeração, a nova formulação é imune à objeção de McKay e Johnson.

Penso, no entanto, que a regra (β)2 de Finch e Warfield seja também inválida. Considere o seguinte argumento:

> 1. Necessariamente, se a Terra for atingida por um cometa gigantesco em 2013, então os seres humanos morrerão em 2013.
>
> 2. Ninguém tem, nem nunca teve, qualquer escolha sobre se a Terra será atingida por um cometa gigantesco em 2013.
>
> 3. Logo, ninguém tem, nem nunca teve, qualquer escolha sobre se os seres humanos morrerão em 2013.

As duas premissas do argumento são verdadeiras. Presumivelmente, em todos os mundos possíveis em que a Terra é atingida por um cometa gigantesco em 2013, os seres humanos morrerão nesse mesmo ano. E pelo menos atualmente não temos o poder de tornar falsa a proposição expressa pela frase "A Terra será atingida por um cometa gigantesco em 2013"; por isso ninguém tem, nem nunca teve, qualquer escolha sobre se a Terra será atingida por um cometa gigantesco em 2013. Entretanto, a conclusão é falsa. Isso porque alguém pode tornar falsa a proposição de que os seres humanos morrerão em 2013; portanto, é falso que ninguém tem, nem nunca teve, qualquer escolha sobre se os seres humanos morrerão em 2013.

Um defensor da regra (β)2 poderá argumentar que, nas circunstâncias em que a premissa 1 é verdadeira, e a conclusão 3 é falsa, a premissa é, na realidade, falsa. Considere o seguinte:

> 1. Necessariamente, se a Terra for atingida por um cometa gigantesco em 2013, então os seres humanos morrerão em 2013.
>
> 2. Mas posso me cuidar e não morrer em 2013, mesmo que um cometa se choque com a Terra.

Acontece que 1 e 2 são proposições inconsistentes. 2 é verdadeira se considerarmos que eu poderia sair da Terra e, portanto, poderia não morrer em 2013. Mas, ao afirmar isso, estamos implicitamente dizendo que 2 é falsa. Isso porque admitimos que é possível que um cometa se choque com a Terra mesmo que alguns seres humanos não morram, o que, obviamente, é a negação de 1.

Esta resposta, entretanto, não parece funcionar. O que o objetor deveria mostrar é que estas duas proposições são inconsistentes:

1. Necessariamente, se a Terra for atingida por um cometa gigantesco em 2013, então os seres humanos morrerão em 2013.

2. Posso tornar falsa a proposição de que os seres humanos não irão morrer em 2013.

Para 2 ser verdadeira, não tenho de ter o poder de fugir da Terra. Eu teria de ter esse poder apenas se, necessariamente, um cometa se chocar com a Terra. Neste caso, sim, teríamos uma inconsistência. Mas 1 não diz isso. Apenas diz que, necessariamente, os seres humanos morrem caso um cometa se choque com a Terra.

Considere o seguinte caso análogo: necessariamente, se cortarem minhas pernas, não caminharei de Ouro Preto a Mariana. Disso não se segue que eu não possa tornar falsa a proposição de que não caminharei de Ouro Preto a Mariana. Eu só não poderia caminhar de Ouro Preto a Mariana se tivéssemos uma premissa adicional, a saber, a de que necessariamente cortarão minhas pernas. Mas não temos essa premissa e, por isso, as proposições não são inconsistentes.

Talvez o defensor da regra (β)2 presuma que a premissa 1 só é verdadeira quando a antecedente da condicional o é. Mas nada exige que isso seja assim. Num mundo possível em que um cometa não se choca com a Terra, a premissa 1 continua verdadeira; afinal, a antecedente da condicional é falsa. Mas a conclusão obviamente é falsa. Como o mundo não acabou, posso me cuidar para não morrer, e, assim, tornar a proposição de que os seres humanos irão morrer em 2013 falsa.

Em suma, para a regra (β)2 ser válida, a forma argumentativa teria de ser a seguinte:

$\Box (P \rightarrow \mathrm{Q})$

$\Box P$

NP

Logo, NQ

Entretanto, essa regra nova regra não funciona para salvar o argumento modal da consequência da objeção de McKay e Johnson; afinal, precisaria da premissa $\Box (H \wedge L)$. Em virtude dessa dificuldade, consideremos outra resposta à objeção.

A regra beta determinismo

O contraexemplo de McKay e Johnson depende da suposição de que alguém poderia arremessar a moeda. Isso porque, se alguém pode arremessar a moeda, então pode tornar a proposição de que a moeda virou cara e coroa falsa; e, portanto, tem escolha sobre se a moeda virou cara e coroa. Assim, a conclusão N($P \wedge Q$) seria falsa, apesar das premissas do argumento serem verdadeiras, pois pelo menos alguém teria escolha sobre $P \wedge Q$. Contudo, se supusermos que o determinismo seja verdadeiro, segue-se que há circunstâncias em que é impossível alguém arremessar a moeda. E como é impossível arremessar a moeda, segue-se que não é possível tornar $P \wedge Q$ falsa. Portanto, não é possível que N($P \wedge Q$) seja falsa. Desse modo, não teríamos um caso de premissas verdadeiras e conclusão falsa, já que é impossível, admitindo o determinismo, que a conclusão N($P \wedge Q$) seja falsa.

Para mostrar isso, admitiremos que P é a proposição de que a moeda não foi lançada.

Em primeiro lugar, temos a definição do determinismo:

$\Box((H \wedge L) \rightarrow P)$ 0

Agora supõe-se que a moeda poderia ser lançada, considerando as mesmas leis da natureza e o mesmo passado:

$\Diamond$ ((H $\wedge$ L) $\wedge$ ¬P) 0

Disso se segue uma contradição:

1.	0⇝1	(2)	
2.	(H $\wedge$ L) $\wedge$ ¬P	(2,3)	1
3.	H $\wedge$ L	(4)	1
4.	¬P	(4)	1
5.	H	(5)	1
6.	L	(5)	1
7.	$(H \wedge L) \rightarrow P$	(1,3)	1

10. $\neg(H \wedge L)$ (9) 1 11. P (9) 1

x

11. ¬H (10) 1 12. ¬L (10) 1

x x

A conclusão a retirar é que é impossível alguém lançar uma moeda num mundo possível com as mesmas leis da natureza e o mesmo passado que o mundo determinista no qual a moeda não foi arremessada. Portanto, o contraexemplo de McKay e Johnson à regra (β) não funciona, caso se admita o determinismo.

McKay e Johnson podem objetar que isso é irrelevante quanto à invalidade da regra (β). Concordo com a objeção e penso que o contraexemplo por eles apresentado realmente mostra que a regra (β) é inválida. No entanto, o contraexemplo não funciona para invalidar a regra (β)D (cf. CRISP e WARFIELD, 2000), ou seja, a regra (β) determinismo:

D, N$(P \rightarrow Q)$, N$P \vdash$NQ.

O contraexemplo de McKay e Johnson não funciona caso se admita o determinismo, pois nesse caso seria impossível alguém arremessar a moeda nas condições expostas. E, se é impossível alguém arremessar a moeda nessas condições, a conclusão N$(P \wedge Q)$ não pode ser falsa. Portanto, não se teria um caso com premissas verdadeiras e conclusão falsa, pois a conclusão N$(P \wedge Q)$, caso se admita o determinismo, é necessariamente verdadeira. E note-se que o argumento de Inwagen supõe o determinismo; é o primeiro passo de seu argumento. Portanto, embora o contraexemplo de McKay e Johnson mostre que a regra (β) seja em si inválida, não mostra que o argumento modal da consequência seja inválido, pois este supõe o determinismo. Assim, teríamos o seguinte argumento, depois de relembrarmos a formulação das regras (α) e (β)D:

(α) $\Box p \vdash \mathbf{N}P$

((β)D) D, $\mathbf{N}(P \rightarrow Q)$, $\mathbf{N}P \vdash \mathbf{N}Q$

1. $\Box((H \wedge L) \rightarrow P)$	premissa
2. $\Box(H \rightarrow (L \rightarrow P)$	1, pela regra de exportação
3. $\mathbf{N}(H \rightarrow (L \rightarrow P))$	2, pela regra (α)
4. $\mathbf{N}H$	premissa
5. $\mathbf{N}(L \rightarrow P)$	**1, 3, 4 pela regra (β)D**
6. $\mathbf{N}L$	premissa
7. $\mathbf{N}P$	**1, 5, 6 pela regra (β)D**
8. $\Box((H \wedge L) \rightarrow P) \rightarrow \mathbf{N}P$	1-7 I→

O erro da formulação original de Inwagen teria sido não introduzir 1 na justificação dos passos 5 e 7. Entretanto, ao introduzir o determinismo, temos que a conclusão $\mathrm{N}(P \wedge Q)$ não pode ser falsa, de modo que não há problema no fato de a regra (β)D implicar o princípio de aglomeração. Afinal, o princípio de aglomeração não é inválido para a regra (β)D.

Há uma réplica que se poderia fazer à resposta aqui oferecida. Talvez seja razoável dizer que alguém teve escolha sobre a moeda não virar cara e não virar coroa mesmo num mundo determinista, independentemente de poder tornar falsa a proposição de que a moeda não virou cara e não virou coroa. Mas não penso que essa objeção seja bem-sucedida. O que está em jogo no problema é o de saber se o determinismo é compatível com o livre-arbítrio. Supor que alguém tem escolha num mundo determinista é já pressupor de antemão o compatibilismo. Portanto, a réplica é circular, e por isso não é bem-sucedida.

O argumento modal da consequência de Inwagen é um argumento poderoso contra o determinismo moderado. Seu ponto mais fraco apresentou-se na regra (β). No entanto, há respostas a esse problema, e procurei mostrar que o argumento modal da consequência pode ser consertado.

Conclusão

O problema do livre-arbítrio é um dos mais intricados em filosofia. Neste artigo, o leitor pôde entrar em contato com um importante argumento incompatibilista e uma de suas principais objeções, além de ter indicações de resposta a essa objeção. Em específico, procurei mostrar que um defensor da versão original do argumento modal da consequência pode levar em conta a resposta à objeção de McKay e Johnson que foi aqui brevemente explorada.[2]

Referências

CRISP, Thomas M. & WARFIELD, Ted A.. *The irrelevance of indeterministic counterexamples to principle beta*. In: *Philosophy and Phenomenological Research* 61, n. 1, p. 173-84, 2000.

CARROLL, John W. *Laws of Nature*. In: *The Stanford Encyclopedia of Philosophy,* 2010.

FINCH, Alicia & Warfield, Ted A. *The mind argument and libertarianism*. In: *Mind* 107, n. 427 p. 515-28, 1998.

HUEMER, Michael. *Van Inwagen's consequence argument*. In: *The Philosophical Review* 109, n. 4 p. 524-44, 2000.

LEWIS, David. *Are we free to break the laws?*. In: *Philosophical Papers*. New York: Oxford University Press, 1986.

MCKAY, Thomas J. & JOHNSON, David. *A reconsideration of an argument against compatibilism* In:. *Philosophical Topics* 24, n. 2 p. 113-22. 1996.

SLOTE, Michael. *Selective necessity and the free-will problem*. In: *The Journal of Philosophy*, v. 79, n. 1 p. 5-24, 1982.

VAN INWAGEN, Peter. *The incompatibility of free will and determinism*. In: *Philosophical Studies*, 25: 185-99, 2002.

[2] Este artigo contou com diversas objeções e sugestões de vários amigos e colegas. Agradeço, em especial, às valiosas objeções e sugestões de Sérgio Ricardo Neves de Miranda, Desidério Murcho, Kherian Gracher, Luiz Helvécio Marques Segundo, Iago Bozza Francisco, Aluízio de Araújo Couto Júnior e Eduardo Cruz.

_____ . *An essay on free will*. Oxford: Clarendon Press, 1983.

VIHVELIN, Kadri. *Arguments for incompatibilism*. In: *The Stanford Encyclopedia of Philosophy*, 2011.

A questão da razão e da responsabilidade e o problema da irracionalidade no agir moral

Ana Paula da Silveira Simões Pedro
Universidade de Aveiro - Portugal

Resumo

A questão central em análise neste artigo é a de discutir que papel a irracionalidade, tomada por meio do exemplo da *acrasia*, ocupa na razão e na responsabilidade do agir moral. A acrasia é aqui entendida como a ação realizada pelo agente, contrária aos seus melhores juízos iniciais; trata-se de uma ação intencional, mas irracional, pois o sujeito parece agir contra aquilo que ele próprio havia deliberado como sendo o melhor. Esse tema foi suscitado a partir do confronto com a tese defendida por Wolf em *Freedom within reason* (1990) e *Asymmetrical freedom* (1980), segundo a qual o sujeito só é livre e responsável se agir em conformidade com a razão. Em consequência, assim será ditada a natureza da responsabilidade moral como assimétrica somente para quem realiza uma ação moral negativa. Essa posição é em tudo semelhante à que Sócrates e Platão defenderam, para quem o sujeito só faz o mal por ignorância. Contrariamente a Wolf, defenderemos a posição de que a ação dos sujeitos pode ser igualmente motivada por causas irracionais. Sustentaremos a ideia de que os sujeitos permanecerão livres e responsáveis mesmo quando agem acraticamente, salvo raras exceções patológicas. Esse fato introduz a ideia de uma certa irracionalidade na ação moral; contudo, defenderemos que o sujeito não pode deixar de ser responsabilizado por isso. Na primeira parte, apresentaremos o princípio wolfiano da razão; na segunda, aduziremos a responsabilidade assimétrica; na terceira, apontaremos limites críticos dos argumentos apresentados, demonstrando a insuficiência de cada uma delas.

Abstract

This article presents an analysis about the the role of irrationality, understood here as acrasia, in questions concerned with reason and responsibility in moral actions. Acrasia is considered as an action performed against its agent´s best first judgements. Even taken as an intentional action it is simultaneously considered irrational since it is not what the agent first deliberated as his or her best action. This argument

goes against Wolf´s thesis on Freedom within Reason (1990) and Asymmetrical Freedom (1980) where a subject is only taken as free and responsible when his or her actions are performed accordingly with reason. This article will defend the the idea that actions can be irrationally caused and still the agent would be considered responsible for such actions.

O princípio da razão

A razão ocupa um lugar central nas explicações de Wolf sobre o modo como poderá contribuir para as ações serem moralmente responsáveis. Depois de ter amplamente criticado o princípio de autonomia[1], que em nada contribuía para o estabelecimento das ações moralmente responsáveis, à medida que as razões do agir autônomo, atuando somente por si, num agir diferente e irracional são, por isso mesmo, de natureza contrária às da própria razão; e depois de se ter dado conta de que o princípio do Eu (*self*) também não constituía uma resposta capaz de fundamentar a responsabilidade dos nossos atos, uma vez que é inaplicável a situações psicológicas específicas que condicionam o uso do "verdadeiro eu" dos sujeitos, ilibando-os, assim, da sua responsabilidade moral (ex: cleptomania, vítimas de hipnose, droga), a resposta ao exercício livre e responsável dos nossos atos só poderia, então, residir na razão.

[1]Apesar de Wolf considerar, inicialmente, a autonomia uma das condições essenciais que caracteriza o agente responsável, entre outras, assim como o controle de si mesmo e a vontade esclarecida, desde logo a crítica, quer porque a autonomia se encontra associada a um requisito impossível de satisfazer – a de *prime mover unmoved* –, *i.e*, a condição *autonomia* nunca seria completamente satisfeita por estar sempre sujeita aos fatores socioeconômicos de que depende para a sua definição – tal como em Kant, em que a vontade só pode ser determinada pela razão e nunca pelas leis naturais que regem os fenômenos – quer porque, inversamente, ser autônomo também pode significar agir de um modo completamente "livre" e não determinado pela razão, – ao contrário do que Wolf e Kant pretendem – de acordo com o qual os **agentes** *not only do make choices on no basis when there is no basis on which to make them, but who also can make choices on no basis even when some basis is available. In other words, they must be agents for whom no basis for choice is necessitating* (WOLF, 1990, p. 55). O que verdadeiramente diferencia as pessoas autônomas das que o não são, é que enquanto as primeiras têm liberdade para agir contra os interesses da razão, as segundas não têm. Ora, agir autonomamente desse modo pode, em última análise, vir a ser um ato de completa irracionalidade porque baseado em escolhas que ignoram totalmente os verdadeiros interesses e razões sem nunca olhar a razão. Por isso, não é de admirar a pergunta que Wolf coloca: *"Why should one want an ability that one never wants to exercise?"* (WOLF, 1990, p. 57). Daí que a autonomia não seja, em seu entender, uma condição necessária à responsabilidade moral, mas sim a razão.

Entendida como *"the highest faculty…that will help us form true beliefs and good values…we need to know whether we have the ability to choose and to act on the basis of the right reasons for choosing and acting"* (WOLF, 1990, p. 70-71), a razão – ao contrário da autonomia, por não se colocar num ponto "neutro" de interesses e de razões morais, condição essa (de liberdade) que lhe é característica (biodirecionalidade) –, só pode agir unidirecionalmente; *i.e*, as da sua vontade. Só que, para isso, terá, antes, de as saber, reconhecer e identificar. Na perspectiva de Wolf, o agente deverá, então, aprender a reconhecer a verdade e o Bem, pois a razão, unidirecional que é, só pode escolher agir de um (único) modo e não de outro(s). Trata-se, portanto, de uma capacidade que permite adquirir crenças verdadeiras em vez de falsas e de crenças boas em vez de más e, consequentemente, *fazer* um bom uso delas. Todavia, não se trata de uma capacidade especial que apenas assiste a alguns seres iluminados e através da qual somente estes tenham acesso privado à Verdade e ao Bem. Ver, apreciar o mundo corretamente significa compreendê-lo não só no que ele tem de verdadeiro ou de falso, mas também no que ao seu valor diz respeito, ou falta dele. Por isso, é que é esta capacidade racional, e não a autonomia, que constitui condição necessária à atuação responsável dos sujeitos, se estes souberem agir em conformidade com as razões daquela. Ser livre e responsável dependerá, assim, da capacidade intelectiva do sujeito para se aperceber das razões corretas (*right reasons*) e para agir em consonância com estas.

Responsabilidade assimétrica

Apesar de, para Wolf, o conceito de responsabilidade constituir *a single feature of agents that makes them suitable objects of either praise or blame* (1990, p. 90), o certo é, porém, que esse está fundamentalmente associado a reações de elogio e não tanto de recriminação. Por exemplo, o sujeito que estiver psicologicamente determinado[2] a agir corretamente e a atender à Verdade e ao Bem, a sua

[2] A esse propósito, a posição compatibilista de Wolf é claramente assumida ao argumentar contra os incompatibilistas: *if we require an agent to be psychologically undetermined, we cannot expect him to be a moral agent. For if we require that his actions not be determined by interests, then a fortiori they cannot be determined by his moral interests* (WOLF, 1980, p. 153). Desse modo, conclui que, em presença da autonomia, a única condição em que o sujeito seria livre, tão livre até dos seus interesses, seria a da liberdade, mas não a dos valores escolhidos pelo sujeito. O dilema que aqui se coloca para os incompatibilistas é o da relação

ação poderá (eventualmente) ser objeto de elogio. É o caso que Wolf apresenta da mulher que salva uma criança de se afogar sem ter de parar para pensar qual deveria ser a melhor forma de atuar[3] naquela circunstância. Mas, se para Wolf a protagonista dessa ação não merece ser (necessariamente) alvo de elogio, pois estava psicologicamente determinada a agir corretamente, já a ação contrária a essa, ou seja, daquele que, por um variadíssimo número de razões, hesita em salvar a criança de se afogar e que acaba por não o fazer, essa ação não implica que o sujeito venha a ser recriminado por isso. O fato de não socorrer alguém que se encontre em apuros não é sinônimo de censura, na perspetiva de Susan Wolf, pois tal significa, apenas, que, naquele momento, o sujeito não é detentor daquela importante capacidade da razão, a que já nos referíamos e, portanto, de *agir* em conformidade com ela. E, porque não nos é possível conhecer exatamente todas as razões que poderão determinar, ou não, se o sujeito é realmente responsável por uma determinada má ação, então, não (podemos) devemos recriminá-lo por isso. Desse modo, a única coisa que importa saber é se, em circunstâncias semelhantes, os sujeitos possuem, ou não, capacidade para discernir de acordo com as razões corretas, dado que, se não conseguiram agir em conformidade com elas, deve ter sido por um conjunto total de vastíssimas e variadíssimas razões às quais raramente teremos acesso. A razão de ser da responsabilidade assimétrica reside precisamente no fato de que quem agiu incorretamente ou mal tinha um conjunto de possibilidades ou de alternativas para agir de outro modo. Contudo, porque não possuía a capacidade racional para conhecer e diferenciar as boas razões das más, não soube agir corretamente, pelo que não pode, sequer, ser responsabilizado pelas más ações cometidas.

---

de exclusão entre liberdade e moralidade, segundo a qual um agente que seja livre não pode ser moral e vice-versa.

[3] Essa posição, que alerta para o fato de não serem necessárias alternativas para agir moralmente em liberdade, é em tudo idêntica à defendida por Frankfurt (1969), como sabemos. Para além disso, a resposta de Wolf assemelha-se, igualmente, em nosso entender, ao raciocínio moral kantiano, à medida que outra solução não parecia existir para aquela mulher que se impusesse com a força de consciência moral (dever) senão a de salvar imediatamente a criança, pelo que não era necessário ponderar quaisquer outras alternativas.

Compreende-se, assim, que, para Wolf, se a determinação psicológica (de agir corretamente) é compatível com a responsabilidade moral de uma boa ação, já o mesmo não se pode afirmar quanto às ações incorretas; porque não houve determinação nessas últimas, não se pode asseverar que os sujeitos que as praticaram sejam moralmente responsáveis pelas suas ações. Assim sendo, *if the freedom necessary for moral responsibility is the freedom to be determined by the True and the Good, then obviously we cannot know whether we have such a freedom unless we know…that there is a True and a Good and…that there are capacities for finding them* (WOLF, 1980, p. 160). Ou seja, para sermos livres e responsáveis necessitamos possuir não só as capacidades desejáveis para reconhecer a Verdade e o Bem, como já afirmamos, mas também uma sensibilidade e percepção que nos permitam reconhecer e identificar as circunstâncias específicas acerca das quais agiremos em conformidade com as razões certas e os valores adequados. Todavia, há que relembrar que esse agir, caracterizado pela determinação do que é a Verdade e o Bem, não deverá ser uma determinação sujeita às influências optativas da liberdade (autonomia); na verdade, esse agir está condicionado pela determinação que escolhe a Verdade e o Bem (razão). Desse modo, um agente só é moralmente responsável se se deixar determinar (influenciar) pelas razões boas e corretas; caso contrário, se não se deixou contaminar intelectualmente por elas ou não as (re)conhece de todo em todo, então, o sujeito que realiza má ação não pode, efetivamente, ser responsabilizado por ela.

Em suma: para Wolf, a condição de atribuição de responsabilidade moral é de natureza assimétrica, pelo que podemos deduzir que não são necessárias possibilidades alternativas (FRANKFURT, 1969) para que uma ação livre e responsável ocorra; o que é absolutamente imprescindível é, pois, a capacidade (liberdade) do sujeito para reconhecer, apreciar e identificar as razões certas, entendida como uma capacidade de *receptivness to the Good* (WOLF, 1990, p. 121) para compreender bem a situação real em que o sujeito se encontra para, depois, agir em conformidade: *the ability to recognize and appreciate the True and the Good refers to…the ability to acquire true beliefs rather than false ones and good values rather than bad ones, and to understand these beliefs and values sufficiently to be able to make proper use of them* (*Idem,* p. 122). Por outras palavras: segundo a tese da assimetria moral, só a pessoa que agiu corretamente pelas razões certas – porque possui essa capacidade e a pôs em ação – é que deve ser responsabilizada e merece

ser reconhecida por tal, ao contrário daquela que não conseguiu, não soube ou não pôde agir do mesmo modo, uma vez que não possui ou não é detentora dessa mesma capacidade[4].

Limites críticos sobre o estatuto da razão e da assimetria moral em Wolf

Estatuto da razão e do desejo nas ações morais

A influência do pensamento socrático-platônico em Wolf é nítida: a concepção intelectualista do agir moral de Sócrates e Platão, segundo a qual o homem erra somente por uma questão de ignorância, é encontrada igualmente em Wolf na justificação do princípio da responsabilidade moral (assimétrica) assente no pressuposto de que os sujeitos que agem erradamente fazem-no, simplesmente, por desconhecimento ou por ausência da capacidade de entendimento (razão); *i.e*, o facto de os sujeitos não agirem corretamente deve-se apenas ao seu desconhecimento do que são a Verdade e o Bem, logo, não podem ser responsabilizados pelo conjunto dessas ações. Essa estrutura de raciocínio é idêntica à que está presente no *Alcibíades*[5], por exemplo, em que o erro é apontado como produto de uma ignorância inconsciente; na *Apologia de Sócrates* (25e), em que é defendida a ideia de que ninguém é mau voluntariamente, e também no *Protágoras*; dentre outros diálogos platônicos, em que Sócrates nega a existência de acrasia, à medida que parte do pressuposto de que é impossível que os agentes não atuem em conformidade com o que consideram ser o *melhor*. Contudo, a ascensão ao *logos* só é permitida a alguns, apenas aos melhores: aqueles que agem mal não o fazem por maldade, mas sim por erro ou desconhecimento. Na continuidade desse pensamento, a acrasia é, então, inexistente e a ação errada é provocada apenas pela ignorância.

[4] Há que entender que, para Wolf, essa é uma questão de natureza normativa e não metafísica e será tanto mais facilmente compreensível e, menos paradoxal, portanto, se considerarmos que a tese da assimetria moral – que confere responsabilidade moral ao sujeito que pratica atos bons e não a confere àquele que pratica atos negativos –parte do pressuposto de que o sujeito atua em função das *right reasons* e não em função da liberdade (autonomia) para agir, segundo a qual o sujeito é autônomo, podendo decidir agir quer num sentido quer noutro, de tal modo que pode mesmo agir contra a própria razão (mas, isso de que lhe serve?). Enquanto a perspetiva normativa, psicológica ou ética é compatível com o determinismo, a segunda já não o é.

[5] Alcibíades, 117d: "não vês tu que os erros da conduta resultam…desse género de ignorância que consiste em crer que se sabe aquilo que não se sabe?".

Mas como compreender que o sujeito (todos os sujeitos), cujas ações não são boas *só* as faz por ignorância, por um lado, e não ter em consideração, por outro, todos os outros casos em que os sujeitos agem contrariamente ao pleno conhecimento da Verdade e do Bem que detêm, sendo movidos por outros fatores como pelo desejo (emoções), por exemplo?

Na verdade, não podemos esquecer que o desejo[6] é o motor essencial das nossas ações e que, enquanto ele move (por vezes) à ação, o intelecto sozinho não o consegue fazer. Tal significa a ação que o agir racional não vai a par do agir moral: frequentemente, o sujeito tem consciência dos juízos morais que enuncia, escolhe e seleciona para si, mas não existe uma relação de proporcionalidade direta entre o pensar e o agir; *i.e*, ao *pensar* não se segue (necessariamente) o *agir*[7]. Só assim se explica que alguns sujeitos, muito embora conhecedores das *right reasons* não querem, deliberadamente, agir em conformidade com aquelas (é o caso, por exemplo, de alguns indivíduos que, embora sabendo que não devem roubar, optam, no entanto, por o fazer)[8].

---

[6] Ponderar essa hipótese equivale a recolocar a questão da existência de ações acráticas, na esteira de Aristóteles - contrariamente a Sócrates, para quem a acrasia era considerada inexistente, reduzindo-a a uma questão de (des)conhecimento – para quem a acrasia é perspetivada mais como um problema moral de conflito entre dois ou mais desejos, tratando-se, portanto, de uma questão de autocontrole, ou falta dele, do que propriamente um problema de conhecimento.

[7] A teoria da motivação para a ação a que acabamos de nos referir formulada a partir da pergunta: *o que é que nos motiva para a ação?,* foi inicialmente enunciada por D. Hume (1711-1776) no *Tratado da Natureza Humana* (1739), segundo a qual são as paixões, ou os desejos, que nos movem para a ação e não a razão; esta, sem aquelas é totalmente incapaz de nos conduzir à ação. Ao invés, Kant, considera que é a razão, e não as emoções, que nos levam a agir. Estas últimas, mais não são do que uma contaminação da razão desviando-nos do seu caminho no cumprimento da lei (dever). Essa teoria da motivação para a ação é caracterizada por alguma controvérsia que oscila entre a aceitação do modelo crença-desejo (Davidson) e a sua rejeição (Nagel). Para um estudo mais pormenorizado das razões pelas quais este autor se opõe ao referido modelo, consultar P. Madeira, (2003). O que é o modelo crença-desejo?, bem como a objeção de Nagel ao modelo crença-desejo (e o realismo moral), em *Intelectu.* n. 9 (2003). Atualmente, cada uma dessas teorias – a teoria sentimentalista de Hume e a teoria racionalista de Kant – conhece os seus representantes mais significativos, respetivamente, em S. Blackburn (humiano) e em T. Nagel (racionalista). Outra vertente relativa à teoria da motivação para a ação é a colocada pelos internalistas (as razões para ação dependem das motivações internas dos agentes) *versus* externalistas (as ações ocorrem independentemente das motivações e têm origem exterior ao sujeito) e que procuram saber qual o estatuto das razões para a ação. (MIGUÉNS, 2004).

[8] Outra questão que vem sendo analisada por alguns filósofos (GREENE & HAIDT, 2002; HAIDT, 2001; ALMADA, 2010) com base nos resultados alcançados pelas neurociências é a

Ora, Wolf é omissa quanto a esses fatores não cognitivos que interferem, transformam e enviesam o agir moral o que, em nosso entender, poderá significar não só maior insuficiência da sua teoria em si e na explicação do agir moral, como também esta concepção conduz, erradamente, em nosso entender, a uma condição assimétrica da responsabilidade moral, como adiante assinalaremos. Para além disso, esse modelo wolfiano de pensamento moral parece basear-se num modelo cartesiano de pensamento que separa a razão das emoções para explicar o comportamento moral livre e responsável quando, afinal, a par da possibilidade do agir intencional, constatamos existirem um conjunto de razões inconscientes para agir, as quais, por sua vez, constituem, igualmente, fonte de explicação para algumas das nossas ações.

Mantém-se, ainda, um outro problema no que ao princípio da Razão diz respeito, ou melhor, no acesso ao que ela permite: *i.e*, à Verdade e ao Bem e, consequentemente, à liberdade e à responsabilidade. A questão é que quando Wolf se refere às primeiras, fá-lo em tom de verdades únicas e instituídas separadas das emoções quando, pensamos nós, aquilo a que o sujeito consegue aceder mais não é do que uma mera interpretação subjetiva daquelas, suportada pelas suas crenças e desejos. O que/quem atribui, então, legitimidade à Verdade e ao Bem? Mais: como conciliar tal postura relativista com a objetividade de valores defendida por Wolf? Na verdade, parece incongruente a utilização do conceito de racionalidade, à medida que é assinalada a importância da sua dimensão mais objetiva de conformidade com as normas e as leis sociais e culturais vigentes num determinado contexto histórico, e não tanto subjetiva, como acontece em Davidson na sua explicação da racionalização da ação. De acordo com essa posição wolfiana, não só se torna possível agir racionalmente bem mas moralmente mal, agir moralmente bem mas sem razões, bem como pode haver uma ação racional e moralmente correta, mesmo que a ação não se dê pelas razões corretas. Ou seja, a incongruência maior que podemos assinalar aqui é que Wolf endeusou tanto a Razão que acabou por cair nas suas próprias malhas, pois, na verdade, eu posso empregar meios racionalmente corretos para responder a uma determinada

que se prende com a teoria da tomada de decisão/ação moral ser fundamental e primeiramente emocional e não cognitiva, bem como sobre o reconhecimento da importância que as emoções têm no estabelecer do raciocínio moral, ao invés do primado da razão.

circunstância moral que exigirá de mim uma ação moralmente responsável sem que, no entanto, eu me identifique racional e moralmente com a ação tomada. Essa postura parece ser condenável em Wolf quer porque a razão se vira contra a própria razão ao pontenciar que tudo (todas as ações) seja racionalmente defensável e permitido, quer porque, do ponto de vista moral, parte do pressuposto de que toda a racionalidade implica (necessariamente uma) moralidade.

Para além disso, apesar de considerar que um agente moralmente responsável deve ter conhecimento do que é o Bem e a Verdade e agir em conformidade, *i.e*, para ser livre e moralmente responsável deverá reconhecer o Bem (*right reasons*), agir segundo os ditames da Razão (*right thing*), obedecendo-lhe (*Reason View*)[9], Wolf não chega a ponderar que os sujeitos necessitem de uma faculdade especial para alcançar a Verdade e o Bem; contudo, o certo é que, de acordo com a potencialidade máxima que ela atribui à Razão, consideramos que essa não estará propriamente ao alcance de todos, à medida que exigirá um aperfeiçoamento e esforço contínuos de consciencialização e reflexão sobre a realidade, garantes únicos (ainda assim relativos) do atingir das *right reasons*. Na verdade, na melhor das hipóteses, essa capacidade estará reservada apenas para alguns, os melhores – tal como em Platão – e são (apenas) esses que conseguem alcançar a Verdade e o Bem. Todavia, parece não só ilógico como impraticável pretender explicar a ação exclusivamente por meio do conhecimento, análise e ponderação de todos os fatores envolvidos e necessários à tomada de decisão/ação, como Wolf defende, à medida que esse princípio de racionalidade pode facilmente ser levado até à exaustão, ou até ao infinito, o que o torna praticamente impossível de realizar.

Ação acrática e irracionalidade

O facto de a razão, por si só, ser insuficiente para explicar o conjunto das ações morais não é, ainda, o bastante para aclarar a natureza da ação moral, à medida que há todo um universo de ações acráticas, por

[9] Gostaríamos de assinalar, igualmente, a proximidade do pensamento de Wolf ao de Kant, à medida que parece tratar-se de um sujeito que age em função de um elevado compromisso moral do dever *i.e*, não só do seu reconhecimento intelectual (*ought*), como também da sua capacidade para a realizar (*can*). Assim sendo, *ought* implica *can* (KANT).

exemplo, que introduz uma certa irracionalidade no agir moral e cujas consequências importa conhecer. Ao abordar essa dimensão, interessa-nos, nesse ponto, demonstrar que a ação acrática, também frequentemente designada de "fraqueza de vontade" ou "incontinência" (MELE, 2010; MAY, J., HOLTON, R. 2010; SANTOS, s/d), não iliba os sujeitos de continuarem a ser responsáveis pelas suas ações, independentemente de agirem ou não contra os seus melhores juízos. Sustentaremos essa ideia nos casos em que se verifique que nenhuma força exterior obrigou o sujeito a agir de forma diferente daquela inicialmente deliberada e decidida por si. Nessas circunstâncias, as condições de raciocínio, intencionalidade e liberdade necessárias e típicas a uma ação moral verificar-se-iam, igualmente, no caso das ações acráticas sempre que o agente agisse contra os seus juízos iniciais, fato esse impeditivo de uma desresponsabilização sobre as consequências das suas ações.

O que define uma ação acrática como tal? Tratar-se-á da disposição que protela uma ação, por isso é uma não ação?[10] Poderá a acrasia ser uma não ação (reação)? Tratar-se-á de realizar uma ação contrária à ponderação e à deliberação dos melhores juízos inicialmente decididos? Por exemplo, o fumante que protela continuamente a sua decisão de deixar de fumar, mas que nunca o consegue fazer, sucumbindo à sua realização. Ou tratar-se-á de uma ação comandada por um conflito de desejos que momentânea e misteriosamente nos invadem e toldam o pensamento anteriormente claro e nos afastam, como um véu cego, da decisão inicialmente tomada, levando-nos a agir contrariamente àquele? Por outro lado, quando confrontados com a necessidade de encontrar uma explicação minimamente racional para essa mudança no percurso de ação previamente delineado e traçado, deparamos-nos com uma espécie de zona "muda", uma caixa de pandora que nos veda o acesso a essa tão desejada explicação. Que sentidos atribuir, então, ao que se passa nessa zona "escura" quando, muito frequentemente, nem nós próprios conseguimos encontrar justificações para as nossas ações que expliquem tal mudança repentina e significativa de comportamento?

[10] Pretende assinalar-se aqui a importância do que consiste em referir o que torna uma ação uma verdadeira ação como tal e não uma mera reação. Nesse sentido, as características de intencionalidade, consciência e voluntariedade são absolutamente essenciais para a sua determinação (DAVIDSON).

Na verdade, se, por um lado, mais parece tratar-se de uma falha de racionalidade (irracionalidade na ação) que nos leva a agir contrariamente ao nosso melhor juízo (DAVIDSON, 1982), por outro lado, entendemos que essa ação não pode ser totalmente desprovida de razões, motivos ou intenções suficientemente bons, válidos e necessários para, naquele momento em particular, servirem de guia à alteração ou à mudança de percurso das nossas ações: "a irracionalidade é uma falha dentro da casa da razão" (DAVIDSON, 1982, p. 289). Mas, nesse caso, tratar-se-ia, então, de um erro de ótica em que o sujeito procede à substituição dos primeiros juízos pelos segundos juízos, estes mais apelativos às suas necessidades e desejos? Ou tratar-se-ia, ainda, de uma "ilusão cognitiva de depreciação" de juízos ou de uma "miopia de preferências", como referem Ainslie & Elster (citados por CORREIA, 2010)? De qualquer modo, parece-nos que o elemento cognitivo está sempre presente nessa explicação da acrasia, como erro ou engano de percepção (entendimento) acerca dos juízos primeiros, motivado e influenciado que está pelos desejos e afetos (irracionalidade cognitiva motivada); ou seja, as decisões cognitivas podem sofrer transformações provocadas pelas emoções e pelos desejos, sendo que estes têm a capacidade de subverter negativamente a racionalidade da ação (ELSTER, 2007). Contudo, é inegável tratar-se de uma ação contrária aos juízos iniciais, pelo que uma certa irracionalidade parece estar fortemente presente.

Simetria *versus* assimetria moral

O princípio até aqui refutado foi o da exclusividade do critério racional para a determinação das ações moralmente livres e responsáveis, quer porque o pensamento não implica necessariamente ação, quer porque o agir humano é naturalmente perpassado por uma certa irracionalidade (acrasia), quer porque, numa apreciação global da teoria wolfiana, tal corresponderia a uma perspectiva cartesiana da mente que separa razão de emoções. O problema que se nos coloca, nesse momento, é o de procurar saber se os sujeitos que agem erradamente e, portanto, acraticamente, são ou não responsáveis pelas suas ações. Se o forem, como assim pensamos, então a tese da assimetria moral de Wolf está errada.

As perguntas a colocar são, para nós, as seguintes: será que a irracionalidade acrática que caracteriza parte das ações de todos os sujeitos poderá ser um fator de impedimento da atribuição de

responsabilidade moral? *i.e*, o sujeito que age contrariamente aos melhores juízos iniciais deixou, por esse fato, de ser livre e autônomo? Se não, como, então, não ser responsável pelos seus atos? Na verdade, e de acordo com Wolf, se o sujeito que age moralmente bem não necessita de possibilidades alternativas para materializar livremente a sua ação, necessitará delas o sujeito que age erradamente ou acraticamente?

Tal como já tivemos oportunidade de nos referir anteriormente, a ideia fundamental da tese da assimetria moral consiste na defesa do princípio da liberdade (razão) para agir em conformidade com a Verdade e o Bem. Todavia, o exercício dessa liberdade, que é só para alguns, como sabemos, é questionável, pois, na perspectiva wolfiana – tal como em Kant – as razões da razão só permitem agir unidirecionalmente; *i.e*, o reconhecimento da Verdade e do Bem impõem-se, por força de lei (moral), à consciência do sujeito que, perante tais razões, só pode agir de uma única forma segundo a qual a vontade se inclina (obedece) perante a Razão (dever). Contudo, para Susan Wolf, só esses agentes livres podem ser considerados responsáveis. Mas, não considerar livres e responsáveis todos aqueles que agiram mal assente apenas no pressuposto do desconhecimento, para além de trazer implicações claramente indesejáveis, não compreende igualmente as principais razões que motivam a ação humana e que radicam no desejo e nas emoções, como já referimos. Afinal, se esses sujeitos que agem erradamente não sofreram constrangimentos exteriores que cortassem a sua liberdade de ação; *i.e*, se não foram coagidos a agir de determinada maneira, então, como não os considerar responsáveis pelos seus atos?

Para além disso, se a razão (pensar) não conduz (necessariamente) à ação, de igual modo, aquela não conduz ao agir moral; *i.e*, o agir racional não é sinônimo do agir moral e responsável e vice-versa.

Entendemos, pois, que tal postura é insuficiente e não pode ser sustentada, a não ser para alguns casos específicos, nomeadamente, aqueles a que Susan Wolf recorre, mas que se enquadram num contexto de patologia clínica (toxicodependência e cleptomania, por exemplo), adquirindo, assim, uma especificidade e um caráter de exceção quanto à natureza da sua consideração. Desse modo, só se justificaria a condição de não atribuição de responsabilidade moral pelos atos cometidos apenas a esses casos específicos, à medida que os sujeitos se encontram, de certa forma, manipulados por um fator de inevitabilidade exterior

que lhes coarta a liberdade de pensamento e de ação, não devendo, portanto, ser alargado a todos os outros casos. Na verdade, a perspectiva assimétrica da responsabilidade moral aqui em causa, sustentar-se-á melhor se pensarmos que os casos a que se refere são casos que procuram explicar alguma excepcionalidade relativa a circunstâncias específicas. Daí que consideremos que todos os sujeitos são moralmente responsáveis pelos seus atos, bons ou menos bons, mesmo que para isso tenham tido a possibilidade (alternativa) de agir diferentemente.

Conclusão

Neste artigo, procuramos criticar quer a posição racionalista de Wolf quanto à natureza do agir moral, quer a sua postura assimétrica no que diz respeito à atribuição de responsabilidade moral do sujeito. Depois de uma breve apresentação inicial das suas principais teorias da Razão e da assimetria moral, argumentamos das suas insuficiências baseadas no modelo crença-desejo, nos princípios de que ao *pensar* não se segue (necessariamente) o *agir* e que o agir racional não é sinônimo do agir moral e responsável e vice-versa, de que as ações acráticas, praticamente experienciadas por todos nós, espelham. Ao mesmo tempo, essas ações também introduzem uma certa irracionalidade no agir moral, bem como assinalam a importância que os desejos e as emoções têm na nossa existência. Por fim, postulamos a natureza simétrica – e não assimétrica – do agir moral, à medida que as ações do sujeito continuam a manifestar condições de racionalidade, de liberdade e de intencionalidade.

Referências

ALMADA, L. F. *As relações neurais de interação e integração entre raciocínio moral e emoções: um diálogo das neurociências com as éticas contemporâneas.* In: *Ethic@: Revista Internacional de Filosofia da Moral*. Florianópolis*.* v. 9, n. 1, pp. 89 – 109, jun. 2010.

CORREIA, V. *Os limites da racionalidade: autoengano e acrasia*. In: *Disputatio.* Lisboa v. III, n. 28, p. 275-291, 2010.

DAVIDSON, D. *Paradoxes of irrationality* In: WOLLHEIM, R. & Hopkins, J. *Philosophical Essays on Freud.* Cambridge: Cambridge University Press, p. 289-305, 1982.

ELSTER, J. *Agir contre soi.* Paris: Odile Jacob, 2007.

FRANKFURT, H. *Alternate possibilities and moral responsibility*. In: *The Journal of Philosophy*. v. 66, n. 23, p. 829-839, Dec. 1969.

GREENE, J. & HAIDT, J. *How (and where) does moral judgment work?.* In: *Cognitive Sciences*. v. 6, n. 12, p. 517-523, Dec 2002.

HAIDT, J. *The emotional dog and its rational tail: a social intuitionist approach to moral judgment*. In: *Psychological Review.* n. 108. pp. 814-834, 2001.

HUME, D. *Tratado da natureza humana.* Tradução de Serafim da Silva Fontes. Revisão científica e prefácio: João Paulo Monteiro. Lisboa: Fundação Calouste Gulbenkian, 2002.

MADEIRA, P. *O que é o modelo crença-desejo?.* In: *Intelectu.* n.9. 2003. Disponível em: http://intelectu.com/intelectu_archive_win_09_04.html. Acesso em 20/jan./12

_____. *A objeção de Nagel ao modelo crença-desejo (e o realismo moral)*. In: *Intelectu.* n.9, 2003. Disponível em: http://intelectu.com/intelectu_archive_win_09_04.html. Acesso em 20/jan./12.

MAY, J. ; HOLTON, R. *What in the world is weakness of will?*. In: *Philosophical Studies* 157. pp. 341-360 2012. Disponível em: http://web.mit.edu/holton/www/pubs/what%20in%20the%20world.pdf.

MELE. *Weakness of will and akrasia*. In: *Philosophical Studies* 150 (3). p. 391–404, 2010.

MIGUÉNS, S. *Racionalidade*. Porto: Campo das Letras, 2004.

PLATÃO. *Alcibíades*. Disponível em: http://www.ancienttexts.org/library/greek/plato/alcibiades1.html. Acesso em 10/dez./11.

_____. *Apologia de Sócrates e Críton. Tradução do grego, Introdução e notas de PULQUÉRIO, Manuel de Oliveira.* Lisboa: Edições 70, 2009.

_____. *Protágoras.* Disponível em: http://www.educ.fc.ul.pt/docentes/opombo/hfe/protagoras/texto/index.html. Acesso em 10/dez./11.

SANTOS, R. *Explicação davidsoniana da akrasia*. In: *Cadernos de Filosofia*. IFL Universidade Nova de Lisboa e Edições Colibir, p.85-104, 1997. Disponível em: http://www.filosofia.uevora.pt/rsantos/RS1997_acrasia.pdf. Acesso em 12 /jan./12.

WOLF, S. *Freedom within reason*. N.Y: Oxford University Press, 1999.

WOLF, S. *Asymmetrical freedom*. In: *Journal of Philosophy* .77 (March), p. 151-66, 1980.

A defesa milliana da liberdade de expressão

Gustavo Hessman Dalaqua
Universidade Federal do Paraná

Resumo

Neste artigo, apresentaremos a defesa da liberdade de expressão empreendida por John Stuart Mill em seu ensaio *On liberty*. Seguindo a divisão original da defesa, separaremos nossa apresentação em três partes. Na primeira, veremos que Mill fundamenta sua doutrina política da liberdade na doutrina epistemológica do falibilismo. Na segunda, explicaremos porque toda opinião, independentemente de sua verdade, deve ser ouvida. Na terceira, ressaltaremos, em oposição a Williams, que a defesa milliana da liberdade de expressão não se confunde com permissividade irrestrita; liberdade de expressão não significa poder falar tudo.

Palavras-chave: John Stuart Mill, liberdade de expressão, falibilismo, Williams

Abstract

In this paper I consider J S Mills free speech defense. First, I explain Mills political doctrine of free speech based on his epistemological doctrine of fallibilism. Then I explain why opinion must be heard regardless of its truth. Finally, I will argue against Williams and sustain that the millean free speech defense does not entail boundless permissiveness: free speech does not mean one can say whatever one may feel like saying.

Keywords: *John Stuart Mill, free speech, fallibilism, Williams*

Introdução

Em 1849, assustado com o recrudescimento do conservadorismo após momento histórico em que a liberdade ganhara as ruas, John Stuart Mill publica a mais apaixonada defesa da liberdade já realizada: o opúsculo *On liberty* (cf. BERLIN, 2002, p. 132). Mill escreve que a liberdade humana compreende três esferas distintas, umbilicalmente laçadas entre si. No corrente artigo, por motivos de concisão,

abordaremos a primeira delas, a saber, a liberdade de pensamento e discussão.

Da liberdade de pensamento e discussão – primeiro ato da defesa

> *No segundo capítulo de sua obra, Mill trata da liberdade de pensamento e discussão. Para ele, a defesa de ambas se dá em conjunto porque uma é inseparável da outra: sem liberdade de expressão não há pensamento livre.*[1] *No entanto, é certo que ambas atuam em domínios diferentes: pela primeira, entende-se "o domínio interno da consciência (...) em seu sentido mais amplo"* [2]*, o que inclui tanto o pensamento quanto o sentimento do indivíduo; pela segunda, a liberdade que o indivíduo tem de "expressar e publicar" esses mesmos pensamentos e sentimentos* (MILL, 1952: 272).[3]

Nessa defesa, a argumentação empregada pelo autor é bipolar. No polo negativo, Mill supõe casos em que não há liberdade de expressão, para daí mostrar o que *ipso facto* se perde. No positivo, o filósofo constrói uma epistemologia para provar a importância dessa liberdade.

Passemos em revista esses casos. Eis o primeiro:

> *Se toda a humanidade partilhasse de uma opinião e apenas uma pessoa partilhasse da opinião contrária, a humanidade não estaria mais justificada em silenciá-la, não mais do que ela estaria justificada em, houvesse o*

[1] "É certo que a condição mais básica para que o pensamento seja livre é a ausência de punições legais para a expressão das opiniões" (RUSSELL, 1996, p. 113, tradução nossa).

[2] É nessa passagem que Arendt se apoia para acusar Mill de ser um dos filósofos responsáveis pelo divórcio entre filosofia e política (cf. ARENDT, 1991, p. 59). No entanto, a frase em questão só é capaz de sustentar semelhante acusação se retirada de seu contexto original, que prega a liberdade de pensamento como indissociável da liberdade de expressão e de associação entre os indivíduos. Assim sendo, é provável que a recusa arendtiana em citar a supramencionada sentença por completo tenha sido deliberada. Pois caso explicasse que a liberdade milliana também compreende a expressão, a ação e a união entre indivíduos, Arendt não poderia, obviamente, matizá-la de apolítica.

[3] Doravante, como essa obra será uma referência constante, limitar-nos-emos a indicar, entre parênteses e no corpo do texto, apenas o número de sua página quando a citarmos. Em contraste, toda e qualquer citação estrangeira a essa obra terá sua fonte citada no pormenor.

> *poder para tanto, silenciar a humanidade* (p. 274-5, tradução nossa).

Era uma vez um homem que pensava que o Sol não girava em torno da Terra. Porém, a seu tempo, ninguém exceto ele pensava assim. Para todos os outros, o contrário é que era verdadeiro. Ora, por que então não silenciar esse homem? Afinal, não haveria mais verdade na crença de muitos que na de um indivíduo isolado?

De modo algum. Primeiro porque, segundo Mill, a crença verdadeira nem sempre é a mais difundida. Como ensina a História, é perfeitamente possível que a crença de todos menos um seja falsa. Aliás, é perfeitamente possível que *qualquer* crença seja falsa:

> *Nunca podemos ter certeza de que a opinião que estamos nos empenhando para reprimir é uma opinião falsa [...]. A opinião que se tenta suprimir autoritariamente pode ser verdadeira. Os que desejam suprimi-la, é claro, negam sua verdade; mas eles não são infalíveis [...]. Recusar-se a escutar uma opinião porque eles têm certeza de que ela é falsa é supor que a certeza deles é a mesma coisa que certeza absoluta. Todo o silenciamento de uma discussão é uma suposição de infabilidade* (p. 275, tradução e grifo nossos).

Eis o primeiro elemento da epistemologia milliana: a concepção falibilista do conhecimento. *Nunca* teremos certeza de que uma dada crença corresponde à verdade. Todo conhecimento é, em princípio, falível – ou o que dá na mesma – falseável. Nada garante, pois, que a opinião da maioria seja verdade. Logo, sem debate livre, a sociedade inteira perde a oportunidade de trocar o falso pelo verdadeiro.

Para ser justo, cumpre-se destacar que a certeza inabalável não tem vez somente no campo do conhecimento. Na religião, certezas absolutas existem e cumprem um papel indispensável: representam o dogma, ponto indisputável e inquestionável da doutrina religiosa que Mill de modo algum pretende negar. O que o filósofo nega, não obstante, é sua veracidade. Diferente do que prega a doutrina católica, os *dogmata*,

contesta o autor, não são "verdades (...) reveladas por Deus"[4]. O dogma não deve ser assumido como verdade porque a

> *[l]iberdade completa de contradizer e desaprovar nossa opinião é justamente a condição que nos justifica em assumir sua verdade [...]; e de nenhum outro modo pode um ser com faculdades humanas possuir qualquer garantia racional de estar certo* (p. 276, tradução e grifo nossos).

A concepção milliana do conhecimento exclui, portanto, o dogma do domínio da verdade: o dogma *não* é falível, *não* é falseável, *não* está aberto à discussão – logo, *não* é verdade.

Assim concebida, a verdade acaba por fazer da liberdade de expressão sua *conditio sine qua non*.[5] O que justifica a verdade de uma opinião é seu teste. A epistemologia milliana, dizíamos, é falibilista. Entretanto, não se deve por isso achar que Mill fosse cético ou relativista. Não se trata de dizer que a verdade inexiste, muito menos de conceder que todas as crenças são igualmente verdadeiras. Para o falibilista, muito embora a certeza absoluta seja inalcançável ao conhecimento, a verdade não precisa ser abandonada. Existem crenças mais verdadeiras que outras, e nós podemos descobri-las, basta testarmo-las. E de que modo testá-las senão por meio do debate livre com teses que lhe sejam contrárias?

De acordo com Mill, o falibilismo não é admitido pela realidade da maioria dos homens:

> *As pessoas [...] depositam [...] confiança ilimitada na [...] infalibilidade do "mundo" em geral. E o mundo, para cada indivíduo, significa a parte que dele ele conhece [...] Tampouco sua fé nesta autoridade coletiva é de algum modo abalada pelo conhecimento de que outras épocas,*

[4] Enciclopédia Católica, 1913, verbete *dogma*. Não obstante essa caracterização, é preciso lembrar que os *dogmata* que integram a doutrina católica, muito embora revelados por Deus, foram aceitos apenas a partir de debate *entre* os homens (cf. SILVEIRA, 2010, p. 49).

[5] Ou, como dirá mais tarde Russell, é a falibilidade de nosso conhecimento que fundamenta a liberdade de expressão: "O argumento fundamental para a liberdade de expressão é a falibilidade de todas as nossas crenças" (RUSSELL, 1996, p.149, tradução nossa). *Vide* também p. 116, em que Russell quase plagia o liame entre liberdade de expressão e falibilismo apontado por Mill.

> *países [...] pensaram, e ainda pensam agora, o exato oposto. [...] Jamais o perturba que foi um mero acidente que decidiu quais destes numerosos mundos seria o objeto de sua confiança* (p. 275, tradução nossa).

Para a maioria, seu conhecimento, porque compartilhado com o "mundo" em geral, é infalível. Porém, na verdade, o mundo não é unívoco. Variadas religiões e países o habitam; várias são, por conseguinte, as "maiorias" que nele vivem. E, ademais, pertencer a esta ou aquela maioria é fruto do acaso. Como bem apontara Montaigne, o fato de ser cristão e católico é fortuito. Por si só, isso basta para evidenciar a falibilidade dessa "autoridade coletiva" da qual a maioria derivaria sua infalibilidade.

Seja como for, talvez fosse possível, reconhece Mill, contestar que quando a maioria pretende silenciar a minoria, ela não pressupõe a infalibilidade de sua opinião. Antes, o que a leva a suprimir a expressão da minoria é a convicção de que as opiniões e ideias desta são falsas. Nesse caso, aventa Mill, a objeção poderia ser assim formulada:

> *Não existe certeza absoluta, mas existe garantia suficiente para os propósitos da vida humana. Podemos e devemos supor que nossa opinião é verdadeira para guiar a nossa conduta: e não estamos a supor nada além disto quando proibimos homens malvados de perverter a sociedade através da propagação de opiniões que temos por falsas e perniciosas* (p. 276, tradução nossa).

Todavia, ainda que admita o falibilismo, ainda que troque "certeza absoluta" por "garantia suficiente", o objetor falha em perceber a promiscuidade que existe entre a doutrina epistemológica do falibilismo e a doutrina política da liberdade de expressão, pois até mesmo a mera "garantia suficiente" da verdade requer um exame atento de suas proposições contrárias. O corolário do falibilismo é, mais uma vez, transformar a liberdade de expressão em condição *sine qua non* da verdade: minha verdade só se justifica conquanto esteja seguro de que "nenhum argumento ou evidência disponível e relevante tenha sido ignorada" (SKORUPSKI, 1991, p. 378). Eis, enfim, a resposta de Mill para a questão central da epistemologia:

> *[O] único modo pelo qual um ser humano pode se aproximar de um conhecimento completo de uma disciplina é escutando o que pode ser dito sobre ela por pessoas das mais variadas opiniões e estudando os vários modos pelos quais ela pode ser vislumbrada [...]. Nenhum homem sábio jamais adquiriu sua sabedoria de algum modo que não este; tampouco está na natureza do intelecto humano tornar-se sábio de qualquer outra maneira. O hábito firme de corrigir e completar sua própria opinião mediante a colisão com a opinião dos outros [...] é a única fundação estável para uma confiança adequada do conhecimento* (p. 276, tradução e grifo nossos).

É pela colisão com opiniões diferentes que o intelecto humano adquire conhecimento. O fundamento do conhecimento é, pois, a liberdade da expressão. Multiplicar a variedade de nossas fontes é, a um só tempo, garantir a veracidade e a expansão do conhecimento. Isso, é claro, se aceitarmos com Mill que conhecimento e verdade são essencialmente incompletos; sempre passíveis de aperfeiçoamento, eles não têm fim. Conhecimento e verdade não provêm da revelação nem da razão pura. Britânico que é, Mill mantém que ambos se originam da experiência. Mas não apenas da experiência – é necessário também "discussão para mostrar como a experiência deve ser interpretada. Práticas e opiniões equivocadas gradualmente cedem aos fatos e argumentos; mas para que fatos e argumentos tenham algum efeito sob o espírito, é preciso que se apresentem" (p. 276). Por isso mesmo, pensamento livre e liberdade de expressão caminham lado a lado: para que o pensamento se liberte do equívoco, ele precisa, antes de mais nada, entrar em contato com os fatos. Minha crença na proposição X é justificada se (i) a evidência fornecida pela experiência me assegura X; (ii) não há, no presente, nenhum argumento ou fato relevante a X que eu ignore. Sem liberdade de expressão, a condição (ii) deixa de ser satisfeita. Meu conhecimento, portanto, perde sua justificação.

Diferente do que prega o Cristianismo, não há verdade revelada por Deus porque não há verdade imutável (salvo as axiomáticas)[6].

[6] Observe-se que, em Mill, o falibilismo, ao contrário de algumas de suas versões contemporâneas, não é radical: ele concede que as verdades axiomáticas são infalíveis. Do mesmo modo, concede que existem verdades objetivas no campo da Matemática.

Abandonando a verdade como Revelação, Mill, por assim dizer, secularia a verdade. A verdade revelada, lembremos, tinha na sobre-humanidade uma de suas maiores características. Ela era revelada *para* os homens e não *pelos* homens. E isso porque, como atestam as Sagradas Escrituras, de duas uma: ou bem a verdade era manifesta "na presença de Deus" (2 Cor 4,2), ou bem a verdade se confundia com o próprio Deus (cf. Jo 14,16). Mill pensa diferente. Segundo ele, a verdade é construída *pelos* homens, mediante o debate *entre* homens. Importante atentar que, no entanto, a verdade não está *nos* homens. É necessário experiência para que se a comprove.

A verdade, assim, desce dos céus e finca seus pés na terra[7]. Terra esta condenada à instabilidade. Noutras palavras, a secularização da verdade ocorre às expensas de sua imutabilidade. Doravante, verdade e conhecimento tornam-se infinitamente perfectíveis. E isso porque se admitem crias de um sujeito igualmente perfectível, qual seja, o homem. O conhecimento brota da experiência, e como a experiência sempre muda, é impossível que ele também não mude. Jamais voltará o homem a pisar em terra firme. O que não deixa de ser, a seu modo, angustiante. Com efeito, o falibilismo milliano tem lá seus aspectos desconcertantes. Com ele, a desconfiança avizinha-se à confiança, perseguindo-nos como uma sombra: a melhor das certezas pode, num abrir e fechar de olhos, desmoronar por completo.

Skorupski matiza a verdade milliana de multifacetada (cf. SKORUPSKI, 1991, p. 382). E é com razão que assim procede: de fato, se o mundo, a experiência e os homens não são unívocos, a verdade também não o é. Ela se deixa mostrar por mais de uma faceta, é acessível por vários lados. Sendo assim, quanto mais homens a perscrutam, mais completa e acurada ela fica. Se a mesma questão se manifesta por vários lados, então por que não analisar todos os lados? O conhecimento é, pois, cumulativo e comunitário. E eis aí outro laço entre pensamento livre e liberdade de expressão: não é sozinho que se conhece a verdade; para alcançá-la, devemos nos juntar aos outros.

[7] Esse ponto é bem captado por Russell e sua distinção entre "a verdade" e "a veracidade" [*truthfulness*]: "*A verdade é para os deuses*; do nosso ponto de vista, ela é um ideal, em direção do qual podemos nos aproximar, porém jamais alcançar. [...] A veracidade [...] consiste no hábito de fundamentar nossas opiniões em evidências e de sustentá-las de acordo com o grau de convicção que a evidência nos assegura. Esse grau *sempre* estará aquém da certeza absoluta, e devemos *sempre* estar prontos a admitir novas evidências contra nossas crenças" (RUSSELL, 1996, p. 149, tradução nossa).

Finalizando a primeira parte de sua defesa, Mill conclui que quem perde com a ausência de liberdade de expressão não é somente a minoria, mas também a maioria. A justificação do conhecimento aumenta conforme a sua extensão: quanto mais fatos conheço, mais verdadeiro e justificado é meu conhecimento. Logo, se a extensão desse domínio me é propositadamente constrangida, meu "desenvolvimento mental é paralisado" (p. 282). Adiante, veremos que isso é uma constante: todos saem perdendo quando a liberdade é cerceada.

Da liberdade de pensamento e discussão – segundo ato

> *Permita-nos ora passar à segunda divisão do argumento e desconsiderar a suposição de que qualquer uma das opiniões recebidas seja falsa. Presumamo-las verdadeiras e investiguemos o mérito da maneira própria para sustentá-las quando não se averigua livre e abertamente a sua verdade* (p. 283, tradução nossa).

Na primeira parte da defesa, estudamos a situação em que todos, menos um, estavam errados. Imaginemos agora o reverso. Suponha-se que noventa e nove vírgula nove por cento creem em X. A despeito de sua assustadora popularidade, há um homem que não acredita em X. Deveríamos, nesse caso, dar ouvidos a esse indivíduo extraordinário? Por que não silenciá-lo? O que ele teria a acrescentar?

É aí que Mill entra em cena e responde: muito. Quando uma pessoa discorda de mim, ela sempre me faz um favor. Pouco importa se sua objeção é certa ou errada – o que importa é que sua objeção está me presenteando com a oportunidade de testar a minha crença. E o teste, vimos alhures, é a única garantia que tenho de que minha crença não é falsa. Se, tendo confrontado uma objeção qualquer, a evidência comprova que minha crença X estava certa, e que a crença não-X era errada, isso não significa que o procedimento tenha sido em vão. Testar nunca é perda de tempo porque é com ele que se aumenta a força de minha convicção. Quanto mais testo X, mais seguro X fica: se repetidamente submeto X a testes, e repetidamente X se prova verdadeiro, então maior é a probabilidade de que X continue a suportar testes futuros.

Suponhamos que haja no mundo um conjunto de fatos (*i.e*, dados empíricos) consecutivos. Digamos que duas teorias tenham sido inventadas a fim de explicar essa conjunção constante, e que ambas, cada

uma a seu modo, satisfaçam este propósito. Qual das duas merece maior confiança? Com Mill, a resposta é simples: aquela cuja verdade tenha sido provada mais vezes. Dito de outro modo, é na teoria mais testada que se deve repousar maior confiança.

Estranho notar que esse princípio milliano possa conflitar com o princípio de economia. Para quem não sabe, o princípio de economia, por vezes também denominado de princípio de parcimônia, afirma que, entre duas teorias igualmente plausíveis, devemos preferir a mais simples. Ocorre que nada impede que a teoria mais testada seja a mais complexa. Aliás, de certo modo, faz sentido pensar que a teoria mais complexa, justamente porque envolve um maior número de teses, seja a mais disputável e, portanto, a mais testada. Nesse caso, a filosofia de Mill, tomando a maior quantidade de testes por critério de seleção entre as teorias, situar-se-ia contra o princípio de parcimônia. Seja como for, visto que esse representaria apenas um dentre vários casos nos quais a aplicação do princípio de economia não é bem-sucedida, não nos delonguemos nesse assunto.

Mill acrescenta outro item na lista de benefícios provenientes da liberdade de expressão. A bem da verdade, trata-se do mesmo item; o que o autor faz é tão somente extrair um novo desdobramento dele. Novamente, é pelo teste que impõe a nossa crença que a objeção equivocada é bem-vinda. A novidade é que agora Mill expõe outro aspecto louvável do teste:

> *Porém, em toda disciplina em que a diferença de opinião é possível, a verdade depende de um balanço a ser dado entre duas séries de razões opostas. [...] deve-se mostrar por que a outra teoria não pode ser verdade; e até que se mostre isto, nós não entenderemos os fundamentos de nossa opinião* (p. 284, tradução nossa).

Além de asseverar a convicção, o teste aumenta a extensão de meu conhecimento de X. Explicando porque não-X não é o caso, eu acabo entendendo melhor porque X é o caso. Nesse sentido, meu entendimento acerca das bases de minha crença é aprofundado.

No limite, o que está em questão é a liberdade de informação: liberar a circulação da informação, tanto as certas quanto as erradas, configura aquela situação que no jargão econômico chama-se de "ganha-

ganha". Quando tudo está aberto à discussão, todos saem ganhando: quem apresenta a tese errada ganha porque depara-se com a oportunidade de trocar o falso pelo verdadeiro; quem apresenta e defende a certa ganha um aprofundamento da compreensão da verdade.

Conhecer crenças diferentes – ou mais amplamente, conhecer o diferente – é, pois, importante. Tão importante que Mill recomenda levantar objeções contra si próprio na ausência de quem o faça:

> *Esta disciplina é tão essencial para um entendimento efetivo dos assuntos morais e humanos que, na ausência de contraditores das verdades importantes, é indispensável imaginá-los e atribuir-lhes os mais fortes argumentos que o mais hábil advogado do diabo poderia maquinar* (p. 284, tradução nossa).

Mill segue à risca seu próprio conselho. Com efeito, tem-se visto até agora que, para toda tese que afirma, a argumentação milliana a conjuga com uma contratese de sua própria autoria. Esse método dialético de proceder não se limita apenas aos primeiros capítulos. Por toda a sua obra, o autor se dedica a formular *ad nauseam* argumentos que o neguem. Se esses argumentos são excessivamente elaborados, é porque Mill considera que os argumentos mais difíceis de serem rechaçados são os que mais proveito oferecem. Não é surpresa, portanto, que o filósofo enxergue no grande orador Cícero um exemplo a ser seguido por todos aqueles que perseguem a verdade.

> *O maior orador de todos [...] sempre estudava a posição do adversário com a mesma intensidade, se não maior, que a sua própria. O que Cícero praticou [...] requer imitação da parte de todos [...] que visam chegar à verdade. Quem conhece do caso apenas o seu lado, pouco conhece dele. [...] ele deve sentir a força da dificuldade na sua inteireza [...]; do contrário nunca possuirá a porção da verdade que supera esta dificuldade* (p. 284, tradução nossa).

Na interpretação de Mill, a objeção deve ser não só ouvida, como também analisada com cuidado; se duvidar, até com mais cuidado do que a nossa. Pois é aí que os argumentos opostos se fazem sentir com maior força. É necessário esforço para colocar-se do outro lado da questão. Se,

por exemplo, meu intento é defender a teoria heliocêntrica, é preciso estudar todas as outras teorias concorrentes com afinco. Para explicar que minha teoria é o caso, tenho de explicar o porquê de as outras não o serem.

Ao fazer isso, adquiro (i) mais convicção, (ii) mais conhecimento e – eis aí um elemento novo – (iii) mais verdade. Outra vantagem, pois, do teste proporcionado pela discussão livre é a ampliação da verdade. Invalidar a objeção amplifica a verdade à medida que sua falsidade requer verdade para ser rechaçada. Entretanto, é óbvio que isso não se aplica a qualquer objeção. As objeções cuja refutação não exige nada de novo de nossa tese, de fato, não a acrescentam verdade alguma. Um exemplo é refutar duas vezes a mesma objeção. Se alguém me desafia hoje com uma contratese, a novidade dela me fará bem porque me apresentará uma dificuldade inédita – o que, por sua vez, incrementará minha convicção, conhecimento e verdade. No entanto, se amanhã sou confrontado com a mesma contratese, o desafio cessa; será impossível auferir qualquer vantagem dessa objeção porque, propriamente falando, ela já não me apresenta obstáculo algum.

Anteriormente, mencionamos que a verdade, segundo Mill, é multifacetada. No trecho descrito, percebe-se que, além de multifacetada, a verdade é fragmentária. Mill a concebe como que dividida em partes. Essa divisibilidade possibilita, por seu turno, uma descontinuidade, o que não nega a cumulatividade do conhecimento. Porquanto mesmo que a passagem de um paradigma teórico para outro ocorra mediante uma ruptura (o que geralmente acontece)[8], é inegável que a formulação deste depende daquele. Não há criação (científica) *ex nihilo* (cf. p. 289). Ainda que a transição a um novo paradigma seja, à medida que não segue qualquer princípio em voga, descontínua a outro antigo, ela, precisamente por não segui-lo, se remete – negativamente e *in absentia*, mas se remete – a este.

Alguém poderia objetar a Mill que, não obstante as vantagens da liberdade de pensamento e discussão, quiçá fosse melhor não largar a difusão de informação ao completo *laissez-faire*. "Sem dúvida" – argumentaria o opositor – "concordo que o debate seja necessário ao progresso e que as informações equivocadas lhe sejam indispensáveis. Mas não julgo que, por conta disso, justifica-se a disponibilização

[8] Ver KUHN, 2005.

indiscriminada destas últimas a todos. Existem pessoas cujo preparo intelectual é insuficiente para enfrentar a falsidade. A elas, a informação equivocada, além de nenhum ganho acrescentar, tem o inconveniente de poder desviá-las para sempre do caminho da verdade. Nesses casos, o debate livre é mais maléfico que benéfico. Portanto, creio que o mais correto é destinar as informações de falsidade corruptora apenas aos mais intelectualmente capacitados".

Skorupski chama o argumento descrito de "falácia do burocrata" (*op. cit*. p. 387) e afirma que ele é característico da política governamental dos países comunistas. Seja como for, certo é que nada há de novo nessa tática. De fato, a mesma medida era praticada pela Igreja Católica no século XIX:

> *A Igreja Católica [...] separa aqueles a que permite receber as suas doutrinas por convicção, dos que devem aceitá-la por confiança [...] o clero pode conhecer os argumentos dos oponentes a fim de os responder, podendo, portanto, ler livros heréticos – o que para os leigos demanda uma licença especial, difícil de obter* (p. 285, tradução nossa).

Mill repugna essa atitude porque a considera contraproducente.[9] Ou seja, ela sabota aquilo mesmo que pretendia defender: o progresso do conhecimento. Ao confinar o desenvolvimento do conhecimento a uma elite, um número considerável de contribuintes é excluído. E é possível que um desses se tornasse um grande gênio, um ajudante excepcional na busca pela verdade. Todavia, se excluído das discussões intelectuais, tal gênio não se desenvolveria, o que, no final das contas, seria ruim para a sociedade inteira.

Em segundo lugar, não esqueçamos que o falibilismo também vale para a *intelligentsia*. Ora, se é assim, eles também são suscetíveis ao erro; logo, nada garante que eles não errem na hora de escolher quem atende e quem não atende ao critério de ser "intelectualmente preparado" (*i.e*, quem pode e quem não pode participar das discussões).

[9] A resposta de Mill à falácia do burocrata ecoa, nesse respeito, àquela dada por Kant em seu opúsculo sobre o Esclarecimento quanto ao porque não devemos permitir que uma elite domine e controle o debate advindo do uso público da razão. (Ver KANT, 2010, p. 411-2)

Ademais, o próprio critério de seleção aqui é profundamente problemático. Que requisitos haveriam de ser cumpridos para que um sujeito se classificasse como "intelectualmente capaz"? E a quem delegar a tarefa de classificar os homens? Essa decisão é ainda mais espinhosa que aquela e, ainda por cima, perigosíssima. Com efeito, é preferível deixar que a informação corra solta a confiar o futuro do conhecimento a uma classe privilegiada da população; porque "qualquer elite (...) irá desenvolver interesses coniventes a sua classe social e doutrinas ideológicas incriticáveis para sustentar estes interesses" (SKORUPSKI, 1991, p. 386). Marilena Chauí definiu ideologia como um "ocultamento da realidade" (CHAUÍ, 2001, p. 8). Nesse sentido, a ideologia presta um desserviço ao conhecimento, regredindo e afastando-o da verdade. E o pior é que, nessa situação, estaríamos completamente à mercê dela. Nenhuma contestação seria capaz de derrubar a ideologia, visto que a condição para contestar seria a adesão a esta.

Resumidamente, vimos que testar uma crença é desejável porque gera (i) mais convicção, (ii) mais conhecimento e (iii) mais verdade. Agora, o último argumento que Mill usa em sua defesa apela para a tese de que (iv) é pela discussão livre que as crenças ganham significado:

> *[P]oder-se-ia pensar que, se isso [sc. a ausência de debate livre] é um dano intelectual, não o é moral, e não atinge o valor das opiniões [...]. O fato, contudo, é que na ausência de debate não apenas se esquecem os fundamentos das opiniões, mas ainda, muito frequentemente, o próprio significado delas* (p. 285, tradução nossa).

Esse quarto efeito pernicioso, advindo da ausência da discussão livre, liga-se aos três primeiros. Aliás, os quatro efeitos completam-se perfeitamente. Se não me é permitido questionar uma determinada crença, meu assentimento a ela, devido a (ii), não é fundamentado. Em consequência, a crença perde (i) seu poder de persuasão; (iii) a confiança que nela deposito, ou seja, a quantidade de verdade que a atribuo. Depois de tudo isso, enfim, não é por acaso que a crença não me seja significativa.

Sendo assim, quando se proíbe a discussão, o próprio caráter da crença é afetado. De acordo com Mill, a crença vira então uma palavra vazia, uma fórmula a que se apega irrefletidamente, um clichê. Tudo se passa como se a proibição da discussão da crença afetasse o próprio

pensamento. A ausência de liberdade de expressão macularia, assim, a liberdade do pensamento. Seu exercício não sendo mais requerido, atrofia-se o pensamento. A consciência fica condenada, pois, a um papel eminentemente passivo:

> *Mas quando ela se torna um credo hereditário, recebido passivamente, e não ativamente – quando o espírito não é mais compelido [...] a exercer seus poderes vitais nas questões que suas crenças o apresentam – há uma tendência progressiva em esquecer toda a crença exceto os formulários, ou a dar-lhe um assentimento amorfo e entorpecido. Como se aceitá-la em confiança dispensasse a necessidade de vivê-la amplamente na consciência, ou de submetê-la à prova da experiência pessoal, fazendo-a perder sua ligação com a vida interior do ser humano* (p. 286, tradução nossa).

Se a consciência não é mais convidada a pôr uma dada teoria em questão, a teoria cessa de ser pensada. Ela passa a ser aceita sem mais, entra por um ouvido e sai por outro, sem acrescentar nada ao ouvinte. Decerto que continua a ser uma crença, mas há uma diferença muito grande, segundo Mill, em acreditar em algo porque você, após ter pensado e inquirido acerca de sua validade, concluiu que ele é o caso, e acreditar simplesmente porque outros lhe mandaram que o fizesse. No primeiro caso, você testou a teoria, ou seja, submeteu-a ao método experimental. Com isso, você a dotou de significado e verdade, duas qualidades que só a experiência é capaz de formar. Conferiu-a, assim, com aquele "respeito sincero que a razão atribui apenas àquilo que se submeteu ao teste do debate livre e aberto" (KANT, C. Juízo, §22, segundo cita ARENDT, 1992, p. 32). No segundo caso, é até possível que o indivíduo – ou até inevitável, dependendo de seu estado mental – creia naquilo a que foi compelido por outrem a crer, assim como é igualmente possível crer em um dogma sem nunca tê-lo questionado. Porém, de acordo com a epistemologia milliana, semelhante crença não "merece o nome de conhecimento", porquanto carece tanto de verdade quanto de significado (p. 288).

Havíamos lido primeiro que Mill suprimira o dogma do domínio da verdade. Estaria ele agora excluindo-o do campo do significativo? Certamente. Se o crente não se incomoda a ponto de questionar o princípio religioso que segue, esse seguimento, afirma o filósofo, não é

significativo. Inclusive, tão forte é essa tese em Mill que o exemplo a que o autor recorre para ilustrar a inocuidade da crença que jamais se questiona é justamente o dogma. Na sua visão, havia uma discrepância entre o discurso e a prática religiosa. Eis o discurso religioso de sua época: "Todos os cristãos acreditam que os abençoados são os pobres, (...) que a eles não compete julgar, que devem amar seus vizinhos como a si mesmos" (p. 286). A conduta, contudo, destoava do discurso; na Inglaterra vitoriana, as pessoas que mais atiravam pedras eram justamente aquelas que acreditavam que não se devia jogar pedras, que oravam "perdoai nossas ofensas assim como nós perdoamos a quem nos tem ofendido". Essa hipocrisia, para o autor, acontecia porque a crença religiosa não era significativa. Logo, ela não influenciava a conduta do crente porquanto estes não tinham aquele "sentimento que emana das palavras para as coisas significadas, e que estimula o espírito a interiorizá-las" (p. 286). É a partir de então – do momento em que deixa de fazer sentido – que a Religião se subtrai do mundo e cessa de influenciar e alimentar o espírito e a conduta dos homens, donde se segue que processos como o "niilismo europeu" e a "secularização" ganhem força.

Crenças irrefletidas, por não exigirem atividade alguma do pensamento, não são interiorizadas pelo indivíduo que as recebe. Apenas se sedimentam sob a consciência, sem transformá-la ou aperfeiçoá-la. É de se supor, portanto, que a sociedade em que os indivíduos são privados de empregar seu senso crítico no debate livre é uma sociedade de cabeças ocas. Com efeito, Mill declara precisamente isto:

> *Assim, veem-se os casos, tão frequentes nesta época do mundo que quase constituem a regra, nos quais o credo permanece como se fora do espírito, incrustando-o e petrificando-o contra [...] as partes mais elevadas de nossa natureza; manifestando seu poder [...] ao nada fazer pelo espírito ou pelo coração, a não ser montar guarda sobre eles a fim de mantê-los ocos.* (p. 286, tradução nossa).

Sedimentando-se tão somente na superfície do pensamento, a crença irrefletida acaba por atrofiá-lo. Petrificado, cada vez mais improvável torna-se sua emancipação, porquanto se encontra sitiado, sob a influência de um poder que o oprime, a fim de garantir que continue vazio. Nesse sentido, não surpreendentemente, a ausência de

liberdade de expressão pode vir a servir interesses escusos. É dessarte que os homens, mantidos na menoridade, são mais facilmente manipuláveis (cf. KANT, 2010, p. 407). Não é mera coincidência, portanto, que a censura seja um denominador comum a toda espécie de fascismo que até hoje se conheceu.[10]

Mill alerta que a maioria das pessoas da sua era vivia sob esse domínio, com a consciência carregada de crenças que nunca se importavam de questionar. Sendo assim, pouco faltava para que a sociedade inglesa caísse na tirania da maioria, e talvez fosse para salvar a Inglaterra desse fascismo que Mill dedicara-se tão apaixonadamente à defesa da liberdade.

No que concerne ao fascismo, é curioso ressaltar o quão próximo, nesse ponto, Mill está a Adorno. Segundo este, a sociedade de massa (a moderna república democrática sendo um exemplo óbvio)

> *Perfaz-se sempre pela subjugação de alguns [sc. a minoria] por muitos: a opressão da sociedade exibe (...) os traços da opressão exercida por um coletivo. É essa unidade de coletividade e dominação (...) que se sedimenta nas formas de pensamento* (ADORNO, 2005, p. 39).

Mais intrigante ainda é perguntar se houvesse Mill vivido no século XX, teria ele concordado com Adorno?

Outro autor cuja menção vem a calhar nesse momento é Ray Bradbury.[11] Tal qual Adorno, Bradbury também não era muito otimista com relação ao futuro das sociedades de massa. Em *Fahrenheit 451*, Bradbury retrata uma sociedade distópica na qual a censura atinge seu grau máximo. Nela, os bombeiros não apagam o fogo. A única coisa que apagam é a consciência das pessoas. Sua responsabilidade é queimar todo e qualquer livro que cruzar o seu caminho, visto que em *Fahrenheit*

[10] "Pessoas que são privadas da discussão livre ficam retardadas e mesquinhas – ficam vulneráveis (...) à paranóia, à agressividade defensiva que decorre da ignorância e falta de autoconfiança, à exploração de demagogos" (SKORUPSKI, 1991, p. 387).

[11] Poder-se-ia criticar que é descabido de nossa parte aliar Mill a Bradbury. No entanto, julgamos que esse alinhamento não é forçado, não somente porque *Fahrenheit 451* ecoa várias das teses millianas, como também porque Skorupski faz o mesmo, e vai ainda mais longe (na página xii de seu *Prefácio*, o comentador compara Mill a um autor distópico mais radical que Bradbury, qual seja, Orwell).

os livros são proibidos, e seus donos são contraventores. O protagonista da estória é Guy Montag, um bombeiro exemplar que começa a ter suas crenças abaladas quando conhece Clarice, uma adolescente diferente que ao invés de ficar trancada em casa ou na escola vendo tevê, vivia na rua rolando na grama, colhendo frutas, cheirando e esfregando flores no queixo, bebendo água da chuva, ou fazendo qualquer outra "esquisitice". Um dia, Montag se intriga com a atitude da moça e pergunta:

> *"Por que você não está na escola? Todo dia te vejo vagando por aí."*
>
> *"Ah, eles não sentem minha falta", ela disse. "Dizem que sou antissocial. Não me misturo. É tão estranho. Eu acho que sou muito social. Tudo depende do que você entende por social, não é? Social para mim significa conversar sobre coisas como estas (...) ou conversar sobre como o mundo está estranho. (...) Mas não acho nem um pouco social botar um bando de gente junto e não deixar que falem, não acha? Uma hora de vídeo-aula, uma hora de basquete ou beisebol ou atletismo, outra hora de (...) esportes, mas sabe que a gente nunca faz pergunta (...); eles só jogam um monte de respostas em você, bing, bing, bing, e a gente sentado lá para mais quatro horas de vídeo. Para mim isso não é nem um pouco social. É um monte de funil e um monte de água jorrando por um lado e saindo por outro, e então eles nos falam que é vinho quando [na verdade] não é. Deixam a gente tão débil no final do dia que não conseguimos fazer nada a não ser ir para cama, ou (...) intimidar as pessoas, quebrar janelas (...) ou destruir carros (...). Não tenho nenhum amigo. Isto já basta para provar que sou anormal. Mas todo mundo que conheço está gritando ou dançando feito maluco (...). Tenho medo dos jovens da minha idade. Eles se matam uns aos outros. (...) Mas o que eu mais gosto," ela disse, "é de ficar olhando as pessoas. Às vezes ando de metrô o dia inteiro e fico as olhando e escutando o que dizem (...) e sabe de uma coisa?"*
>
> *"O quê?"*
>
> *"As pessoas não falam sobre nada."*

> *"Ah, claro que falam!"*
>
> *"Não, nada mesmo. Elas falam um monte de nomes de carros ou roupas (...) e dizem que legal! Mas falam sempre a mesma coisa e ninguém nunca diz uma coisa diferente* (BRADBURY, 2003, p. 29 – 30, tradução nossa).

No trecho descrito, uma triste convergência desponta entre as observações de Clarice e as de Mill. Triste porque embaralha as fronteiras entre o surreal e o real; *a fortiori* o ambiente fictício de *Fahrenheit* não está muito longe de nossa sociedade, assim como não estava muito longe da Inglaterra de Mill. Segundo Clarice, a escola se resumia a um banho incessante de respostas prontas e mentirosas. Os alunos não tinham liberdade para questionar. A escola, assim, reproduzia a ausência de liberdade de expressão da sociedade, ausência esta que tornava os alunos débeis.[12]

Clarice comenta que todas as pessoas eram iguais (falavam "sempre a mesma coisa"). Seria forçoso estabelecer uma relação causal entre esse fato e a ausência de liberdade de expressão? Segundo a nossa interpretação, é provável que não. De fato, frustrar o debate livre, à medida que abafa a pluralidade e homogeneíza o discurso da sociedade, contribui para a estandardização (e também idiotização) dos homens. Todos passam então a entoar o mesmo canto. Não é sem razão, portanto, que Adorno pensasse que "a opressão da sociedade" visava à eliminação do diferente, criando "a falsa identidade da sociedade e do sujeito" (ADORNO, 1985, p. 127).

Todavia, note-se que para Mill uma ressalva aqui seria preciso. Conforme sua filosofia, como vimos, a diferença e a pluralidade são inerentes à natureza humana. Logo, é inexequível, dada à constituição do homem, extirpar a diferença por completo. A própria constatação de Clarice atesta isso (pois se realmente *ninguém* pensasse diferente, Clarice jamais teria formulado essa frase). Dito de outro modo, indivíduos excepcionais são capazes de florescer em "uma atmosfera geral de escravidão mental" (p. 283). O mesmo, no entanto, não se aplica

[12] Debilidade que, de acordo com o relato de Clarice, ocasionaria a banalização do mal. Seguramente, Arendt concordaria com Clarice a esse respeito. Afinal, não seria Eichmann, ícone da banalidade do mal, um sujeito débil que seguia as ordens impostas feito mula, sem jamais questioná-las, que abusava dos clichês (*i.e*, palavras vazias de significado repetidas à exaustão pela maioria) para justificar suas ações? (cf. ARENDT, p. 2005).

ao grosso da população. Via de regra, para que os homens se libertem, ou mais especificamente, para que se sintam estimulados a desenvolver sua individualidade própria, eles dependem da manutenção de um espaço público receptivo à mudança. Nesse sentido, um sujeito comum dificilmente conseguiria ser livre em uma sociedade governada por um estado de escravidão mental.[13]

Da liberdade de pensamento e discussão – terceiro ato

> *Consideramos até aqui apenas duas possibilidades: que a opinião recebida seja falsa e, consequentemente, alguma outra opinião verdadeira; ou que seja verdadeira a opinião aceita, caso em que um conflito com o erro oposto é essencial a uma apreensão clara e a um sentimento profundo de sua verdade. Mas há um caso mais comum que qualquer um destes.* (p. 288, tradução nossa).

Os dois casos que analisamos até agora não representam, de acordo com Mill, o mais comum. Geralmente, não é o caso que todos menos um estejam certos, tampouco que estejam errados. No mais das vezes, o que acontece é que todos estão um pouco certos: "as doutrinas conflitantes, ao invés de ser uma verdadeira e a outra falsa, partilham a verdade entre elas" (p. 288).

Num tal caso, a liberdade de expressão permanece tão imprescindível quanto antes. A manutenção do debate diversificado de opiniões fornece a cada uma "o restante da verdade da qual cada doutrina possui apenas uma parte" (p. 288). Tudo se passa, portanto, como se a verdade se encontrasse dividida entre os homens – o que novamente reforça seu caráter fragmentário. A verdade completa então resultaria da confluência das verdades parciais espalhadas nas opiniões individuais. Sendo assim, mais uma vez, o diálogo livre entre os homens é *conditio sine qua non* da verdade. Visto que cada opinião em isolado é parcial, um intercâmbio com opiniões diferentes é preciso. Ademais, nada mais natural que nossas opiniões estejam, via de regra, fadadas à parcialidade, uma vez que a verdade somente nos é acessível a partir de *um* ângulo circunscrito.

Disso se segue que toda opinião deve ser valorizada. Por mais bizarra que uma opinião contrária soe, é sempre possível que ela

contenha um fundo de verdade. No concernente a esse ponto, um exemplo é útil para aclarar nossa compreensão:

> *[N]o século dezoito, quando quase todos os instruídos [...] admiravam o que se chama de civilização [...] com que salutar abalo explodiram os paradoxos de Rousseau [...] deslocando a massa de opinião unilateral e forçando seus elementos a se reajustarem em melhor forma e com ingredientes novos* (RIESMAN, p. 289, tradução nossa).

Quando da publicação do *Discurso sobre as ciências e as artes*, Rousseau escandalizou a Academia. Elaborado como resposta à questão "O Reestabelecimento das ciências e das artes terá contribuído para aprimorar os costumes?", proposta pela Academia de Dijon, o primeiro *Discurso* chocou a *intelligentsia* oitentista justamente porque optou por uma negativa num momento histórico em que a glorificação do Renascimento era unânime. Tamanho foi o choque que, num certo sentido, o filósofo acenado por Mill seria mais bem aproveitado se invocado no primeiro ato da defesa. Seja como for, muito embora não fosse inteiramente acertada, Mill crê que, não obstante, a filosofia rousseauniana foi de suma importância para o debate da época porque chamou atenção para um fragmento da verdade que o discurso corrente obliterava. Nomeadamente, este se resume no fato de que os costumes e a polidez desmoralizam a natureza humana. Na visão de Mill, afirmá-lo "é preciso mais do que nunca" porque, como já tivemos oportunidade de explicar, na Inglaterra do século XIX, era senso comum encarar costume e natureza humana como sinônimos. A esse respeito, Mill concorda com Rousseau: para o inglês, o costume, longe de equivaler a um preceito do qual o indivíduo não deve desviar, não integra a natureza humana. O costume é adquirido, e não natural.

Tendo explicitado esses três casos, o filósofo conclui a defesa da primeira liberdade estipulando um limite para ela: as "opiniões perdem sua imunidade quando (...) sua expressão constituiu uma instigação positiva para algum ato danoso" (p. 293). Liberdade, portanto, não se confunde com permissividade irrestrita. Desmentindo as críticas com que Williams o ataca, Mill não considera que um debate livre anárquico é garantia da verdade (cf. WILLIAMS, 2002, p. 212f.). Decerto que para ele a "discussão livre" é condição *sine qua non* da verdade.

Todavia, o que por aqui se designa como discussão livre comporta limites, por mais vagos que sejam,[13] até porque, como argumenta Williams e como demonstram alguns verbetes polêmicos da *Wikipedia*, um debate livre permissivo é capaz de prestar mais serviço à calúnia e à mentira que à verdade (cf. *ibidem*., p. 213 e GOPNIK, 2011, p. 126).

Um sujeito que, por exemplo, descobre como fabricar uma bomba misturando sal e óleo de cozinha é passível de ter sua liberdade de expressão constrangida ao querer divulgar sua receita em rede nacional, porquanto, nesse caso, a opinião expressada instigaria dano à sociedade. O cerceamento da liberdade de expressão é, pois, justificável quando provoca dano à sociedade. Eis que nos deparamos, pois, com o princípio do dano: é legítimo, segundo Mill, cercear a liberdade individual se e somente se o exercício dela constitui dano à sociedade.

Conclusão

Conclui-se, pois, que a liberdade de expressão é imprescindível para o desenvolvimento científico, e também moral, dos seres humanos. A liberdade de expressão é o único meio de justificar e salvaguardar o conhecimento e a própria humanidade dos indivíduos. Desmentindo a noção de que o liberalismo implica a atomização dos homens, Mill postula que o conhecimento e a descoberta da verdade são atividades essencialmente comunitárias. Nesse sentido, defender a liberdade de expressão é preciso porque é a partir dela que conheço a verdade e a minha própria pessoa. A ausência de liberdade de expressão confisca dos indivíduos não só a verdade como também aquela bênção tão cara a cada um de nós: a chance de poder descobrir, conhecer e expressar sua individualidade. A chance de, numa palavra, poder viver.

[14] Eis os vagos limites que Mill estipula para a liberdade de expressão: *free expression of all opinions should be permitted, on condition that the manner be temperate, and do not pass the bounds of fair discussion* (p. 292). Essa vagueza, contudo, é proposital; a fim de delinear um modelo político aplicável a uma infinidade de contextos, Mill formulou sua doutrina da liberdade em termos largos. No que exatamente consiste uma *fair discussion*, e pelo que se define o conturbado conceito de "dano", cumpre à sociedade discutir. De sorte que o que constitui dano a uma sociedade não necessariamente o constituirá noutra (na Alemanha, por exemplo, devido às suas circunstâncias históricas, decidiu-se que a negação do Holocausto não é protegida pelo debate livre, uma vez que é danosa à sociedade. No Brasil, ao contrário, semelhante atitude é protegida por nossa liberdade de expressão, já que, dada a nossa situação, a negação do Holocausto não implica dano).

Referências

MILL, J. S. *On liberty*. In: J. S. Mill, Col. Great Books of the Western World, vol. 43. Illinois: Enciclopédia Britânica, 1952.

ADORNO, T. & HORKHEIMER, M. *Conceito de Iluminismo*. In: *Textos escolhidos*. Col. Os Pensadores. São Paulo: Abril Cultural, 2005, p. 17–64.

_____. *A indústria cultural: o esclarecimento como mistificação das massas*. In: *Dialética do esclarecimento*. Rio de Janeiro: Zahar, 1985, p. 99–138.

ARENDT, H. Eichmann and the Holocaust. In: *Col. Great ideas*. Nova Iorque: Penguin, 2005.

_____. *Freedom and politics*. In: *Liberty (Oxford readings in Politics and Government)*. Nova Iorque: Oxford University Press, 1991.

_____. *Lectures on Kant's political philosophy*. Illinois: University of Chicago, 1992.

BERLIN, I. *John Stuart Mill and the ends of life*. In: *J. S. Mill On Liberty in focus*. Londres: Routledge, 2002.

BRADBURY, R. *Fahrenheit 451*. Nova Iorque: Ballantine, 2003.

CHAUÍ, M. *O que é ideologia?* In: Coleção Primeiros Passos. São Paulo: Brasiliense, 2001.

GOPNIK, A. *The information – does the internet change how we think?* In: *New Yorker*, edição de aniversário 2011, p. 124–130.

KANT, I. *Resposta à questão: o que é o esclarecimento? (Trad. de Vinicius de Figueiredo)*. In: *Antologia de textos filosóficos*. Paraná: SEED, 2010.

KUHN, T. *A estrutura das revoluções científicas*. São Paulo: Perspectiva, 2005.

MILL, J. S. *On liberty*. In: J. S. Mill, Col. Great Books of the Western World, v. 43. Illinois: Enciclopédia Britânica, 1952.

RIESMAN, D. *The lonely crowd*. Connecticut: Yale University Press, 1989.

ROUSSEAU, J-J. *Discurso sobre a origem da desigualdade entre os homens*.

In: _____, *Coleção Os Pensadores*. São Paulo: Abril Cultural, 1978.

_____. *Discurso sobre a ciência e as artes*. In:_____, *Coleção Os Pensadores*. São Paulo: Abril Cultural, 1978.

RUSSELL, B. *Freedom versus authority in education*. In: *Sceptical essays*. Londres: Routledge, 1996.

_____. *Free thought and official propaganda*. In: *Sceptical essays*. Londres: Routledge, 1996.

_____. *Freedom in society*. In: *Sceptical essays*. Londres: Routledge, 1996.

SILVEIRA, D. *A elaboração dos dogmas*. In: *O sagrado na história do cristianismo*. São Paulo: Duetto, 2010.

SKORUPSKI, J. John Stuart Mill. In: *Col. The arguments of the philosophers*. Nova Iorque: Routledge, 1991.

WILLIAMS, B. *Truth and truthfulness*. Nova Jersey: Princeton, 2002.

A reformulação do liberalismo clássico por John Rawls

Leno Francisco Danner
Universidade Federal de Rondônia

Resumo

Este artigo tem por objetivo refletir sobre a reformulação do liberalismo clássico (especificamente na variante política de Locke e na variante econômica de Adam Smith), no intuito de demonstrar, com base nas críticas de Hegel e de Marx ao liberalismo clássico, que Rawls retoma sua própria posição em relação a esse mesmo liberalismo clássico. Com isso, defenderei que, muito mais do que visar a uma crítica à versão clássica, Rawls tem em mente a crise do estado de bem-estar social e, em relação a ela, uma crítica à resposta neoliberal que apelaria para a volta de alguns dos princípios básicos da economia de *laissez-faire*.

Palavras-chave: Rawls, liberalismo clássico, economia de *laissez-faire*; estado de bem-estar social, neoliberalismo

Abstract

This paper aims at reflecting on Rawls's reformulation and critics to classical liberalism (specifically Locke's classical political and Adam Smith's classical economic liberalism). It will be demonstrated that since Hegel and Marx's critics to classical liberalism, Rawls retakes his own conception related to classical liberalism. I will defend that Rawls is fundamentally critic to neo-liberalism and not of classical liberalism and that the welfare state's crisis and the neo-liberal to this issue is the central point of Rawls's.

Keywords: *Rawls, classical liberalism, economy of laissez-faire, welfare state, neo-liberalism.*

É interessante perceber que os trabalhos de Rawls, à medida que se propõem reformular algumas das premissas básicas do liberalismo clássico (seja o liberalismo político de Locke, seja o liberalismo econômico de Adam Smith), buscam oferecer uma resposta tanto à

crítica de Hegel quanto à crítica de Marx ao liberalismo[1]. É óbvio que Rawls é devedor de uma tradição liberal mais vasta, que, além de John Locke e Adam Smith, também é dependente ou crítica de Jeremy Bentham, de James e John Stuart Mill, de Henry Sidgwick, etc., enquanto teóricos do liberalismo clássico (séculos XVII, XVIII e XIX). Para o que aqui interessa, entretanto, partir da crítica de Hegel e de Marx ao liberalismo implica fundamentalmente ter como pano de fundo a posição de Locke e de Adam Smith, que são os autores liberais que Hegel e Marx visam com suas críticas. Num outro sentido, vou perseguir a tese de que a retomada de algumas premissas do liberalismo clássico pelos neoliberais ou neoconservadores, a partir da segunda metade do século XX, permite concentrarmo-nos exclusivamente naqueles dois autores, que já delineiam, portanto, algumas premissas centrais do próprio arcabouço neoliberal, ou que a teoria neoliberal retoma.

Então, como vou procurar deixar claro, a reformulação do liberalismo clássico, por parte de John Rawls, no meu entender apenas de maneira secundária visa retomar ou reformular aquela teoria (liberalismo político e econômico clássicos) própria do capitalismo moderno ou de *laissez-faire*; tal como o entendo, Rawls é crítico do neoliberalismo (e do neoconservadorismo), que retoma aquelas teses clássicas. Ora, não faria sentido pura e simplesmente retomar aquelas premissas do liberalismo clássico exatamente porque elas serviram como fundamentação e legitimação do capitalismo de *laissez-faire*, não cabendo mais, no que diz respeito ao estado de bem-estar social, as premissas ligadas à economia *laissez-faire*, por exemplo. Rawls não ignora as mudanças levadas a cabo pela reformulação do Estado e da economia norte-americanos por parte de John Maynard Keynes no governo de Franklin Delano Roosevelt. E não ignora também que o advento da teoria neoliberal – Friedrich Hayek, Milton Friedman e a Escola de Chicago, Robert Nozick, para citar os exemplos que considero mais proeminentes e, portanto, mais desafiadores – põe em xeque o cerne do Estado de bem-estar social (intervenção e controle dos mercados, pleno emprego, seguridade social, ou seja, uma espécie de planificação da sociedade em suas mais diversas esferas) com base em uma *premissa moral*, já elaborada por Locke e corroborada por Adam

[1] Sobre a análise e a resposta de Rawls em relação à crítica de Hegel ao liberalismo, conferir RAWLS, John. *História da filosofia moral*, p. 419-427; sobre a análise e a resposta de Rawls em relação à crítica de Marx ao liberalismo, conferir: RAWLS, John. *Justiça como equidade*: uma reformulação, §45, p. 210-211; e §52, p. 250-253.

Smith: o individualismo, ou seja, a liberdade individual, contra as instituições objetivas – o que apontaria, como o quer Hayek, na própria impossibilidade de uma planificação ainda que mínima da sociedade por parte do Estado (e, nesse contexto, o *laissez-faire* e a *mão invisível* de Adam Smith encontrariam todo o seu sentido). Perseguirei isso no que se segue.

John Locke: Direitos individuais fundamentais, trabalho e estado mínimo

Com efeito, pode-se perceber em John Locke exatamente essa ideia de que os direitos individuais estão na base da estruturação da sociedade política e delimitam tanto o sentido quanto o papel dessa mesma sociedade política. Locke tem em mente, com essa sua afirmação dos direitos individuais fundamentais como fundamento da sociedade (nos seus mais variados aspectos: políticos, sociais, culturais, econômicos, etc.), o absolutismo monárquico e sua ideia de que o governante teria poderes absolutos que lhe davam direito de vida e de morte sobre os súditos, bem como lhe davam o direito de dispor da propriedade desses súditos a seu bel-prazer; e também teria em mente uma crítica à íntima associação entre religião e política, seja no sentido de a religião (e, em particular, a Igreja Católica) fundamentar um poder político de caráter absoluto (à semelhança do caráter absoluto do poder religioso – a Igreja representaria o poder espiritual e o rei, o poder temporal, em íntima conexão), seja no sentido de o poder político defender um credo religioso em particular como a doutrina única e abrangente em termos de sociabilidade (LOCKE, 2005. p. 237-289). Ora, Locke, no *Segundo tratado sobre o governo civil* aponta de maneira direta para o fato de que os indivíduos têm direitos fundamentais que sob hipótese alguma podem ser violados pelas instituições; na verdade, tais direitos, ainda segundo Locke, é que devem estar na base de fundação daquelas instituições, que somente encontrariam legitimidade, portanto, *a partir* deles (isto é, de tais direitos). E é interessante de se perceber que a ênfase lockeana nos direitos fundamentais é tão forte – ou seja, o caráter desses direitos fundamentais é tão forte – que a própria derrubada da sociedade política está justificada à medida que as instituições públicas violarem aqueles direitos (LOCKE, 2005. p. 213-234).

Locke pode, no meu entender, ser considerado o iniciador da modernidade política porque é com ele – e não, por exemplo, com

Thomas Hobbes – que a formulação de uma sociedade fundada no caráter *absoluto e universal* dos direitos individuais fundamentais apontaria para um poder constitucional representativo, no qual a cidadania política daria o tom da própria dinâmica política (e somente ela) e no qual o pluralismo moral daria o tom em termos de sociabilidade[2]. É interessante perceber que o S*egundo tratado sobre o governo civil* é, conforme palavras do próprio Locke, um ensaio sobre *a origem*, *os limites* e *os fins verdadeiros* do governo civil[3]; e é interessante perceber exatamente que esse ensaio comece com a afirmação de que o poder político deve ser diferenciado em relação ao poder de um pai sobre seus filhos, de um patrão sobre seus empregados, de um marido sobre sua esposa e de um senhor sobre seus escravos[4].

Ora, esta confusão entre poder despótico, patriarcal e senhoril *e* poder político é o cerne do absolutismo monárquico. O poder absoluto do rei não se justificava apenas pela fundamentação religiosa desse mesmo poder ou pela nobreza daquele soberano; seu poder absoluto justificava-se, muito mais, pelo fato de o rei representar o pai do povo. *Pai* adquire o sentido de tutor dos filhos, isto é, o termo *pai* denota maioridade, ao passo que o termo *filho* denota menoridade. Nesse sentido, o maior efetivamente tem o direito – e até o dever – de guiar, de orientar e de governar o menor, de protegê-lo de si (do próprio menor) mesmo e dos demais (uma questão que, inclusive em nosso contexto, é incorporada ao direito). O rei seria o *pai do povo* e o seu (do rei) poder como que absoluto em relação ao povo advém desse sentido básico dado à relação entre poder político, que o rei encarna, e povo. A afirmação da menoridade do povo, no fim das contas, é a marca do absolutismo político, seja no sentido de que o povo não sabe conduzir adequadamente o poder político e, portanto, não sabe conduzir-se politicamente (o poder político seria, então, prerrogativa do monarca absoluto, que teria por missão guiar politicamente o povo), seja no sentido de que o povo não sabe conduzir-se adequadamente em termos

[2] A modernidade política, assim entendo, seria marcada pela formulação de direitos individuais fundamentais; de um poder político constitucional e representativo, no qual a cidadania política seria esse elemento fundamental de justificação da sociedade política; e pela afirmação paulatina do pluralismo moral – por isso que, como disse anteriormente, Locke poderia ser entendido como o fundador da modernidade política.

[3] Esse, aliás, é o título do capítulo primeiro do *Segundo tratado sobre o governo civil*, p. 81-82.

[4] Cf.: LOCKE, John. *Segundo tratado sobre o governo civil*, Cap. I ("Ensaio sobre a origem, os limites e os fins verdadeiros do governo civil), p. 82.

morais e de sociabilidade (e, por isso, o poder religioso guiaria esse mesmo povo pelo caminho adequado).

Locke posicionando-se contra essa percepção do absolutismo monárquico, propõe tal distinção entre o poder de um pai sobre seus filhos, de um patrão sobre seus empregados, de um marido sobre sua mulher e de um senhor sobre seus escravos, de um lado, e a relação política e o poder político de outro lado. Ele procura sustentar que o poder político é estabelecido a partir de outra relação e de outros fundamentos que não as relações despóticas, patriarcais e senhoriais e o fundamento do sangue e da religião, próprios do poder político absolutista e das relações de maioridade e de menoridade pressupostas e reproduzidas. Essa forma específica de compreender o poder político pode ser percebida, segundo Locke, naquilo que ele chama de *igualdade natural*, ou seja, no fato de que todos nós usufruímos desde nosso nascimento das mesmas vantagens da natureza e *do uso das mesmas faculdades*[5]. O fato de desfrutarmos de um espaço natural comum e, mais ainda, de possuirmos as mesmas capacidades (tanto físicas quanto intelectuais) nos torna iguais e, portanto, detentores de direitos individuais fundamentais – a universalidade dos direitos individuais fundamentais é decorrente do fato de que possuímos as mesmas capacidades, o que apontaria tanto para a radical igualdade que os indivíduos desfrutam entre si quanto para, consequentemente, relações "[...] sem sujeição ou subordinação [...]"[6] como a única forma válida de relações, de sociabilidade entre esses indivíduos radicalmente iguais em capacidades.

O poder político, para Locke, contrariamente às relações de poder verticais do absolutismo monárquico, fundadas na pressuposição de um poder despótico do rei (que expressaria maioridade) em relação ao povo (que, por sua vez, expressaria menoridade em relação ao rei), é um poder entre indivíduos iguais, que somente pode ser fundado em relações horizontais de uns no que diz respeito aos outros. A própria percepção lockeana de que a sociedade teria começado como fruto de um acordo entre indivíduos, iguais em todos os aspectos relevantes (especificamente iguais em termos de capacidades), apontaria para a perda da naturalidade das instituições e, portanto, para a recusa da

[5] Cf.: LOCKE, John. *Segundo tratado sobre o governo civil*, cap II ("Do estado de natureza"), p. 83.
[6] *Idem, Ibdem*, Cap II, p. 83.

teoria clássica que concebia a sociedade como uma *comunidade natural*[7]. No caso de Locke, não se trata da questão da sociedade *como* comunidade (e muito menos como comunidade natural), mas sim de um individualismo puro e simples (utilizo esse termo sem conotações morais negativas) *a partir do qual* as instituições são fundadas e a partir do qual a ideia de uma comunidade natural é substituída pela ideia de sociedade jurídica e civil daqueles indivíduos egoístas que não teriam nenhum outro vínculo entre si que não aquele do consenso em termos de fundação das instituições políticas, garantidoras daquela sua igualdade natural (um vínculo civil, jurídico, portanto). A sociedade de Locke, fundada nos direitos individuais fundamentais e em sua universalidade, aponta para relações jurídicas como o *médium* da sociabilidade, e não para relações morais – a própria ideia de tolerância encontraria aqui o seu lugar, à medida que a sociedade *não é uma comunidade moral ou natural*; ela é, por assim dizer, uma associação daqueles indivíduos egoístas com vistas à realização da justiça de maneira objetiva, por parte das instituições públicas. O objetivo da sociedade é a proteção, por parte das instituições, daqueles direitos individuais fundamentais: as instituições realizam justiça punitiva de maneira objetiva, imparcial[8]. Sua função, aliás, é exatamente esta de realizar a justiça punitiva de maneira objetiva, imparcial; a realização de uma *vida boa* não lhe cabe, à medida que é algo ligado à privacidade e à liberdade de cada indivíduo e grupo de crença. Isso que poderíamos entender como *bem-estar social* ou *bem público* seria entendido, no caso de Locke, tão-somente como a correta aplicação das leis por parte das instituições públicas, que administrariam e regulariam a sociedade e as relações entre os indivíduos. Bem público seria, portanto, o correlato de *ordem social*, estabelecida pelo Estado em termos de coordenadas jurídicas.

[7] É possível percebermos em Platão, por exemplo, a ideia de que os males sociais advêm da injustiça política – portanto, de que os *deficits* das instituições levariam à injustiça na sociedade (é interessante se perceber que, n'*A República*, a noção de uma *cidade boa*, *justa*, é colocada como fundamental e como condição para a estabilidade da sociabilidade e mesmo da própria formação humana; e de Aristóteles se pode perceber exatamente essa primazia ontológica da sociedade, que é concebida como comunidade natural, em relação aos indivíduos, que somente podem ser entendidos a partir daquela – o homem não poderia ser pensado como um ser isolado dos demais, fora da sociedade, nem poderia subsistir sozinho. Cf., respectivamente: PLATÃO. *A república*, livro IV e V, p. 112-178; ARISTÓTELES. *Política*, livro I, p. 141-142.

[8] Cf.: LOCKE, 2005, Cap VII, p. 133-134.

Isso me parece muito importante, porque o sentido da associação política, no caso de Locke, está *na defesa mútua* dos indivíduos uns em relação os outros (e uns contra os outros) e, por conseguinte, o *sentido do poder político* está *na realização da justiça punitiva*. Locke é muito claro em relação a essa questão sobre o sentido e os fins do poder político:

> *Por poder político, então, eu entendo o direito de fazer leis, aplicando a pena de morte, ou, por via de consequência, qualquer pena menos severa, a fim de regulamentar e de preservar a propriedade, assim como de empregar a força da comunidade para a execução de tais leis e a defesa da república contra as depredações do estrangeiro, tudo isso tendo em vista apenas o bem público*[9].

O poder político, no que diz respeito à esfera interna do próprio país, não partiria de nenhum vínculo moral entre os indivíduos/cidadãos. O poder político se funda exatamente naquele individualismo (novamente pediria que esse termo não fosse avaliado a partir de conotações morais negativas) que se caracteriza pela fruição da vida privada/produtiva como o elemento basilar não somente da vida dos indivíduos, mas também da própria sociedade. A sociedade, nesse aspecto, seria uma associação entre indivíduos privados, voltados ao seu interesse pessoal; e o poder político, que daí se origina, objetiva exclusivamente instituir regras jurídicas (possuindo, inclusive, a força repressiva capaz de tornar efetivas aquelas leis) que regrariam, que regulariam as relações entre esses indivíduos privados, que se relacionam entre si em termos produtivos. Ora, é essa recusa de um vínculo moral entre os indivíduos o ponto de partida para a afirmação, como disse anteriormente, de um poder político como o marcado por uma relação fundamentalmente instrumental entre os indivíduos (no sentido de que eles relacionam-se como sujeitos de direitos – portanto, em igualdade jurídica –, voltados à persecução do seu interesse pessoal), bem como o ponto de partida para a própria colocação da justiça punitiva como esse elemento basilar dos próprios fins do Estado. Nesse segundo ponto, o sentido do Estado, à medida que se concebe a sociedade como uma associação de indivíduos egoístas, voltados para

[9] *Idem, Ibdem*, cap I, p. 82.

sua esfera privada de vida, sem nenhum vínculo moral maior entre eles que a própria associação jurídica em vista da defesa mútua, reside tão-somente em garantir que os direitos fundamentais de cada indivíduo possam ser respeitados por todos os demais, e vice-versa. Entre esses direitos fundamentais, está especificamente o direito de cada um fazer o que quiser com sua propriedade, isto é, com sua vida, sua liberdade e seus bens. O objetivo da esfera político-jurídica estaria, então, na proteção da esfera individual de liberdade contra uma possível intromissão dos demais (inclusive do próprio Estado) – e a igualdade entre todos seria a própria equiparação jurídica de uns em relação aos outros.

> *E, assim, a comunidade social adquire o poder de estabelecer a punição merecida em correspondência a cada infração cometida entre os membros daquela sociedade, que é o poder de fazer leis, assim como também o poder de punir qualquer dano praticado a um de seus membros por qualquer um que a ela não pertença, que é o poder de guerra e de paz; ela o exerce para preservar, na medida do possível, os bens de todos aqueles que fazem parte daquela sociedade*[10].

Como disse, na minha percepção é exatamente essa *juridificação da sociedade*, a partir da recusa de vínculos morais entre os próprios indivíduos, que permanecem fundamentalmente presos à sua privacidade, que justifica não apenas o Estado mínimo, mas também o próprio conservadorismo político, por parte de Locke. Senão vejamos. A ideia de que todos os indivíduos/cidadãos são iguais em termos de capacidades decreta, segundo Locke, a igualdade jurídica dos indivíduos/cidadãos entre si, que, ao mesmo tempo em que não podem ter seus direitos individuais fundamentais violados pelos demais indivíduos e pelas instituições, podem seguir sua vida do jeito que quiserem. Portanto, à medida que o Estado garante a igualdade jurídica entre todos como ponto de partida da sociedade/sociabilidade, o ponto de chegada é definido pelo desenvolvimento das próprias capacidades, dos próprios talentos individuais, *por meio do trabalho*. O ponto de partida é uma questão da sociedade política (em termos de garantia da igualdade jurídica entre todos e para todos); o ponto de chegada é uma

[10] *Idem, Ibdem*, cap VII, p. 133.

questão exclusiva dos próprios indivíduos produtivos, e não do Estado (ponto de chegada que *não pode*, inclusive, *ser determinado pelo próprio Estado*).

Ora, em primeiro lugar, está claro que a igualdade jurídica *não significa*, no caso de Locke, igualdade material entre os indivíduos, de modo que a garantia da igualdade jurídica a todos esses indivíduos, por parte do Estado, não significa que esse mesmo Estado deva equalizá-los em termos materiais (ou seja, em termos de posse de riqueza e de propriedade). Se nos reportarmos ao absolutismo monárquico, perceberemos que a sociedade (aqui, sim, concebida como *natural*) estaria de antemão organizada a partir de uma estrutura hierárquica de poder, concebida como natural (por ser cultural ou religiosamente fundamentada) e, portanto, possuindo um caráter inquestionado (utilizamos o termo *natural* como significando o caráter não histórico, ontológico, da própria organização social, que, exatamente por isso, adquiriria aquele sentido de justificação absoluta e, por conseguinte, apontaria para a impossibilidade da crítica àquela hierarquia e mesmo às instituições). Os indivíduos e as relações que eles estabeleceriam de antemão estariam definidos pelo seu lugar de nascimento: as desigualdades em termos de hierarquia – desigualdades sociais, políticas, culturais e mesmo econômicas – e todas as relações surgidas a partir dessas desigualdades em termos de hierarquia estariam definidas de maneira prévia aos próprios indivíduos, por meio do apelo às tradições culturais e à religião; e essas desigualdades, como disse, definiriam o sentido dos próprios indivíduos, o que eles poderiam esperar vir a ser e as relações que eles travariam, e isso de maneira absoluta. Mas é interessante percebermos que o liberalismo político de Locke (Locke que também é, conforme Hume[11], o fundador do empirismo clássico, que recusa de maneira veemente a fundamentação metafísico-teológica do conhecimento), à medida que recusa essa ideia de uma sociedade hierarquicamente constituída a partir do sangue e da fundamentação religiosa (ou seja, à medida que recusa uma hierarquia social de caráter natural e, portanto, um poder vertical absoluto), partindo, como já disse, de uma igualdade radical entre todos em termos de talentos e capacidades, tem de justificar de outro modo – ou pelo menos tem de justificar política e juridicamente – as desigualdades sociais que surgem (e elas com certeza surgiriam). A estratégia de Locke,

[11] Cf.: HUME, David. *Investigação concernente ao entendimento humano*, Seção I, p. 27.

que já não poderia mais recorrer ao modo de fundamentação metafísico-teológico, já foi indicada por mim: a garantia jurídica da liberdade entre todos, que é o fim do Estado, é o ponto de partida da sociabilidade; o ponto de chegada é definido por parte de cada indivíduo, a partir do desenvolvimento maior ou menor desses indivíduos em termos de atividade produtiva, ou seja, em termos de trabalho. O trabalho, nesse caso, e nossa maior ou menor capacidade em relação a ele, definiriam o paulatino surgimento de desigualdades sociais, que seriam legítimas. À medida que apela ao individualismo egoísta, que se baseia no trabalho, Locke pode substituir a ideia de uma totalidade social fundada na cultura, na religião, em vínculos morais e/ou estamentais, etc. A sociedade é uma associação jurídica – que tem por fundamento o direito positivo, e não a cultura ou a religião – entre indivíduos iguais voltados fundamentalmente à sua esfera privada, que, esta sim, acaba definindo o *status* público e o grau desse *status* possuído pelos indivíduos.

Desse modo, aparece um sentido positivo ao trabalho, que não podia ser percebido no absolutismo político ou Antigo Regime (utilizo ambos os termos com o mesmo sentido). É que, no absolutismo monárquico, o trabalho (em seu sentido negativo de *tripalium*, ou seja, instrumento de tortura) seria algo específico do povo, também compreendido no sentido negativo de *plebe*, ou seja, a classe que vive do trabalho de suas mãos e que, exatamente por isso, possuiria um *status* inferior em relação à nobreza e ao clero – até porque, na compreensão clássica do conceito de trabalho, este, à medida que não contribuiria no desenvolvimento das capacidades intelectuais e mesmo à medida que impediria o exercício da cidadania política por parte da plebe, deformaria aqueles que o realizavam e que dele viviam. E aqueles que viviam do trabalho, por causa da deformação intelectual e mesmo moral que sofriam devido a realizarem por toda a sua vida esse trabalho braçal, estavam condenados ao estrato mais baixo da pirâmide social – seu objetivo estava em sustentar a nobreza e o clero. Ora, é interessante que Locke, na esteira da Reforma protestante levada a cabo principalmente por Martinho Lutero e por Ítalo Calvino[12], aponta para um sentido

[12] Chamo a atenção, no que diz respeito à percepção protestante em relação ao trabalho, para *A ética protestante e o 'espírito' do capitalismo*, de Max Weber. Nessa obra, Weber procura mostrar – e é isto que me parece muito interessante – que *não é mera coincidência* o fato de que a revolução industrial e o desenvolvimento do capitalismo deram-se fundamental e prioritariamente naqueles países de religião e cultura protestante, e não nos países católicos. Cf.: WEBER, Max. *A ética protestante e o 'espírito' do capitalismo*, p. 29.

positivo ao trabalho. É por meio do trabalho que os indivíduos se desenvolvem: o trabalho dignificaria o homem. Como já disse de passagem, à medida que concebe a sociedade como sendo marcada por relações horizontais entre indivíduos livres e iguais, e à medida que o fato de esses indivíduos serem iguais juridicamente falando não implica a igualdade material entre todos eles, Locke tem de justificar o surgimento de desigualdades – e tem de justificar a legitimidade dessas desigualdades. Nesse sentido, tal como eu o entendo, penso que Locke concebe o surgimento das desigualdades sociais ou, em outro sentido, da hierarquia social como sendo resultado do desenvolvimento maior ou menor de nossas capacidades em termos produtivos, de modo que as desigualdades de poder e riqueza, à medida que surgissem, a partir exatamente desse maior ou menor desenvolvimento, por parte de cada indivíduo, de seus talentos, seriam todas legítimas. O trabalho, portanto, conferiria progressiva e paulatinamente *status* social diferenciado aos indivíduos – e esse *status* diferenciado estaria ligado, como disse, ao desenvolvimento desigual das nossas capacidades naturais, desenvolvimento este que se deve fundamentalmente ao nosso empenho ou à falta dele[13]. Ora, à medida que é o trabalho que confere *status* econômico, social e político aos indivíduos, temos que, para Locke, o *homo oeconomicus* define não apenas o sentido, mas também o próprio *status* do *homo politicus*, à medida que a importância do *homo oeconomicus* determina a concomitante proporcional importância do *homo politicus*. Não mais o sangue, o lugar de nascimento ou mesmo a religião determinariam uma estrutura social e diferenças de hierarquia rígidas, mas sim o trabalho seria esse fator de distinção social e de hierarquia entre os indivíduos – inclusive em termos de hierarquia de poder.

O trabalho seria o fundamento da propriedade privada, porque é por meio dele que os indivíduos, desenvolvendo suas capacidades, têm condições de, aos poucos, alcançarem patamares maiores de poder e acumular capital. E a defesa da propriedade privada, assim, já, com Locke, se coloca como o objetivo fundamental do Estado. Ora, é dessa especificidade em termos de entendimento da sociedade/sociabilidade, por parte de Locke, que o próprio sentido de um *Estado mínimo*, centrado basicamente na realização da justiça punitiva, fica explicitado. É que, no caso de Locke, a igualdade jurídica entre todos seria o ponto de

[13] Cf.: LOCKE, 2005, cap V, p. 97-112.

partida tanto da esfera política, em particular, quanto da esfera produtiva/social de uma maneira geral. O ponto de chegada seria responsabilidade de cada indivíduo, em termos de desenvolvimento dos seus talentos na esfera produtiva, por meio do trabalho. Nesse sentido, as desigualdades que surgem por meio do desenvolvimento dessas capacidades individuais a partir do trabalho conferem plena legitimidade para a hierarquia de poder e para as desigualdades daí advenientes. As desigualdades sociais, portanto, não se devem a *deficits* nas instituições públicas, mas sim ao trabalho realizado por parte de cada indivíduo e ao modo como cada indivíduo desenvolveu suas capacidades em termos de trabalho. À medida que tais desigualdades não são causadas pelas instituições, mas estiveram fundadas no desenvolvimento das capacidades individuais dos próprios indivíduos, elas são todas legítimas e, portanto, sobre elas o poder do Estado não incide, no sentido de corrigi-las. Basta ao Estado que realize justiça punitiva, garantindo o respeito e o cumprimento dos pactos e dos contratos. O Estado deveria – e essa seria sua única função – proteger a propriedade privada dos indivíduos/cidadãos, adquirida por meio do trabalho de suas mãos. Aqui aparece o sentido conservador do pensamento de Locke: o Estado não tem condições nem legitimidade para corrigir as desigualdades sociais pelo fato de que elas não surgiram por causa de *deficits* nas instituições; as desigualdades sociais, à medida que o Estado garantiu a igualdade jurídica entre todos os indivíduos, se devem exclusivamente ao maior ou menor desenvolvimento dos talentos de cada um e, por isso, são todas legítimas. Basta, então, um Estado restrito à função de realização da justiça punitiva, ou seja, à realização do direito privado[14].

Adam Smith e o princípio do laissez-faire: da sociedade econômica para a sociedade política

Ora, na minha percepção, o liberalismo político clássico de Locke é o complemento ou mesmo o fundamento filosófico e político do liberalismo econômico clássico de Adam Smith, exatamente porque oferece os fundamentos normativos de uma sociedade econômica que, em sua centralidade em relação ao Estado, à esfera política, determina não apenas o próprio sentido dessa esfera política como secundária e subordinada em relação à esfera econômica, bem como do Estado como Estado mínimo, cuja função básica consiste em realizar justiça punitiva,

[14] Cf.: HABERMAS, Jürgen. *Direito e democracia*: entre facticidade e validade (vol. I), p. 109.

mas também, e até fundamentalmente, (sociedade econômica) que tem condições de estabilizar-se e, ao fazer isso, de estabilizar a sociedade como um todo em termos de distribuição da riqueza socialmente produzida e mesmo em termos de sociabilidade (capitalismo de *laissez-faire*). A esfera econômica, e não a esfera política, adquiriria centralidade – e a partir dela a sociedade como um todo se desenvolveria. É nesse sentido que a economia política clássica, tal qual concebida por Adam Smith, consolida termos como *Estado mínimo* ou *Estado de laissez-faire*, *mão invisível*, *competição*, *individualismo*, *egoísmo*, etc.

Tentarei, no que diz respeito a Adam Smith, relacionar de maneira dinâmica esses conceitos nesta seção.

É interessante perceber, em relação a Adam Smith, que a esfera das atividades produtivas, isto é, o mercado, possui uma dinâmica interna que por si só tem condições de estabilizar-se e, à medida que essa esfera do mercado, da produção, é central para a sociedade como um todo, também tem condições de estabilizar a sociedade como um todo, de satisfazer as aspirações individuais e sociais como um todo. E o modelo de sociedade concebido por Adam Smith é o mesmo modelo formulado por Locke. Trata-se fundamentalmente de uma associação jurídica entre indivíduos livres e iguais, que se relacionam de maneira egoísta e competitiva no mercado, buscando cada um seus objetivos pessoais e, indiretamente, contribuindo para a satisfação das necessidades dos demais. O individualismo egoísta (também neste caso pediria que não o compreendêssemos como possuindo um sentido moral negativo, pejorativo) e, consequentemente, a competição entre esses indivíduos egoístas daria o tom da esfera produtiva da sociedade, à medida que cada um, conforme já salientado, buscaria, antes de tudo, seus interesses pessoais, a partir de suas relações com os demais; mas, interessantemente, essa competição entre indivíduos egoístas, que para Adam Smith dá a tônica em termos de relações produtivas, em termos de mercado, conduz tanto à satisfação das necessidades de cada indivíduo quanto à estabilidade social[15]. Ora, como é possível, em uma situação de competição entre indivíduos egoístas, a satisfação das necessidades de cada um desses indivíduos e a estabilidade social?

[15] Cf.: SMITH, Adam. *A riqueza das nações*, v. I, cap II ("Do princípio que dá origem à divisão do trabalho"), p. 94-95, sobre a questão do egoísmo, da busca do bem-estar pessoal e, consequentemente, da consecução do bem-estar social.

Em primeiro, lugar é interessante perceber que a esfera do mercado, em Adam Smith, assim como a esfera do trabalho, em John Locke, foi concebida como uma esfera marcada eminentemente por relações instrumentais, técnicas, destituídas de uma ligação moral entre os indivíduos. Como salientado de passagem citada, nesta seção, para Adam Smith a sociedade se caracteriza como uma associação jurídica entre indivíduos livres e iguais, mas profundamente egoístas, cujo objetivo é a defesa mútua, a partir de uma instituição central que teria por função realizar justiça punitiva de maneira objetiva e cujo objetivo (dessa associação jurídica e mesmo da fundação do Estado) é a instauração de uma esfera de relações produtivas – no caso, o mercado – a partir da qual se daria o desenvolvimento da sociedade e a satisfação das necessidades de cada indivíduo, por meio do trabalho. Então, fica claro, e Adam Smith insiste nisso, que o sentido do Estado moderno está efetivamente na defesa da propriedade: a associação político-jurídica moderna encontra seu sentido exatamente em torno da defesa e da promoção da propriedade[16], à medida que é a partir desta e do trabalho que se desenvolve em torno dela que a riqueza e a estabilidade social são produzidas. Isso é muito importante de se perceber, porque, à medida que o sentido do Estado moderno está na defesa e na promoção da propriedade (vida, liberdade e bens), aparece de maneira clara a própria ênfase na juridificação da sociedade e mesmo nos próprios valores do trabalho, valores estes ligados ao caráter primordial da esfera privada da vida em relação à esfera pública (esfera privada significando tanto a esfera do gozo pessoal quanto do trabalho). Assim, uma associação jurídica entre indivíduos livres e iguais entre si, que também são profundamente egoístas, basta para garantir, no caso de Adam Smith, a estabilidade da sociedade exatamente pelo fato de que é na esfera privada e a partir dela que os indivíduos, por meio do trabalho, desenvolvem-se, realizam seus interesses e, por conseguinte, contribuem no desenvolvimento da sociedade de uma maneira geral. Nesse caso, o poder público seria uma espécie de apêndice do poder econômico privado, pelo fato de estar subordinado a esse poder privado: o objetivo do poder político-estatal e da associação jurídica é proteger a esfera privada de vida, ou seja, a esfera produtiva, o mercado, à medida que essa esfera é condição da própria sociedade política.

[16] Para A. Smith foi o surgimento da propriedade que fez necessário a criação do governo civil. Cf.: SMITH, Adam. *A riqueza das nações*, v. II, segunda parte ("Das despesas com a justiça"), p. 315.

Então, e isso é muito interessante, a esfera do mercado é colocada como uma esfera privada, e não como uma esfera pública. Isso significa duas coisas: (a) tanto que a esfera privada determina o sentido da esfera pública, à medida que aquela é mais primordial que esta; (b) quanto que a esfera privada possui uma autonomia em relação à esfera pública, que lhe permite total independência em relação à qualquer intervenção social, política, de modo que a própria ação do Estado em relação à esfera privada, em relação ao mercado, fica interrompida. O Estado, à medida que sua tarefa é proteger a propriedade, deve retirar-se, de acordo com Adam Smith, da própria intervenção na propriedade. Ora, a ênfase em um Estado marcado pela justiça punitiva em nível interno aponta, de maneira clara, para essa percepção de que o Estado está bem estruturado quanto menos papel tiver em relação ao livre-mercado. Basta ao Estado que garanta, como sua função fundamental e básica, a realização da justiça punitiva a todos os indivíduos e para todos eles; o restante é responsabilidade dos próprios indivíduos e do trabalho que realizam no mercado[17]. Ou seja, em relação a este último ponto, a questão-chave, no caso de Adam Smith, está em que é na esfera produtiva que os desafios, que as necessidades em termos de produção da riqueza social são satisfeitos, e não por meio da esfera política; seria aquela, e não esta, quem estabeleceria a dinâmica social, o que significa que é a esfera econômica o centro da sociedade, a base a partir da qual a sociedade se origina e se desenvolve (até porque é nela e por meio dela que se dá a produção da riqueza social – e é isto que lhe confere primazia em relação à esfera política). Mas, mais importante ainda, é suficiente que o Estado tenha por função básica realizar justiça punitiva pelo fato de que são os próprios indivíduos, iguais entre si em capacidades e, por conseguinte, iguais também em termos de *status* jurídico, que, no uso de sua liberdade criativa, conseguem conquistar, por meio do trabalho que realizam e do desenvolvimento de suas capacidades, as posições sociais e políticas. O argumento de Adam Smith é interessante: são os próprios indivíduos os únicos responsáveis por sua situação social – e a competição entre esses indivíduos egoístas no mercado teria inclusive o mérito de "forçar" cada indivíduo a desenvolver cada vez mais suas capacidades, para que não seja suplantado pelos demais (portanto, a

[17] A. Smith diz, no primeiro volume de *A riqueza das nações*, parte II ("As desigualdades que resultam da política da Europa"), p. 263 e seguintes, que a falta disso que ele chama de "liberdade total" do mercado é a causa da grande desigualdade da Europa de então. Os Estados europeus de seu (A. Smith) tempo, segundo ele, põem "[...] obstáculos à livre circulação do trabalho e do capital".

competição teria esse aspecto positivo, de modo que ela levaria não ao surgimento de processos de exclusão e de marginalização, mas sim à necessidade, por parte de cada indivíduo, de desenvolver ao máximo suas próprias capacidades). Ou seja, e era isso que eu queria significar, as distinções sociais surgem por obra dos próprios indivíduos, e não por obra do Estado: consequentemente, são todas legítimas.

É então que a ideia de um Estado mínimo, intrinsecamente ligada às ideias de economia de *laissez-faire* e de *mão invisível*, encontram todo o seu sentido. Numa situação de competição entre indivíduos egoístas, na qual esses mesmos indivíduos se veem obrigados a desenvolver suas capacidades em um grau máximo, cada um deles consegue satisfazer suas necessidades pessoais e, por meio do desenvolvimento dessas suas capacidades, acabam contribuindo, direta ou indiretamente, para a satisfação das necessidades dos demais, estabilizando as expectativas sociais como um todo. Haveria um princípio, uma lógica subjacente à produção econômica – marcada pela competição entre indivíduos egoístas na esfera da produção, do mercado – que levaria à estabilidade da sociedade exatamente a partir daquela competição entre indivíduos egoístas, que, numa situação como esta, não teriam alternativa que não o desenvolvimento cada vez mais acentuado de suas capacidades: trata-se da ideia de *mão invisível* – a consolidação do mercado como o elemento por excelência não apenas da produção da riqueza social, mas também, por causa disso, da própria estabilidade da sociedade como um todo[18].

Nesse aspecto, o mercado estabiliza-se exatamente por meio da competição entre aqueles indivíduos egoístas, obrigados a desenvolver seus talentos em grau máximo para não serem colocados às margens do próprio mercado. Ora, à medida que é a base da sociedade, o mercado, ao estabilizar-se, a partir de sua dinâmica interna, estabiliza a própria satisfação das necessidades sociais, levando ao bem-estar geral. Por isso mesmo, o sentido fundamental da sociedade política consiste exatamente em ser uma associação jurídica; e o sentido do Estado consiste em realizar a justiça punitiva. Basta que o Estado garanta, por meio do exercício dessa justiça punitiva, a efetiva garantia da igualdade jurídica entre todos como ponto de partida não apenas da sociedade, mas também das relações de produção: como em Locke, o ponto de chegada será determinado pelos próprios indivíduos, a partir do maior ou menor

[18] Cf.: SMITH, Adam. *A riqueza das nações*, v. I, 1999, livro III, cap II ("Do desencorajamento da agricultura no antigo Estado da Europa após a queda do Império Romano), p. 668.

grau de desenvolvimento de suas capacidades no mercado de trabalho. O Estado, interessantemente, adquire esse papel restrito de realizar justiça punitiva tanto pelo fato de o mercado estabilizar-se e estabilizar, por conseguinte, a sociedade como um todo (o mercado, nesse caso, seria a esfera básica da sociedade, o centro da dinâmica social e política) quanto pelo fato de as desigualdades sociais que surgirem não serem ocasionadas por *deficits* nas instituições, mas sim pelo maior ou menor grau de desenvolvimento das capacidades dos próprios indivíduos. As desigualdades sociais, políticas e econômicas resultam da própria capacidade dos indivíduos nas relações produtivas que eles estabelecem entre si; e, à medida que o Estado garante a igualdade jurídica entre todos, todas as desigualdades que surgirem a partir daí serão consideradas legítimas, porque o Estado, por meio da justiça punitiva, teria imunizado as relações de produção de desigualdades injustificadas em termos de posse de poder (os indivíduos relacionam-se entre si como livres e iguais juridicamente falando – essa garantia jurídica de sua liberdade e igualdade lhes capacitaria a, por si próprios, desenvolverem suas capacidades e satisfazerem suas necessidades privadas).

Assim, há uma ideia importante que aparece com a economia de *laissez-faire* e seu correlato Estado de *laissez-faire*: a ideia, já exposta em linhas gerais, de que a esfera do mercado é uma esfera marcada por relações eminentemente jurídicas como regulando a competição entre os indivíduos livres e iguais (livres e iguais em termos jurídicos), de modo que a esfera do mercado seria um espaço que não está fundado em relações morais, mas sim pelo direito privado[19]. Ora, isso é fundamental para compreendermos porque, numa situação de garantia jurídica da liberdade e da igualdade entre todos os indivíduos, qualquer desigualdade que surgir será responsabilidade única e exclusiva dos próprios indivíduos, o que implica não apenas em sua legitimidade, mas, como consequência na ilegitimidade de qualquer tentativa de correção daquelas desigualdades. Aquele que caiu na dinâmica da competição econômica é, no fim das contas, o único culpado do seu fracasso. Pode-se, nesse sentido, olhar para ele sem nenhum remorso exatamente porque, diferentemente daqueles que foram bem-sucedidos, ele não soube ou não quis desenvolver efetivamente seus próprios talentos: o individualismo puro e simples apontaria para o direito privado como

[19] Cf.: ROSANVALLON, Pierre. *O liberalismo econômico*: História da Ideia de mercado, p. 08; HABERMAS, Jürgen. *Direito e democracia*: entre facticidade e validade (V. II), p. 144.

uma forma de regulação técnica da sociedade e do próprio mercado que excluiria qualquer sentido mais amplo tanto para a esfera política e para a sociabilidade de uma maneira geral quanto para a própria esfera produtiva em particular – desigualdades sociais, políticas e econômicas seriam legítimas porque expressam a própria capacidade ou incapacidade dos indivíduos, e não *deficits* nas estruturas políticas, sociais ou mesmo econômicas; não são desigualdades passíveis de ajuizamento moral nem (tais desigualdades) afetariam nossa compaixão moral.

A reformulação do liberalismo clássico por John Rawls

E é interessante perceber, nesse sentido, que Rawls considere que a ideia fundamental, da qual a sua teoria da justiça como equidade parte, é a de que a sociedade é um sistema equitativo de cooperação social ao longo do tempo entre pessoas livres e iguais para benefício recíproco[20]. De imediato se pode perceber algumas afirmações importantes: (a) a sociedade é um sistema cooperativo; (b) entre pessoas livres e iguais; (c) equitativo e voltado ao benefício recíproco; (d) que se mantém ao longo do tempo. Contra a visão liberal clássica de uma sociedade marcada pelo individualismo competitivo em torno da acumulação da propriedade, que, no fim das contas, acaba perdendo o próprio sentido de coletividade, a partir da absolutização do atomismo social (como o haviam percebido tanto Hegel quanto Marx), e contra a legitimação da validade das desigualdades sociais exatamente pela compreensão de que a esfera econômica é uma esfera privada, não política, Rawls enfatiza a ideia de que a cooperação social não é apenas a base a partir da qual a riqueza social é produzida, mas também a base da própria estabilidade da sociedade, à medida que uma sociedade entendida, conforme o liberalismo clássico, como uma associação de indivíduos egoístas em permanente competição entre si conduziria à instabilidade social, à marginalização social. "O bem-estar de cada um", diz Rawls, "depende de um esquema de cooperação social sem o qual ninguém teria uma vida satisfatória" (2002a. p.110).

Além disso, pelo fato de todos contribuírem na produção da riqueza social, segue-se que esta deve ser repartida equitativamente. A ideia de sociedade como sistema de cooperação social entre pessoas livres e

[20] Cf.: RAWLS, John. *Uma teoria da justiça*, § 14, p. 90; RAWLS, John. *Justiça e democracia*, p. 256-257.

iguais ao longo do tempo para o benefício recíproco, em suma, aponta para a recusa de que o individualismo puro e simples, que tem por base a competição de uns com os outros, possibilite efetivamente o desenvolvimento das capacidades de cada indivíduo, bem como a satisfação de suas (de todos os indivíduos) necessidades particulares. Somente por meio da cooperação é que se consegue tanto o desenvolvimento pessoal quanto a justiça social. Diz Rawls:

> *As pessoas precisam umas das outras, pois é apenas com a cooperação ativa dos outros que o talento de cada um em particular pode ser realizado e, por conseguinte, em grande parte, com os esforços de todos. Somente nas atividades da união social o indivíduo pode ser completo*[21].

Segundo Rawls, a sociedade "[...] é uma união social de uniões sociais" (2002b, p.375). Para entender isso, podemos considerar o exemplo dado por nosso pensador. Trata-se de um grupo de músicos talentosos, que poderiam, cada um deles, tocar todos os instrumentos da orquestra, mas que, em virtude das limitações humanas, decidem, cada um deles, se especializar em um instrumento específico. Se, individualmente, esses músicos não conseguem desenvolver todas as suas capacidades, em grupo eles alcançam o desenvolvimento de todas elas (RAWLS, 2002b, p.376-377). A ideia de cooperação, por isso mesmo, implica a recusa de um individualismo puro e simples, tal como no liberalismo clássico. É certo que o individualismo não adquire em Rawls um sentido negativo, à medida que aponta para a liberdade crítica e criativa de cada um dos indivíduos, que têm o direito a seguir sua vida do jeito que quiserem. Mas é interessante que, como condição para uma sociedade democrática, as virtudes cívicas são absolutamente necessárias para a própria estabilidade dessa sociedade. Ora, a cooperação social pressupõe necessariamente o respeito e o benefício mútuos[22], que somente podem ser conquistados por meio do esforço mútuo[23]. Nesse sentido, uma sociedade democrática bem-ordenada aponta para objetivos comuns, coletivos, especialmente o objetivo de realizar justiça política mútua. Diz Rawls: "em uma sociedade bem-

[21] RAWLS, John. *O liberalismo político*, p. 377.
[22] RAWLS, John. *Justiça e democracia*, p. 213; RAWLS, John. *Justiça e democracia*, p. 156; RAWLS, John. *Uma teoria da justiça*, § 76, p. 557.
[23] RAWLS, John. *O liberalismo político*, p. 337.

ordenada [...], os cidadãos têm fins últimos em comum, entre os quais está o de propiciar justiça política uns aos outros" (2002b, p.257). Em *Justiça e democracia*, Rawls diz:

> *[...] uma sociedade bem-ordenada (assim como definida pela justiça como equidade) não é, portanto, uma 'sociedade privada', pois, nela, os cidadãos têm fins últimos em comum. Se é verdade que eles não abraçam as mesmas doutrinas abrangentes, em compensação adotam a mesma concepção política de justiça. Isso quer dizer que compartilham um fim político, inteiramente fundamental e prioritário, que consiste em defender as instituições justas e em proporcionar justiça de acordo com elas, sem contar os numerosos outros fins que devem igualmente partilhar e efetivar através de sua organização política. Ademais, a justiça política pode fazer parte dos objetivos mais fundamentais dos cidadãos, graças aos quais eles definem o tipo de pessoa que querem ser*[24]*.*

Também chamaria a atenção para a percepção – conforme explicitada pela ideia de sociedade concebida como sistema equitativo de cooperação social ao longo do tempo entre pessoas livres e iguais – de que a produção da riqueza social se dá de modo cooperativo, e não a partir de um individualismo puro e simples, fundado na competição mútua. Exatamente por isso, a produção social deveria ser distribuída equitativamente, embora de maneira desigual (trato disso mais adiante). No liberalismo clássico, a questão da distribuição da riqueza e da produção social de antemão ficava deslegitimada pelo fato de a produção dessa mesma riqueza ser uma atividade privada, sob responsabilidade de cada indivíduo em particular.

Contrariamente, em segundo lugar (à medida que o primeiro ponto que tratei, acerca de Rawls, foi o da ideia de sociedade como sistema equitativo de cooperação entre pessoas livres e iguais ao longo do tempo para benefício recíproco), à afirmação de um Estado mínimo, restrito à aplicação do direito privado e concentrando em si o aparato repressivo, a ênfase, por parte de Rawls, em que o objeto central da justiça política seria isso que ele chama de *estrutura básica da sociedade* aponta para a

[24] RAWLS, John. *Justiça e democracia*, p. 321. Cf., ainda: RAWLS, John. *O liberalismo político*, p. 250-251.

sua (de Rawls) compreensão de que as instituições políticas, econômicas e sociais influem no que diz respeito à determinação tanto da formação da personalidade quanto da instauração da sociabilidade. Se estiverem desorganizadas, podem levar a grandes *deficits* sociais e de formação. Diz Rawls:

A estrutura básica da sociedade é a maneira como as principais instituições políticas e sociais da sociedade interagem formando um sistema de cooperação social e a maneira como distribuem direitos e deveres básicos e determinam a divisão das vantagens provenientes da cooperação social. A constituição política com um judiciário independente, as formas legalmente reconhecidas de propriedade e a estrutura da economia (na forma, por exemplo, de um sistema de mercados competitivos com propriedade privada dos meios de produção), bem como, de certa forma, a família, tudo isso faz parte da estrutura básica. A estrutura básica é o contexto social de fundo dentro do qual as atividades de associações e de indivíduos ocorrem. Uma estrutura básica justa garante o que denominamos de *justiça de fundo*[25].

A estrutura básica da sociedade, constituída pelas principais instituições políticas, econômicas e sociais é o objeto básico da justiça política exatamente pelo fato de que ela define as regras a partir das quais a sociabilidade se dá. Nesse sentido, levantaríamos a ideia de que, no caso de Rawls, nós, em contraposição ao liberalismo clássico, temos a primazia do direito público em relação ao direito privado, no sentido de que se reconhece claramente que há uma estrutura a partir da qual as relações sociais se regulam – e de que essa estrutura influi poderosamente no que diz respeito às desigualdades sociais. Sendo assim, essa estrutura deve ser o objeto básico da justiça política porque é a partir dela que a cooperação social se dá e as desigualdades progressivamente são instauradas. Nesse aspecto, nós temos a percepção de que há uma área social, política e econômica que não pode ser concebida como esfera meramente privada, exatamente pelos efeitos macro estruturais que ela leva a efeito. Ora, as instituições econômicas, políticas e sociais, tais quais citadas, influem de maneira poderosa e decisiva no desenvolvimento, na evolução da sociedade – e as

[25] RAWLS, John. *Justiça como equidade*: Uma reformulação, §04, pp. 13-14. Conferir, ainda: RAWLS, John. *Justiça e democracia*, p. 03; RAWLS, John. *Justiça e democracia*, p. 203; RAWLS, John. *O liberalismo político*, p. 309; RAWLS, John. *Uma teoria da justiça*, §02, p. 08; RAWLS, John. *Justiça como equidade*: Uma reformulação, §12, p. 56.

desigualdades e hierarquias sociais surgem *por causa delas*, *por meio delas*, de modo que, inclusive pressupondo o caráter político e social dos direitos individuais, tais instituições e a regulação conveniente dessas instituições devem fazer parte da agenda política democrática.

Ora, pode-se perceber – e vou trazer outras ideias de Rawls para comprovar isso – que a questão-chave da justiça política, em uma sociedade democrática, é a colocação do direito público como a base a partir da qual o direito privado encontra seu sentido. Especificamente à questão do mercado e da propriedade, o direito público adquire primazia. Vou perseguir isso no que se segue, a partir da apresentação dos dois princípios da justiça como equidade. Ora, a primazia do direito público como objeto básico da justiça política aponta tanto para a regulação daquelas instituições políticas e econômicas que produzem desigualdades e hierarquias sociais quanto para o caráter fundamental, em termos de organização democrática da sociedade, daqueles direitos sociais de cidadania sem os quais os próprios direitos individuais não teriam o mínimo sentido. Em primeiro lugar, no que se segue, apresento os dois princípios de justiça, cujo objetivo seria a regulação das instituições políticas e econômicas, propostos por Rawls; depois, esboço algumas consequências desses mesmos princípios de justiça política e econômica.

> *(a) Cada pessoa tem o mesmo direito irrevogável a um esquema plenamente adequado de direitos e de liberdades básicas iguais, que seja compatível com o mesmo esquema de liberdades para todos; e (b) as desigualdades sociais e econômicas devem satisfazer duas condições: primeiro, devem estar vinculadas a cargos e a posições acessíveis a todos, em condições de igualdade equitativa de oportunidades, e, segundo, têm de beneficiar ao máximo os membros menos favorecidos da sociedade (o princípio de diferença)*[26].

[26] RAWLS, John. *Justiça como equidade*: Uma reformulação, §13, p. 60. Conferir, ainda: RAWS, John. *Uma teoria da justiça*, §39, p. 275; RAWLS, John. *Justiça e democracia*, p. 144-145; RAWLS, John. *Justiça e democracia*, p. 207-208; RAWLS, John. *O liberalismo político*, p. 47; RAWLS, John. *O liberalismo político*, p. 324; RAWLS, John. *O liberalismo político*, p. 345; RAWLS, John. *Uma teoria da justiça*, §14, p. 64; RAWLS, John. *Uma teoria da justiça*, §46, p. 333-334.

Como se pode perceber, tais princípios, como diretores da sociedade política em suas tarefas de planificação da sociedade, e diretores da realização pública da justiça política, apontam para a justiça distributiva como questão central da sociedade política. A ideia de sociedade como sistema equitativo de cooperação social entre pessoas livres e iguais ao longo do tempo para benefício recíproco aliada à ideia de estrutura básica representando a poderosa influência das instituições políticas e econômicas na consolidação de desigualdades e hierarquias sociais exigem que certos bens sociais sejam repartidos de forma *exatamente igual* – nesse caso, direitos e liberdades básicas individuais; exigem também que as oportunidades de aceder aos cargos públicos sejam equitativas a todos e conquistadas unicamente pelo mérito pessoal; e exigem que as desigualdades sociais e econômicas, para serem legítimas, devam produzir melhora nas condições de vida dos menos favorecidos.

Portanto, e é isso que gostaria de enfatizar no tocante a Rawls, há uma intrínseca ligação entre a organização das instituições políticas, econômicas e sociais, definidas a começar da estrutura básica, e a realização dos direitos individuais e dos direitos sociais de cidadania. E há uma ligação intrínseca exatamente pelo fato de ser uma organização adequada das instituições políticas, econômicas e sociais a que leva essas mesmas instituições a não serem dominadas, determinadas pela arbitrariedade da competição capitalista pura e simples, nem pela força política dos grupos economicamente hegemônicos, ou mesmo culturalmente hegemônicos. A ênfase em termos de justiça distributiva, como questão-chave da sociedade política no que diz respeito à realização da justiça política, aponta de maneira clara para a percepção de que os direitos individuais fundamentais e, no mesmo sentido, uma igualdade material equitativa, que minimizariam as relações de poder arbitrárias entre os indivíduos e grupos (relações de poder arbitrárias que levariam ao aumento das desigualdades sociais, políticas e econômicas injustificadas), e que possibilitariam a esses mesmos indivíduos e grupos seu florescimento e desenvolvimento, aponta para o fato de que os direitos individuais fundamentais somente se tornam possíveis à medida que os direitos sociais de cidadania sejam efetivos e universalizados a todos – sem estes, aqueles se tornam privilégios restritos aos grupos hegemônicos, na exata medida em que desigualdades acentuadas em termos de poder econômico e político tendem não apenas a solapar um mínimo de igualdade material entre

todos necessária para a efetividade das liberdades políticas, mas também, e até fundamentalmente, tais desigualdades acentuadas tendem a criar um espaço de marginalização que submete os excluídos a um processo de degradação física e moral verdadeiramente destruidor de si mesmos e corrosivo em termos de sociabilidade.

Loïc Wacquant, no que diz respeito à relação entre direitos sociais de cidadania e direitos individuais fundamentais, em seu livro *As prisões da miséria*, pesquisou os resultados de duas décadas de políticas neoliberais nos EUA e na Inglaterra, a começar, respectivamente, dos governos Reagan e Thatcher. Sua conclusão foi a de que a retirada desses Estados em termos de investimento nas áreas sociais foi inversamente proporcional ao aumento dos investimentos no aparato repressivo. Ou seja, quanto mais o Estado de bem-estar deixou de investir em inclusão social, mais teve de investir, e num grau concomitante ao da retirada dos investimentos sociais, em termos de repressão – o aparato punitivo cresce concomitantemente à diminuição dos investimentos públicos ligados à inclusão social[27]. E não é mero acaso que a emergência de um discurso neoconservador, para o qual a pobreza é um problema dos próprios pobres, surge e se consolida exatamente em uma situação de ascensão do neoliberalismo[28].

É nesse contexto de intrínseca ligação e complementaridade entre direitos sociais de cidadania e direitos individuais fundamentais – e, portanto, de centralidade da justiça política – que podemos entender algumas consequências da afirmação dos dois princípios de justiça, especificamente (a) a formulação de um mínimo social, que deveria ser realizado a todos (em especial aos menos favorecidos); (b) a especificação daquilo que significa *bens sociais primários* como base das reivindicações sociais em termos de justiça política; (c) o caráter social da propriedade; (d) a ideia de que os talentos naturais somente à medida que propiciam desenvolvimento social podem levar a desigualdades

[27] WACQUANT, Loïc. *As prisões da miséria*, p. 77-118.

[28] Em relação ao Brasil, é interessante perceber que o governo FHC, à medida que lançou o Brasil no redemoinho da globalização econômica, implicou paulatinamente a idolatria da segurança pública, pelo fato de que a população não tem mais segurança e, por isso mesmo, exige maior investimento público em termos de aparato repressivo – sentimento de insegurança e apelo ao aumento do aparato repressivo que denotam exatamente a emergência do neoconservadorismo e mesmo do xenofobismo social (os culpados da violência social são os pobres, sem uma contextualização mais clara dessa questão em termos de organização social, política e econômica).

sociais (os talentos naturais seriam um bem comum) ou de que eles não poderiam fundamentar desigualdades injustificadas; (e) bem como a ideia de cuidado para com as gerações futuras. Em breves palavras, explicarei cada um desses pontos. Os cidadãos democráticos têm direito a um conjunto de direitos sociais de cidadania, abaixo do qual ninguém poderia cair. Sem esse conjunto mínimo de direitos sociais e mesmo de direitos individuais, eles não conseguem desenvolver-se adequadamente nem possuir uma igualdade material em relação aos demais[29]. Esse conjunto mínimo de direitos sociais de cidadania fica especificado a partir da noção rawlsiana de bens sociais primários, que consistem em direitos e liberdades básicas, oportunidades equitativas de assumir cargos públicos, renda e riqueza, autoestima e autorrespeito, educação, assistência médica e seguridade social[30]. A propriedade, nesse contexto, possui um caráter social no sentido de que a produção da riqueza, à medida que é realizada cooperativamente por todos, deve ser repartida equitativamente, Rawls não defende um igualitarismo radical (tal qual tematizado pelo comunismo, por exemplo), mas também não defende aquelas desigualdades iníquas do liberalismo clássico, que concebia a propriedade como uma esfera fundamentalmente privada, não política. Aparece em Rawls, portanto, o caráter em grande medida político e social da propriedade dos meios de produção (aquele ao qual me refiro quando falo em *propriedade*) e da própria produção da riqueza, o que aponta para o fato de o controle e a regulação públicos deles, em certo nível, serem absolutamente necessários[31]. Contra o liberalismo clássico, Rawls afirma que os talentos naturais não podem dar origem a desigualdades sociais acentuadas ou injustificadas. Devemos colher os frutos do desenvolvimento de nossos talentos à medida que eles contribuem para a melhoria da situação dos demais cidadãos[32]. Note-se, em relação a isso, que, no liberalismo clássico, o desenvolvimento dos

[29] Cf.: RAWLS, John. *O liberalismo político*, p. 213.
[30] Cf.: RAWLS, John. *Justiça como equidade*: Uma reformulação, §17, p. 82-83; RAWLS, John. *Justiça e democracia*, p. 62-63; RAWLS, John. *Justiça e democracia,* p. 166-167; RAWLS, John. *Justiça e democracia*, p. 302; RAWLS, John. *O Direito dos povos*, p. 18; RAWLS, John. *O liberalismo político*, p. 121; RAWLS, John. *O liberalismo político*, p. 228; RAWLS, John. *O liberalismo político*, p. 363; RAWLS, John. *Uma teoria da justiça*, §11, p. 66; RAWLS, John. *Uma teoria da justiça*, §15, p. 98-99; RAWLS, John. *Justiça como equidade*: Uma reformulação, §51, pp. 240-241.
[31] RAWLS, John. *Justiça como equidade*: Uma reformulação, § 32, p. 161; RAWLS, John. *O liberalismo político*, p. 325.
[32] Cf.: RAWLS, John. *Justiça como equidade*: Uma reformulação, §21, p. 106; RAWLS, John. *Uma teoria da justiça*, §17, p. 111.

talentos naturais não apenas leva a desigualdades em termos econômicos e políticos, senão que as justifica e, consequentemente, legitima a instauração de um Estado mínimo, restrito às funções de realização da justiça punitiva, bem como deslegitima a realização da justiça distributiva por parte desse mesmo Estado. No caso de Rawls, o maior ou menor desenvolvimento de nossos talentos não pode significar que os mais bem dotados tenham direito a um esquema cooperativo no qual as vantagens de sua posição sejam priorizadas em relação aos menos favorecidos. Por fim, em relação à delimitação a que me propus acima, deve-se, segundo Rawls, buscar preservar conquistas públicas em termos de cultura, de inclusão social, de proteção ambiental e inclusive em termos de uma poupança direcionada diretamente a garantir o bem-estar das gerações futuras (2002a, p.329).

De tudo isso, fica perfeitamente claro que Rawls rejeita de maneira peremptória o princípio básico do *laissez-faire*, a saber, a ideia de uma esfera econômica, não política, que tem condições, devido à sua dinâmica interna, de autorregular-se, de autoestabilizar-se e, consequentemente, de estabilizar e de regular a sociedade como um todo. Diz Rawls: "[...] a *mão invisível*, antes de socializar seus frutos, possui uma tendência oligopolista e excludente" (2002a, p.77). Assim, o controle público na concentração da propriedade e da riqueza assume um papel fundamental na percepção de Rawls: "a ampla dispersão da propriedade [...] é, ao que parece, uma condição necessária à manutenção das liberdades iguais" (2002a, p.306). E complementa:

> *[...] a interpretação liberal dos dois princípios busca mitigar a influência das contingências sociais e da boa sorte espontânea sobre a distribuição das porções. Para atingir esse objetivo, é necessário impor ao sistema social condições estruturais básicas adicionais. Devem ser estabelecidas adaptações ao mercado livre dentro de uma estrutura de instituições políticas e legais que regule as tendências globais dos eventos econômicos e que preserve as condições sociais necessárias para a igualdade equitativa de oportunidades. Os elementos dessa estrutura são bastante familiares, embora possa ser útil relembrar a importância de se evitarem acúmulos excessivos de propriedade e de riqueza, bem como de se manterem iguais oportunidades de educação para todos. As oportunidades de se atingir conhecimento cultural e*

> *qualificações não deveriam depender da posição de classe de uma pessoa e, assim, o sistema escolar, seja público ou privado, deveria destinar-se a eliminar as barreiras de classe"*[33].

Deve-se ter claro, e com isso finalizo essa parte, que, para Rawls, há uma primazia da justiça social como condição tanto de uma sociedade e sociabilidade estáveis e voltadas ao benefício recíproco quanto como condição para o desenvolvimento pleno de cada indivíduo. Nesse sentido, a justiça política, com sua ênfase na realização da justiça distributiva, aponta para a percepção de que os problemas sociais, políticos e econômicos, no entender de Rawls, "[...] decorrem da injustiça política, com todas as suas crueldades e brutalidades" (2001, p.7-8). A justiça é a primeira virtude, ainda segundo Rawls, das instituições sociais, de modo que instituições injustas devem ser reformuladas ou, caso isso não seja possível, abolidas, já que é por meio delas que as injustiças sociais são causadas e se solidificam[34].

Rawls: uma crítica ao neoliberalismo

Ao finalizar esse já longo artigo, gostaria de levantar a tese de que a crítica de Rawls ao liberalismo clássico e, sob muitos aspectos, sua reformulação desse mesmo liberalismo a partir de pressupostos da crítica hegeliano-marxista dirigem-se não de maneira direta ao próprio liberalismo clássico, mas fundamentalmente ao neoliberalismo, à medida que este, em sua crítica ao Estado de bem-estar social, retoma algumas das teses elaboradas por Locke e, principalmente, por Adam Smith. Efetivamente, em primeiro lugar, não faria sentido uma retomada daquelas críticas num contexto de consolidação do Estado de bem-estar social e, nesse sentido, de consolidação de uma economia de mercado em algum aspecto poderoso planificada por parte do Estado, (Estado de bem-estar) que passa a adquirir, junto a isso, a tarefa de realizar um mínimo de justiça distributiva. O capitalismo de *laissez-faire* e seu correlato, o Estado liberal clássico (como tematizados por John Locke e, principalmente, por Adam Smith), receberam seu golpe de morte com a crise econômica de fins da década de 1920. A reformulação keynesiana da economia norte-americana no governo de Franklin Delano Roosevelt aponta para uma centralidade do próprio Estado tanto no que diz

[33] RAWLS, John. *Uma teoria da justiça*, § 12, p. 77.
[34] Cf.: RAWLS, John. *Uma teoria da justiça*, §01, p. 03-04.

respeito à condução e à promoção do desenvolvimento econômico (pressupondo até um controle, ainda que mínimo, da acumulação da riqueza e da propriedade) quanto no que diz respeito à realização da inclusão social. É interessante, nesse sentido, que o próprio Rawls saliente *deficits* do próprio Estado de bem-estar social, especificamente no que diz respeito ao fato de não regular convenientemente a acumulação da propriedade e da riqueza, permitindo que elas progressivamente acabem sendo concentradas em poucas mãos. Ou seja, o Estado de bem-estar social apenas imperfeitamente consegue realizar uma conciliação entre desenvolvimento econômico e inclusão social, à medida que não ataca frontalmente o problema da concentração da riqueza e da propriedade em poucas mãos, que leva à própria hegemonia política desses grupos já economicamente hegemônicos[35].

Ora, digo que a reformulação do liberalismo clássico, por parte de Rawls, tem por objetivo uma crítica ao neoliberalismo pelo fato de, a partir da década de 1970, emergir nas democracias desenvolvidas – na Inglaterra e nos Estados Unidos, em particular – a doutrina neoliberal como crítica exatamente do Estado de bem-estar social e defendendo uma volta a alguns princípios da economia de *laissez-faire*. Friedrich Hayek, por exemplo, tem no conceito de *evolução espontânea* da sociedade o principal argumento para afirmar que a planificação do Estado em relação a essa mesma sociedade de uma maneira geral e em relação à economia em particular é descabida e ilegítima, seja porque esse mesmo Estado não tem condições de ter uma visão abrangente e a sabedoria necessárias para planificar de maneira satisfatória a sociedade como um todo, seja porque a sociedade evolui a partir das *ações não intencionais dos próprios indivíduos*. Aliás, e isso é muito importante, Hayek afirma que a sociedade, em sentido estrito, não existe; existem apenas os indivíduos, que se associam com vistas à própria proteção[36].

Se tivermos claro que as obras *O liberalismo político* e *Justiça como equidade*: uma reformulação, de Rawls, foram escritas, respectivamente, nos anos de 1993 e 2001; se tivermos claro que, da década de 1970 em diante, o Estado de bem-estar social (seu papel, sua crise) foi o centro da

[35] Cf.: RAWLS, John. *Justiça como equidade*: Uma reformulação, §42, pp. 196-197.
[36] Cf.: HAYEK, Friedrich August von. *O caminho de servidão*, p. 109-122; HAYEK, Friedrich August von. *Arrogância fatal*: os erros do socialismo, p. 27-59; HAYEK, Friedrich August von. *Law, constitution and liberty* (V. III): the political order of a free people, p. 50; BUTLER, Eamon. *A contribuição de Hayek às ideias políticas e econômicas de nosso tempo*, p. 33-36.

dinâmica política das democracias ocidentais, em que temas como *ingovernabilidade*, *privatização*, *desestatização* e mesmo *globalização* passaram a ter lugar central na agenda pública de discussão política com cada vez mais intensidade; e se tivermos claro que, começando com Ronald Reagan, a partir do início de 1980, reformas de cunho neoliberal passaram a ser gradativamente instauradas nos Estados Unidos, poderemos perceber o sentido de muitas críticas em relação ao regime de *laissez-faire*, por parte de Rawls, e a sua defesa de um controle público da acumulação da propriedade e da riqueza, com ênfase em educação, em distribuição de renda e mesmo na própria necessidade de maior equalização das liberdades políticas entre os cidadãos. Mas perceberemos, principalmente, em se tratando da obra *Justiça como equidade*: uma reformulação, a recusa rawlsiana em relação aos princípios teóricos (*mão invisível*, Estado mínimo, competição entre indivíduos egoístas e evolução espontânea) do capitalismo de *laissez-faire* e a sua defesa da intrínseca relação entre direitos individuais fundamentais, direitos políticos e direitos sociais de cidadania, como mutuamente dependentes e complementares. Nesse caso, o objeto central da justiça política, em uma sociedade democrática, é a realização dessas três dimensões como condição para a efetividade de cada uma em particular. E isso passa necessariamente, contra as pretensões do neoliberalismo, por uma reestruturação do Estado de bem-estar social, (uma reestruturação) que corrija seus (do Estado de bem-estar) defeitos (principalmente este de não atuar como freio no que diz respeito à concentração monopolística da propriedade e da riqueza nas mãos de poucos grupos), mas que enfatize o controle e a regulação públicos da acumulação ou, em outro sentido, da distribuição da propriedade e da riqueza – uma *tarefa pública* sob muitos aspectos *permanente*.

Referências

ARISTÓTELES. *Política*. 3. ed. Tradução, Introdução e Notas de Mário da Gama Kury. Brasília: Editora da UNB, 1997.

BUTLER, Eamon. *A contribuição de Hayek às ideias políticas e econômicas de nosso tempo*. Tradução de Carlos dos Santos Abreu. Rio de Janeiro: Instituto Liberal, 1987.

HABERMAS, Jürgen. *Problemas de legitimación en el capitalismo tardio*.

Traducción de José Luis Etcheverry. Madrid: Ediciones Cátedra, 1999.

_____. *Direito e democracia – entre facticidade e validade* (2 vol.). Tradução de Flávio Benno Siebeneichler. Rio de Janeiro: Tempo Brasileiro, 2003.

HAYEK, Friedrich August von. *O caminho de servidão*. Tradução e revisão de Anna Maria Capovilla, José Ítalo Stelle e Liane de Morais Ribeiro. Rio de Janeiro: Instituto Liberal, 1987.

_____. *Arrogância fatal*: *os erros do socialismo.* Tradução de Anna Maria Capovilla e de Candido Mendes Prunes. Porto Alegre: Editora Ortiz, 1995.

_____. *Law, legislation and liberty: the political order of a free people*. Chicago (2 vol.). The University of Chicago Press, 1990.

HEGEL, G. W. F. *Escritos de juventud*. Edición, introducción y notas de Jose M. Ripalda. México: Fondo de Cultura Económica, 1988.

_____. *Lecciones sobre la filosofía de la história universal*. Prólogo de José Ortega Y Gasset. Traducción de José Gaos. Madrid: Alianza Editorial, 1982.

_____. *História da filosofia*. Tradução de Maria Rodrigues e de Hans Harden. Brasília: Editora da Universidade de Brasília, 1995.

_____. *Princípios da filosofia do direito*. Tradução de Orlando Vittorino. São Paulo: Martins Fontes, 1997.

HUME, David. *Investigação acerca do entendimento humano*. Tradução de Anoar Aiex. São Paulo: Editora Nova Cultural, 1999.

LOCKE, John. *Segundo tratado sobre o governo civil e outros escritos*. Tradução de Júlio Fischer. São Paulo: Martins Fontes, 2005.

_____. *Carta sobre a tolerância*, p. 235-239. In: LOCKE, John. *Segundo tratado sobre o governo civil*. Tradução de Júlio Fischer. São Paulo: Martins Fontes, 2005.

MARX, Karl. *Manuscritos econômico-filosóficos*. Tradução de Alex Marins. São Paulo: Martin Claret, 2006.

_____. *O capital*: *crítica da economia política* - (v. I, livro I): o

processo de produção do capital. Tradução de Régis Barbosa e de Flávio R. Kothe. São Paulo: Abril Cultural, 1988.

MARX, Karl; ENGELS, Friedrich. *A ideologia alemã*. Tradução de Luiz Claudio de Castro e Costa. São Paulo: Martins Fontes, 2008.

PLATÃO. *A república*. Tradução de Pietro Nassetti. São Paulo: Martin Claret, 2007.

RAWLS, John. *Uma teoria da justiça*. Tradução de Almiro Pisetta e de Lenita Maria Rímoli Esteves. São Paulo: Martins Fontes, 2002a.

_____. *Justiça e democracia*. Tradução de Irene Paternot. São Paulo: Martins Fontes, 2000.

_____. *O liberalismo político*. Tradução de Dinah de Abreu Azevedo. Brasília: Instituto Teotônio Vilela; São Paulo: Editora Ática, 2002b.

_____. *O direito dos povos*: *seguindo de A ideia de razão pública revisitada*. São Paulo: Martins Fontes, 2001.

_____. *História da filosofia moral*. Tradução de Ana Aguiar Cotrim. São Paulo: Martins Fontes, 2005.

_____. *Justiça como equidade*: *uma reformulação*. Tradução de Cláudia Berliner. São Paulo: Martins Fontes, 2003.

ROSANVALLON, Pierre. *O liberalismo econômico*: *história da ideia de mercado*. Tradução de Antônio Penalves Rocha. Bauru: EDUSC, 2002.

ROUSSEAU, Jean-Jacques. *Discurso sobre a origem e os fundamentos da desigualdade entre os homens*. Tradução de Lourdes Santos Machado. São Paulo: Nova Cultural, 1999.

SMITH, Adam. *A riqueza das nações* (v. I). Tradução e Notas de Teodora Cardoso. Lisboa: Fundação Calouste Gulbenkian, 1999.

_____. *A riqueza das nações* (v. II). Tradução de Luís Cristóvão de Aguiar. Lisboa: Fundação Calouste Gulbenkian, 1999.

TOSEL, André. *Hegel: o bem para além da necessidade*, p. 517-528 (esta citação encontra-se na página 522; o grifo é nosso). In: CAILLÉ, Alain; LAZZERI, Christian; SENELLART, Michel (Orgs.). *História argumentada*

da filosofia moral e política. Tradução de Alessandro Zir. São Leopoldo: Editora da Unisinos, 2004.

WACQUANT, Loïc. *As prisões da miséria*. Tradução de Andre Telles. Rio de Janeiro: Jorge Zahar Editor, 2001.

WEBER, Max. *A ética protestante e o "espírito" do capitalismo*. Tradução de José Marcos Mariani de Macedo. São Paulo: Companhia das Letras, 2004.

O pluralismo cultural no currículo e a universalidade dos direitos morais sob o ponto de vista da crítica habermasiana

Claudia Castro de Andrade
Universidade do Estado do Rio de Janeiro

Resumo

Neste trabalho discuto a questão curricular como processo político no qual estão envolvidas as lutas ideológicas que buscam preencher de sentidos os documentos e práticas curriculares. Considerando a escola como espaço de interação capaz de produzir e reproduzir valores, reflito sobre os movimentos identitários a favor do reconhecimento ao pluralismo cultural, para viabilizar o questionamento sobre a igualdade de direitos e do reconhecimento à diferença. Considerando que as tentativas de fixação de sentidos não ocorrem pacificamente, cumpre ressaltar as disputas político-ideológicas que tentam ocupar espaço nas negociações curriculares. Em relação ao pluralismo cultural, recorro a Habermas em suas considerações sobre facticidade e aceitabilidade racional, além de suas reflexões sobre a diferença entre os discursos dos direitos morais universais e o discurso do direito democrático à pluralidade cultural. Trago também as leituras de Alice Casimiro Lopes e Stephen Ball, em relação ao ciclo contínuo das políticas curriculares. Questiono o racionalismo dogmático que desconsidera a diferença cultural, pautando-se por uma construção de cultura com sentido universal, e também o irracionalismo do relativismo cultural que não problematiza hábitos e valores culturais devido ao fato de justificar toda e qualquer cultura como válida e aceita.

Palavras-chave: currículo, diferença, políticas educacionais, universalismo moral, pluralidade cultural

Abstract

This article discusses school curriculum as a political process that involves ideological struggles seeking to respond to official curricular practices. Considering school as a space for interaction that reproduces the existing values and also transforms them, I try to foster the debate about identity movements favoring the recognition of cultural pluralism,

defending equal rights and the recognition of difference. Accepting that attempts to fix meanings to curriculum do not occur in a peaceful manner, one must note the political and ideological disputes that exist in curriculum choices. Taking into account cultural pluralism, I turn to Jürgen Habermas on facticity and rational acceptability and how social order occurs even in pluralistic societies. I also take Habermas' reflections on the difference between speeches about rights and the speech about universal moral democratic right in cultural diversity. This article also deals with the work of Stephen Ball and Alice Casimiro Lopes on the ongoing cycle of curriculum policies. I also question (1) a dogmatic rationalism that ignores cultural differences considering culture in a universal sense and (2) the irrationality of a cultural relativism that does not discusses habits and cultural values since it justifies any and every culture as valid and accepted.

Keywords: *curriculum, difference, educational policy, moral universalism, cultural plurality*

O pluralismo cultural como proposta curricular

É importante ressaltar de antemão que será considerado neste trabalho a pluralidade cultural como característica intrínseca ao conceito de sociedade democrática. Diante disso, entende-se que recusar a pluralidade cultural é um tipo de violência que afeta grupos "minoritários" que não têm suas características devidamente reconhecidas. Obviamente que ao se fazer tal afirmação, ampliamos o conceito "violência", extrapolando o sentido da clássica interpretação reducionista que considera violência somente como violência física. Inegavelmente isso tem um preço, pois a ampliação do termo "violência" pode levar a uma banalização do uso de seu conceito, podendo causar até mesmo um esvaziamento de seu sentido. Mas, ao mesmo tempo, considerar somente a violência física como violência (sentido clássico do termo) é reducionismo, porque impede a problematização de outras práticas abusivas e produz uma hierarquização entre essas práticas, tornando umas mais aceitas que outras, à medida que algumas são consideradas violências e outras, não.

A questão abordada neste trabalho refere-se à ausência do reconhecimento ao pluralismo cultural presente no currículo escolar. Entendendo, portanto, o currículo como um mecanismo definidor da

realidade escolar e não só dos documentos que determinam as políticas públicas para a educação, mas também do cotidiano escolar, pensaremos a pluralidade em relação à escola e aos documentos capazes de viabilizar uma educação mais igualitária que possa ser instrumento para uma sociedade plural, a partir da compreensão de que as práticas do cotidiano escolar transpõem, na verdade, os muros da escola.

Em vista disso, podemos dizer, sem medo de errar, que pensar o currículo é pensar a imensa rede identitária que busca conquistar coro no espaço escolar. Isso decorre do fato de que vários grupos, estimulando as trocas interculturais, discutiram (e discutem) a necessidade de uma problematização acerca do pluralismo cultural, como também a necessidade de um reconhecimento aos diversos discursos contra-hegemônicos de grupos "minoritários" pelo direito à diferença. Desse modo, esses grupos organizaram-se para pensar e questionar o papel da escola frente à urgência de um cenário social pluralista que concebe perspectivas culturais tão distintas entre si. Tal concepção, vale ressaltar, parte do entendimento de cultura, não como algo positivo[1], mas sim como algo construído por todos nós e em constantes mudanças.

A concepção de cultura como algo previamente dado, resvala na ideia de *a priori,* isto é, de algo a ser descoberto, revelado, ou seja, que preexiste ao homem, cabendo a ele apenas descobrir essa cultura preexistente. Entretanto, no mundo podemos facilmente perceber manifestações culturais hegemônicas e dominantes e outras que são, até mesmo, desvalorizadas, como se determinadas culturas fossem certas e verdadeiras e outras fossem erradas e falsas. Não compreender que a cultura é construída corrobora para esse entendimento. A ideia de cultura *a priori* camufla a vitória de uma cultura que se tornou dominante, fazendo-nos esquecer que, na verdade, ela foi construída e resultante de um embate vitorioso, levando-nos, ao mesmo tempo, a crer que a cultura é transcendente ao homem e, portanto, independente de sua própria vontade. Vista sob esse aspecto, a cultura preexistiria ao homem e não seria resultado de relações de poder e lutas ideológicas, mas sim uma cultura autônoma, que se autodefine, e autotélica, que tem fim nela mesma, à medida que existe independentemente das ações e do querer do homem.

[1] Do latim *positum,* que significa "o que está posto", "o que está dado".

Preferindo o uso do termo identificação, que leva a um entendimento de identidade com movimento, ou seja, como um processo, em vez do termo identidade, que remete a algo fixo e estável, podemos dizer que, da mesma forma que a cultura, essa concepção apriorística também interfere no processo de identificação do indivíduo, o qual também passa a ser legitimado a partir da adesão da maioria. O modelo identitário dominante se naturaliza, tal qual a cultura, de modo que não se percebe que sua ampliação e estabelecimento, considerados certos e verdadeiros, já foram, na verdade, resultados de uma luta ideológica que busca uma hegemonia com vistas à universalização de seus conceitos. O resultado é, portanto, uma identificação padronizada que conseguiu representação e que será considerada como processo comum partilhado por todos, com vistas à homogeneização, à medida que se encontra naturalizada. Como explica Hall,

> *Na linguagem do senso comum, a identificação é construída a partir do reconhecimento de alguma origem comum, ou de características que são partilhadas com outros grupos ou pessoas, ou ainda a partir de um mesmo ideal* (HALL, 2000, p. 103).

Ao contrário, na concepção construtivista de cultura (e processo identitário), entende-se que somos nós que a significamos, sendo ela passível, portanto, de diferenças relativas ao contexto espaço-tempo. Assim, compreendendo-se que a cultura é, então, construída por nós e que, além disso, é relativa ao contexto, compreende-se também que não há uma cultura universal, mas sim contextual, circunstancial. Não há, portanto, uma cultura certa ou verdadeira, falsa ou errada, pois considera-se que todas elas têm, cada uma, sua respectiva validade ontológica.

Contudo, vale lembrar que não se pode pensar ingenuamente que o reconhecimento de uma cultura que visa legitimar suas múltiplas manifestações ocorrerá na escola ou na sociedade de forma pacífica e sem resistências. Cada organismo investido de seus conceitos e valores defenderá seu posicionamento de qualquer outro que lhe contraponha. Isso porque os valores de um determinado sujeito chocam-se com os valores de outro sujeito, quer seja na tentativa de definir um sistema social homogêneo ou heterogêneo, quer seja na tentativa de definir um modelo curricular com vistas a homogeneizar ou a heterogeneizar o

espaço de convivência escolar, o qual, é importante destacar, transpõe os limites dos muros da escola, podendo reproduzir ou modificar os valores contidos nesse espaço.

> *[...] a interpretação da pluralidade cultural como pluralidade de razões permite que se compreenda a cultura como um campo de diversas e múltiplas culturas, constituídas por múltiplas racionalidades em constante embate e conflito* (LOPES, 1999, p. 68).

Ciente, portanto, do papel transformador da escola e de seu poder de produção e reprodução, como também das relações de poder que tentam nortear as propostas educacionais no âmbito curricular é que a heurística sobre as políticas curriculares torna-se tão urgente e relevante. Desse modo, a urgência, por exemplo, das propostas multiculturais que tencionam contemplar discursos "minoritários", que não são contemplados nem textual nem discursivamente, ressalta a necessidade de analisar as lutas hegemônicas e as relações de poder que envolvem essas políticas curriculares.

O currículo pode, pois, contribuir para a perpetuação de valores como também pode modificá-los. Por essa razão, precisamos perceber a questão curricular como um processo político que envolve a tentativa de fixação de sentidos, tanto por meio de textos, como sistema definidor de um modelo padrão, tal como as cartilhas educacionais que buscam universalizar um modelo de ensino quanto por discursos que buscam definir a realidade por meio de mecanismos simbólicos.

Nesse sentido é que se considera de grande valia as reflexões de Ball e Bowe (1992), que analisam o processo de formulação e implementação das políticas educacionais como um ciclo contínuo que envolve variados contextos: um contexto de influência, referente aos discursos de determinados grupos de interesses ideológicos que vão tentar influenciar os rumos do processo político; um contexto de produção, referente a um campo de disputas político-ideológicas em que se encontram presentes os paradoxos e contradições constantes dessas disputas; e um contexto da prática, o local para onde se dirigem os interesses e objetivos dessas influências e produções, no qual as resoluções resultantes desse processo serão possivelmente reinscritas, negociadas e passíveis de serem até mesmo alteradas.

> *Investigar os discursos implica investigar as regras que norteiam as práticas. Assim, ao pensar as políticas como discursos, Ball adverte que os conhecimentos subjugados não são completamente excluídos da arena política, mas certos discursos nos fazem pensar e agir de forma diferente, limitando nossas respostas e mudanças. Os efeitos das políticas como textos e como discursos são contextuais e estabelecem constrangimentos para as políticas. Na medida em que são múltiplos os contextos produtores de textos e discursos – incluindo Estado, governos, meio acadêmico, práticas escolares, mercado editorial –, com poderes assimétricos, são múltiplos os sentidos e significados em disputa* (LOPES, 2007, p. 207).

Ao entender, então, os textos e os discursos como efeitos de segmentos políticos diversos, Ball vai considerar, para a análise de seus estudos, os princípios estruturalistas e pós-estruturalistas (teoria discursiva) presentes nas negociações curriculares. Como destaca Lopes, "Ball (1994) trabalha com as definições políticas como textos e como discursos, associando princípios estruturalistas e pós-estruturalistas" (2007, p. 206).

A partir das considerações sobre currículo e pluralismo cultural, podemos concluir que a proposta de inserir temas desse âmbito choca-se, por assim dizer, com disputas ideológicas que resistem a esses tipos de discursos, ditos pós-modernos. O pluralismo cultural como proposta curricular esbarra, portanto, em uma série de dificuldades, entre as quais, a de impedir o reconhecimento dos discursos pela defesa à diferença, negando, ao mesmo tempo, a importância deles.

Retomando o tema da questão cultural, podemos concluir que toda e qualquer cultura é construída, mas embora se reconheça essa construção "tal concepção não significa, contudo, a defesa do relativismo, segundo o qual qualquer método, qualquer teoria, qualquer política, qualquer ética, qualquer cultura podem ser vistos como válidos." (LOPES, 1999, p. 67).

Segundo as palavras de Lopes, reconhece-se, portanto, ontologicamente (e epistemologicamente), a relatividade da cultura como construto humano, mas com isso não se pretende relativizar os

hábitos e costumes culturais ao ponto de não compreendermos seus limites. O fato, pois, de considerar a cultura como algo relativo não pressupõe que ela será socialmente válida, e aceita. Busca-se, com isso, ressaltar que toda e qualquer cultura é criada e estabelecida *a posteriori*, mas disso não se pode inferir sua validade e aceitabilidade.

Partindo, então, da compreensão de que uma sociedade democrática implica a urgência de reflexões sobre pluralismo e diferença, considera-se relevante que a educação, tanto em relação ao seu conceito geral quanto na forma de um segmento organizado, possa questionar sua própria função nessa sociedade (democrática) e que seja capaz de corroborar para uma política curricular democrática.

Em vista disso, vale ressaltar que a legitimação dos ideais democráticos ocorre, entre outras coisas, pelo reconhecimento às diferenças e aos ideais de uma sociedade plural, e da participação política de múltiplos segmentos para o pleno exercício e fortalecimento da cidadania. Entretanto, a cidadania, como característica de uma sociedade democrática, é entendida aqui como a representação das várias e diferentes manifestações culturais que buscam conquistar espaço, e não como representação de um "todo" que suprime as diferenças de suas partes. Mas, em contrapartida, é a representação do "todo" que tende a caracterizar nosso entendimento de cidadania (e de uma soberania popular) que se fundamenta, por sua vez, no ideal de uma intersubjetividade capaz de representar plenamente os anseios de todo um conjunto, sendo, pois, considerada válida por isso. Em outras palavras, a cidadania é entendida e validada por ser a representação de um todo social coletivo. Desse modo, a relação entre sujeitos é entendida, então, como algo equivalente que se torna produto de um acordo uniforme e homogêneo. Nesse sentido,

> *A cidadania é vista através do modelo da pertença a uma comunidade ético-cultural que se determina a si mesma, ou seja, os indivíduos estão integrados na comunidade política como partes de um todo, de tal maneira que, para formar sua identidade pessoal e social, necessitam do horizonte de tradições comuns e de instituições políticas reconhecidas* (ARAÚJO, 2010, p. 130).

Mais uma vez, a ideia de características partilhadas se destaca como relevante para tornar um determinado conceito em um conceito hegemônico. Assim como a identidade descrita por Hall busca por essas características partilhadas, o mesmo ocorre com a cidadania que, de acordo com Araújo, busca ser representada a partir do que é reconhecido por todos, ou seja, como algo já naturalizado no senso comum, que se torna facilmente reconhecido e aceito, mas que não contempla a realidade de vários indivíduos.

O filósofo alemão Jürgen Habermas considera, por exemplo, que a cidadania pode tornar-se soberana – como também ocorre com a própria soberania popular –, e pode-se dizer que isso acontece quando ela se naturaliza, à medida que se retira "para o anonimato dos processos democráticos e para a implementação jurídica" que "resulta das interações entre a formação da vontade institucionalizada constitucionalmente e esferas públicas mobilizadas culturalmente" (1997, p. 24). Porém, esse entendimento reduz a própria cidadania a um conjunto de leis, que se legitimam e se naturalizam como verdadeiro representante do próprio conceito de cidadania. A cidadania é, nesse sentido, entendida pelos ideais comunitários que se legitimam por meio da representação do todo pressupondo, ao mesmo tempo, a supressão das partes, pois não considera a dimensão do indivíduo como parte desse todo.

Para Habermas, a soberania não se reduz a uma representação totalizante do coletivo, o que pode descaracterizar o indivíduo, nem pode ser ocultada pelas funções legislativas das instâncias políticas. Como o próprio Habermas comenta, "a soberania não precisa concentrar-se no povo nem ser banida para as competências jurídico-constitucionais" (1997, p. 24). A cidadania considerada apenas em seus aspectos legislativos, ou melhor, políticos, reduz-se ao conceito de algo determinado contratualmente, retirando, desse modo, qualquer possibilidade de se compreender a cidadania como algo que se estabelece na *práxis* do cotidiano por consenso entre as partes. A cidadania, sob esse aspecto, seria garantida apenas pelo estabelecimento de regras impostas à sociedade.

Entretanto, a cidadania considerada por Habermas, ao contrário de uma regulação institucional, estaria fundamentada no princípio da "democracia deliberativa", na qual os pressupostos normativos são

definidos pela própria sociedade civil e não pelos mecanismos políticos que a representam. A cidadania então, para Habermas, fundamenta-se pela ideia de cidadãos livres que possam legitimar suas decisões na esfera pública, o que conflita com o fato comentado anteriormente que diz respeito aos mecanismos políticos envolvidos nos documentos que regulam a educação, como, por exemplo, no que concerne o reconhecimento à pluralidade cultural como ação afirmativa para formação do cidadão.

Porém, a proposta de Habermas sobre o tema "cidadania" não se determina nem em uma cidadania soberana nem em uma cidadania particularista, mas sim em uma cidadania democrática (1997, p. 304). Por cidadania democrática, podemos tomar a liberdade de concebê-la como a representação de todas as variantes culturais contidas no interior de um Estado.

Os conceitos de Habermas sobre a diferença entre o pluralismo cultural e o universalismo moral

Entende-se, com base em Habermas, que o reconhecimento ao pluralismo cultural não, necessariamente, implica um relativismo extremo das questões culturais. O que se destaca, desse modo, é que o conhecimento e a racionalidade não podem ser tomados como verdades universais *a priori* e, desse modo, não se pode considerar que haja uma cultura verdadeira e certa, e outra falsa e errada. Nesse aspecto, a ideia de pluralismo converge (e se justifica) para o entendimento habermasiano do "agir comunicativo", que rejeita a noção de normas morais fundadas na perspectiva transcendental de uma concepção totalizante da realidade que pretende nomear e definir uma cultura como legítima e as demais como falsas, ao mesmo tempo em que recusa a ideia de não problematizar os fenômenos advindos dos diversos tipos de cultura.

Em outras palavras, pode-se dizer que Habermas considera a legitimidade de uma diversidade cultural, sem dúvida, mas considera que disso não se pode abstrair sua validade ética e moral. Da mesma forma que o homem é responsável pela construção da cultura, ele é responsável também pelos atos que venha a fazer em nome dessa cultura. Assim sendo, a cultura e as ações humanas são, pois, passíveis de

verificação para uma validação normativa. A legitimação da diversidade cultural não pressupõe, portanto, sua validação normativa.

Habermas (1992), então, nega uma razão dogmática, fundamentada por leis *a priori,* mas nega também uma razão irracional, fundamentada por extremo relativismo, que não problematiza o *modus operandi* de determinadas culturas, sob a justificativa de enxergar toda e qualquer cultura como socialmente válida e aceita, e, além disso, por considerar essa cultura como não sendo nem mesmo passível de discussões a respeito da legitimidade de seu uso e aplicação.

Diante disso, ele diferencia, então, facticidade e validade, ou seja, comenta a diferença que há entre o que é passível de ser feito (facticidade) e sua aceitabilidade racional (validade) que se naturaliza nas práticas discursivas, nos fazendo encarar os fatos como válidos. É a naturalização dos fatos, aliás, que nos faz entendê-los como válidos, ou melhor, que nos faz aceitá-los sem nenhum questionamento. Desse modo, Habermas vai pensar o pluralismo cultural em relação à universalidade dos direitos morais; aliás, é importante logo lembrar, que, para Habermas, a universalidade não anula as diferenças existentes na pluralidade cultural.

Primando pela ordem social, Habermas identifica a validade dessa ordem diante da inevitável complexidade das sociedades pluralistas, pois essa complexidade pode levar a um indeterminismo na própria concepção ética, e conduzir, dessa forma, a um dissenso. Nesse sentido, há que se considerar o que é coletivo (relações entre indivíduos) e o que é individual, isto é, o pluralismo cultural precisa garantir a universalidade dos direitos morais e também as individualidades de cada um. É assim que Habermas diferencia o que ele chama de perspectiva horizontal, referente às relações da coletividade, da perspectiva vertical, que se refere à individuação do sujeito.

Um hábito cultural, como construção humana e relacionada ao espaço de interatividade humana, não pode, por um lado, ocorrer de forma arbitrária contra o indivíduo e, por outro lado, não deve ser visto como garantia de qualquer possibilidade de ações desse mesmo indivíduo. Habermas admite que, sem dúvida, a autonomia produzida pelo pluralismo da sociedade moderna rompeu com o modelo tradicional que representava uma ideia universalista do real e o transcendentalismo

das verdades consideradas universais. Mas ele lembra, por outro lado, que essa autonomia, e esse pluralismo, não implicam uma anomia social e uma autonomia completa das ações humanas, que não considera o homem como responsável por elas.

Vale lembrar novamente que, para Habermas, o pluralismo cultural não está em oposição a um universalismo moral. Pode-se pensar, portanto, num pluralismo cultural que não desconsidere um universalismo moral. A questão não é de oposição, mas de problematização e insere uma necessidade de investigação acerca da validade normativa dos valores num contexto cultural tão pluralista.

A teoria discursiva (e sistêmica) de Habermas inscreve-se no debate entre direitos morais universais e o direito democrático à pluralidade cultural. Conciliar a diferenciação e a heterogeneização propostas por esse pluralismo cultural ao igualitarismo e à homogeneização propostos pelo universalismo moral é, para ele, tarefa do "agir comunicativo", que acontece na linguagem como algo relacional, ou seja, uma integração entre indivíduos no cotidiano, cuja normatividade da ordem social é garantida não de forma contratual, mas nessa mesma dinâmica social.

Desse modo, influenciado pela "teoria do discurso", Habermas considera que o entendimento entre esses indivíduos não é construído pelo papel de sujeitos privados nem por um modelo contratual, mas sim quando assumem "a perspectiva de participantes em processos de entendimento que versam sobre as regras de sua convivência" (1997, p. 323). Há um consenso coletivo que se determina no cotidiano por meio de normas universais que tornam possível o ser humano viver e conviver socialmente com outros indivíduos, e que ocorre, não a partir de um modelo contratual capaz de ser mantenedor absoluto do bem-estar da humanidade, mas pela argumentação discursiva entre sujeitos.

De acordo com tudo o que foi discutido, percebe-se que a questão do pluralismo cultural traz uma série de discussões de caráter filosófico, como as diferenças entre o universal e o individual; o *a priori* e o *a posteriori*; o relativo e o totalizante; o homogêneo e o heterogêneo. Assim, ainda na questão da perspectiva horizontal (relacional) e vertical (individual), cumpre ressaltar que para Habermas, a ética do bem comum conduz à perda da unidade, isto é, à perda da perspectiva vertical, em função de sua característica holística e homogeneizante.

Diferentemente, a ética habermasiana, sendo uma ética argumentativa, não privilegia nem os extremos de uma ideia tomista (particularista) nem os extremos de uma ideia holística (geral).

A ética do discurso proposta por Habermas não é a ética dos extremos. Ela não considera de forma unívoca uma ideia particularizada que defende uma perspectiva individualizante, como também não considera exclusivamente uma ideia globalizante que possui caráter monista, homogêneo e universal.

A solução para essas diferenças entre a parte e o todo é resolvida no fato de que para Habermas a ética é reflexiva, pois está vinculada a uma ação comunicativa que se funda, por sua vez, no processo da vida social e está, desse modo, inserida nas ações práticas do cotidiano, não sendo nem individualizante e subjetivista e nem coletivista e materialista, ou seja, não pode haver para Habermas uma ética do bem comum que desconsidere as particularidades nem uma ética totalmente particularizada que desconsidere a coletividade e as relações humanas.

> *O princípio da ética do Discurso refere-se a um procedimento, a saber, o resgate discursivo de pretensões de validez normativa; nessa medida, a ética do Discurso pode ser corretamente caracterizada como formal. Ela não indica orientações conteudísticas, mas um processo: o Discurso prático. Todavia, este não é um processo para a geração de normas justificadas, mas, sim, para o exame da validade de normas propostas e consideradas hipoteticamente* (HABERMAS, 1989, p. 126, grifos nossos).

Além disso, a característica pragmática da ética habermasiana pressupõe uma ética pós-metafísica, e ressalta a responsabilidade do homem no âmbito de seu "agir comunicativo". Na concepção metafísica, entretanto, essa responsabilidade ocultava-se no transcendentalismo que poderia retirar do sujeito a imputabilidade por suas ações. Isso ocorria porque o pensamento metafísico, segundo ele, tende a justificar a moral pela religião e pela própria metafísica, enquanto a ética argumentativa de Habermas parte do pressuposto de que as questões morais devem ser analisadas sob a luz da autonomia das ações humanas,

sem se prender a modelos religiosos prescritivos e universais, para que com isso o homem não seja considerado inimputável por suas ações.

Por esse motivo é que Habermas posiciona-se contra a idéia kantiana de um *aufklä rung* no qual o conhecimento humano depende ainda de um esclarecimento a ser conquistado pelo homem, como um ideal a ser alcançado. Diferentemente da perspectiva kantiana, Habermas entende o conhecimento e a responsabilidade das ações humanas como constitutivos da própria vida, do agir, do cotidiano. Dessa forma, Habermas se aproxima da corrente pragmática ao considerar que a noção de um determinismo *a priori* é inconciliável com a compreensão de indivíduos-agentes responsáveis, pois o que importa para Habermas é ressaltar a autonomia individual do sujeito e a imputabilidade por suas ações.

> *Na medida em que os participantes da comunicação compreendem aquilo sobre o que se entendem como algo em um mundo, como algo que se desprendeu do pano de fundo do mundo da vida para se ressaltar em face dele, o que é explicitamente sabido separa-se das certezas que permanecem implícitas, os conteúdos comunicados assumem o caráter de um saber que se vincula a um potencial de razões, pretende validade e pode ser criticado, isto é, contestado com base em razões* (HABERMAS,1989, p. 169).

Em outras palavras, Habermas compara o saber intuitivo que implica uma pretensão de validade sem que nunca se tenha problematizado essa mesma validade, com o saber construído, que problematiza essa pretensão de validade pelo uso da razão. O conceito, então**,** de razão**,** no sentido habermasiano, não tem sentido, *a priori*, mas tem, sim, um sentido de racionalismo pragmático. A razão para Habermas não pode ser reduzida a um caráter prescritivo de produção de normas nem transcendentais nem contratuais, mas sim uma razão comunicativa fundada no cotidiano, sendo, pois, capaz de validar e legitimar determinadas normas morais a fim de estabelecer sua aceitabilidade racional. Assim, Habermas transpõe o conceito de razão para a linguagem, como um processo da vida no meio social. Além disso, sobre o "ponto de vista moral", ele também retira qualquer pretensão de um entendimento transcendental.

> *O "moral point of view" ("ponto de vista moral") não pode ser encontrado num "primeiro" princípio ou numa fundamentação "última", ou seja, fora do âmbito da própria argumentação. Apenas o processo discursivo do resgate de pretensões de validez normativas conserva uma força de justificação; e essa força, a argumentação deve-a em última instância ao seu enraizamento no agir comunicativo. O almejado "ponto de vista moral", anterior a todas as controvérsias, orienta-se de uma reciprocidade fundamental embutida no agir orientado para o entendimento mútuo* (1989, p. 197).

Para Habermas, a produção de normas não se reduz a algo transcendental que antecede ao homem nem a algo contratual como fundamento último e regulador do comportamento humano. Afinal, "o modelo do contrato é substituído por um modelo do discurso ou da deliberação: a comunidade jurídica não se constitui através de contrato social, mas na base de um entendimento obtido através do discurso" (1987, p. 309).

Desse modo, conclui-se que o "ponto de vista moral" da crítica habermasiana não está ancorado numa democracia soberana ou particularista nem numa ética contratualista e nem numa cidadania institucionalizada, mas sim:

> 1. No conceito de "democracia deliberativa" que se instaura no "agir comunicativo", no qual os indivíduos possuem autonomia no que concerne à regulação de sua vida social;
>
> 2. No conceito de "ética argumentativa", que considera tanto o indivíduo em sua coletividade quanto o indivíduo em sua subjetividade;
>
> 3. E, por fim, num conceito de cidadania que, fundada nos princípios da "democracia deliberativa", ressalta o consenso estabelecido entre os indivíduos em seu cotidiano.

Assim temos: o sujeito delibera seu próprio poder pelo uso de seu "agir comunicativo", o qual não pode se efetivar senão por meio das relações que ele mantém com outros sujeitos, mediante, vale lembrar, uma "ética argumentativa", que seja reguladora dessas relações.

A discussão dos conceitos habermasianos sobre pluralismo e universalismo e sua implicação nas políticas e práticas curriculares

Trazendo agora a discussão para o nosso contexto, em que o conceito de democracia está implícito em nossa cidadania e em nosso entendimento de sistema de governo legítimo e soberanamente representativo do nosso povo, podemos dizer que, não respeitar os direitos morais é não reconhecer o pluralismo cultural.

Nesse caso, a discussão em torno do pluralismo cultural não é somente um motivo para analisar possíveis descumprimentos das regras morais, ou seja, motivo para analisar se o pluralismo cultural confronta e sobrepõe-se aos direitos morais, mas sim motivo para reconhecer, sobretudo, que é a ausência e o não reconhecimento ao pluralismo que implica o descumprimento às regras morais, ou seja, que a falta de reconhecimento ao pluralismo é que deve ser analisado como algo que confronta e sobrepõe-se aos direitos morais universais.

Considerando, então, as observações de Habermas sobre a facticidade e a aceitabilidade racional, podemos pensar "democracia" de duas maneiras:

> 1. De acordo com sua natureza, os ideais democráticos, via de regra, isto é, necessariamente devem reconhecer e considerar a pluralidade cultural contida no interior de uma sociedade como a nossa, por exemplo;
>
> 2. Vale, contudo, pensar os limites do uso desse termo, ou seja, pensar em até que ponto a democracia por sua característica imanente não estaria investida (por alguns segmentos) da missão de sobrepor uma dada cultura (no caso, uma cultura hegemônica) em detrimento das demais formas de cultura (grupos "minoritários").

Nesse sentido, pela leitura de Habermas, podemos dizer que a imanência democrática que legitima toda e qualquer prática como reconhecidamente válida não implica necessariamente uma legitimação da validade normativa dessa mesma prática em sua aplicabilidade social. Disso se pode pensar então nos limites da democracia, pois nem toda

prática pode ser considerada uma norma legítima em face de possíveis arbitrariedades da forma objetiva pela qual essa prática se instrumentaliza na sociedade.

Desse modo, então, a cultura democrática precisa ser problematizada. O direito democrático de uma cultura hegemônica se sobrepor às demais, não pode esbarrar, sob pena de ser arbitrário, no direito, também democrático, de grupos não hegemônicos conquistarem sua representatividade.

Percebe-se, então, que o direito comporta um paradoxo. Com base em Habermas, podemos pensar em uma democracia a partir de uma *práxis* argumentativa com a necessidade, é claro, de uma análise quanto a sua facticidade e sua aceitabilidade racional, para que não se relativize extremamente o próprio conceito de "democracia".

Assim como o "ponto de vista moral", o processo de socialização também ocorre na própria discursividade pela socialização comunicativa que, apesar de contingente, não é ilógica. Esse processo chama atenção para os limites de nossa liberdade e direitos democráticos. Eis o que diz Berten:

> *O processo de socialização comunicativo é um processo histórico. Porém, precisa de um "ponto de vista" que permite distinguir "as condições que possibilitam a socialização comunicativa" (condições que embora historicamente contingentes desencadearam processos de desenvolvimento lógico) e as limitações ou restrições contingentes. É nesse sentido que, nas ações com os outros se faz a experiência não somente dos limites de minha liberdade (definição liberal da liberdade negativa), mas a experiência positiva de uma liberdade "social", quer dizer a descoberta de uma liberdade que se constitui através da socialização* (BERTEN, 2010, p. 14).

Nesse caso, não basta apenas o entendimento acerca dos limites de minha liberdade, mas o reconhecimento de que esta liberdade se constrói e só existe à medida que se refere às relações entre sujeitos por meio de processo de sociabilização.

Propositalmente, desviei o "olhar" de Habermas sobre a universalidade dos direitos morais para o cerne da concepção de "democracia". O pluralismo cultural, visto com desconfiança e posto em suspensão por Habermas quanto a sua validade e aceitabilidade (não por negar sua validade ontológica, mas por questionar sua validade normativa em relação à universalidade dos direitos morais), foi tratado aqui, na verdade, como "degrau" para a conquista dos direitos morais universais, à medida que considerou-se, neste trabalho, que o pluralismo cultural, examinados os seus limites, é um direito moral universal constituído no cotidiano por sujeitos agentes e construtores da realidade. Desse modo, voltamos à primeira frase deste texto que afirma que será considerada, neste trabalho, a pluralidade cultural como característica intrínseca ao conceito de sociedade democrática.

Trazendo a discussão para as manifestações pluralistas de nosso contexto social e espacial, considerei, portanto, não somente a desconfiança em relação ao pluralismo cultural e às ações possivelmente arbitrárias que poderiam ser cometidas em nome desse pluralismo, mas também, e, sobretudo, a desconfiança de uma moral que pretende ser hegemônica e que se aproveita da imanência dos ideais democráticos, a fim de apelar pelo direito a um universalismo totalizante que não reconhece as diferenças culturais existentes em nossa sociedade e que se utiliza do conceito de democracia para justificar uma democracia que homogeneíza e que, em nome de ideais igualitários, desconsidera particularidades, suprimindo a pluralidade, à medida que renega as diferenças.

Assim sendo, a intenção deste trabalho foi destacar que é a ausência de um pluralismo cultural que induz a um não reconhecimento dos princípios morais do indivíduo, ou dos grupos de indivíduos, de culturas não hegemônicas. Negar o pluralismo cultural, além de negar um direito comum a todos de se manifestarem democraticamente e terem representatividade igualitária na sociedade, é negar, ao mesmo tempo, a aceitabilidade racional do próprio ideal de democracia[2], ou seja, é descumprir o conjunto de regras válidas de uma sociedade democrática.

Contemplar o pluralismo cultural na sociedade ou nos documentos curriculares é, pois, garantir o cumprimento dos ideais democráticos. A

[2] Considerado aqui os limites do uso do termo, como proposto por Habermas.

pluralidade cultural é o caminho para a legitimação dos direitos morais universais – que estabelece o reconhecimento do outro como sujeito de direito – bem no estilo habermasiano do termo, ou seja, direitos morais universais resultantes de um escrutínio entre o que é passível de ser feito (facticidade) em uma sociedade democrática[3] e o que é passível de ser considerado como aceitavelmente válido, de acordo com as normas e princípios morais nessa mesma sociedade democrática.[4] Assim, além de ressaltar a relação entre uma perspectiva vertical (sujeito) e uma perspectiva horizontal (coletivo), e considerar que o universalismo não se contrapõe ao pluralismo, Habermas também nos lembra que a relação entre direitos humanos e soberania popular são complementares entre si e que é o exercício da soberania popular que garante os direitos humanos (1997, p. 259).

A soberania popular para Habermas é produzida nas práticas discursivas, nos debates, nas discussões, enfim, nos intercâmbios comunicacionais constantes do "agir comunicativo". Pensando então na soberania popular como última instância de uma conquista pelo reconhecimento à pluralidade e, por conseguinte, como resultado de uma vitória de manifestações culturais que lutam por esse reconhecimento, pode-se dizer que a conquista dessa soberania acontece no campo da discursividade, ou melhor, nos intercâmbios comunicacionais inseridos no espaço público de formação das opiniões, cujas decisões não só constituem o estado de direito, como também são capazes de influenciar desde a prática e formulação dos documentos curriculares até as práticas cotidianas em sua dinâmica social.

Referências

ARAÚJO, L. B. L. *Pluralismo e justiça: estudos sobre Habermas.* São Paulo: Ed. Loyola, 2010.

BALL, S. J. *Cidadania global, consumo e política educacional*. In: _____. *A escola cidadã no contexto da globalização*. Tradução: Tomaz Tadeu da Silva. Petrópolis, RJ: Vozes, 1998b, p. 121-137.

[3] Como, por exemplo, não impedir manifestações pluri-culturais

[4] Como, por exemplo, considerar como válido e aceito que uma sociedade democrática, necessariamente, pressupõe o reconhecimento ao pluralismo cultural.

BERTEN, A. *Por que Habermas não é e não pode ser contratualista*. In: Revista *Ensaios Filosóficos*, v. 1- abril/2010.

BOWE, R.; BALL, S.; GOLD, A. *Reforming education & changing schools:* case studies in policy sociology. London: Routledge, 1992.

HABERMAS, J. *Consciência moral e agir comunicativo*. Tradução: Guido Antônio de Almeida. Rio de Janeiro: Tempo Brasileiro, 1989.

_____. *Direito e democracia: entre facticidade e validade,* v. II. Tradução Flávio Beno Siebeneichler. Rio de Janeiro: Tempo Brasileiro, 1997.

HALL, S. *A questão da identidade cultural*. Textos didáticos. São Paulo: IFHC/Unicamp, 1998.

_____. *Quem precisa de identidade?* In: _____. *Identidade e diferença***.** Tradução: Tomaz Tadeu da Silva (org.). Petrópolis: Vozes, 2000. cap. 3, p. 103.

_____ .*A identidade cultural na pós-modernidade.* Rio de Janeiro: DP&A, 2003.

LOPES, ALICE R. C. *et al. Currículo: políticas e práticas.* Antônio Flávio Barbosa Moreira (org.). Campinas: Papirus, 1999.

_____ *Currículo e epistemologia.* Ijuí: Unijuí, 2007.

Currículo e Diversidade Cultural: uma abordagem a partir do Ensino Religioso nas escolas públicas

Alberes de Siqueira Cavalcanti
Instituto Federal de Educação, Ciência e Tecnologia do Maranhão

Resumo

O artigo tem por objetivo analisar a relação entre currículo e diversidade cultural especificamente naquilo que concerne à inserção da disciplina Ensino Religioso nas escolas públicas. Apresento algumas análises conceituais sobre o tema currículo, uma síntese do seu percurso histórico, evidenciando a relação entre currículo e política, mostrando a não neutralidade do currículo. Em conclusão, apresento uma reflexão sobre a diversidade religiosa e defendo o princípio da laicidade da escola pública.

Palavras-chave: ensino religioso, currículo, diversidade cultural, escola pública

Abstract

This article analyses the relation between curriculum and cultural diversity specifically aiming at religious teaching in public schools in Brazil. I approach school curriculum in its conceptual and historical perspectives highlighting its link with politics what shows that neutrality is not an accomplishment in curriculum choices. In conclusion, it is presented a reflection on religious diversity and my defense of a secular public school.

***Keywords**: religious teaching, curriculum, cultural diversity, religious diversity, public school*

Introdução

A relação entre educação formal e cultura está na ordem do discurso educacional contemporâneo. Em grande parte, isso se deve ao contexto pós-colonial e ao processo de globalização que trouxe para o cenário internacional a questão da identidade cultural dos povos e dos grupos minoritários que migraram dos chamados países periféricos e

formaram colônias nos países desenvolvidos. Assim, a questão do respeito à diversidade cultural passou a ser discutida por educadores do Norte e do Sul. Expressões como multicultural, pluricultural, transcultural passaram a fazer parte das discussões que envolvem a política educacional e cultural, envolvendo assim as discussões sobre currículo.

No Brasil, um país caracterizado pela pluralidade cultural, a reivindicação de uma educação inclusiva que considere a diversidade regional, racial, étnica e religiosa da sua população também se faz sentir. Os Parâmetros Curriculares Nacionais (PCN), com toda a polêmica que o envolve, além de propor o pluralismo cultural como um dos temas transversais, o define como um dos objetivos gerais do ensino fundamental.

Um dos elementos da cultura brasileira que mais sofreu mudanças nas últimas décadas foi o campo religioso. De uma sociedade tida como catolicamente hegemônica e homogênea, passamos para uma sociedade religiosamente complexa e diversificada, com o considerável aumento de instituições evangélicas das mais variadas matizes, a efervescência de movimentos religiosos de origem oriental, esotéricos, além da resistência e valorização das religiões afro-brasileiras.

Currículo: uma aproximação teórica

No âmbito educacional, o currículo constitui-se num campo complexo cujos limites conceituais são bastante alargados. Ribeiro, na busca de definir o currículo, faz a distinção entre as "acepções comuns" e "concepções típicas". Nas primeiras, o currículo é definido como "o elenco e sequência de matérias ou disciplinas para todo o sistema escolar" (RIBEIRO, 1999, p. 11), confundindo-se com "plano de estudos"; ou como "programas de ensino num determinado nível ou área de estudo do sistema escolar" (p. 12); ou ainda a junção dessas acepções, "identificando o currículo com o conjunto estruturado de matérias e de programas de ensino num determinado nível de escolaridade, ciclo ou domínio de estudos" (p. 12). Nas "concepções típicas", o autor destaca as definições de currículo como "conjunto de experiências educativas vividas pelos alunos, sob a tutela da escola" (p. 13); "a noção de currículo como plano e organização do ensino-aprendizagem" (p. 15).

A partir dessas várias nuances do currículo, é possível perceber que ele envolve questões de ordem teórica e prática, referentes à educação formal e que dizem respeito ao processo de ensino-apredizagem, ao conhecimento escolar, à vivência da escolarização. As "concepções típicas", no entanto, já oferecem uma distinção entre o que se vivencia e o que se planeja em termos de currículo. Alguns autores chamam a atenção para esta distinção. Pacheco, por exemplo, define o currículo como "um projeto, cujo processo de construção e desenvolvimento é interactivo, que implica unidade, continuidade e interdependência entre o que se decide ao nível do plano normativo, ou oficial, e ao nível do plano real, ou do processo de ensino-aprendizagem" (1996, p. 20). Já Sacristán, enfatiza que o "currículo tem que ser entendido como a cultura real que surge de uma série de processos, mais que como um objeto delimitado e estático que se pode planejar e depois implantar" (1995, p. 84).

Santos (1998), considerando o currículo como "um artefato social e cultural", diferencia o currículo formal (escrito em forma de propostas ou guias curriculares) do currículo real (todo material que permita o estudo da prática de uma determinada matéria ou disciplina escolar). Segundo essa autora, o currículo escrito mostra os interesses e as influências que atuam no nível das formulações das políticas educacionais e estabelecem parâmetros para a realização da prática pedagógica; já a realização do currículo em sala de aula pode ser analisada por meio da história de vida de professores. De tal forma, pode haver discrepância entre as propostas curriculares oficiais e aquilo que realmente é ensinado na escola.

Nesse ponto, chegamos então a estabelecer a relação entre currículo e política educacional e cultural. O currículo não pode ingenuamente ser visto como um documento estático, desinteressadamente estabelecido, como se a escola e o que nela se faz, se vivencia e se aprende ou desaprende não possuísse relação com o seu entorno social, político, econômico e cultural. Disso resulta a importância do currículo para qualquer sistema educacional, uma vez que alguém define formal e/ou realmente o que se ensina ou não, para quem se ensina ou não, como se ensina ou não, quando se ensina ou não. Obviamente que esse alguém que define o currículo possui uma visão de mundo, de sociedade, de homem e de educação que se quer fomentar no processo de ensino-aprendizagem.

A organização do conhecimento escolar, portanto, não obedece unicamente ao interesse estritamente educacional. Santos (1998), analisando a história das disciplinas escolares, observa o processo de mudanças curriculares, fazendo-nos compreender que as disciplinas escolares são apenas aparentemente estáveis, mas que historicamente elas sofrem mudanças. Tais mudanças dependem de fatores internos e externos, de eventos políticos e sociais que tornam plausíveis ou implausíveis certas ideias já existentes em um campo do currículo, do papel da psicologia na construção do currículo e da instrução, pela influência crescente do livro didático, vinculado às empresas editoriais, e das mudanças de orientação nos vestibulares.

Assim, a organização curricular do conhecimento escolar em disciplinas não é algo historicamente imutável e ideologicamente neutro. A própria organização disciplinar do conhecimento escolar, a inserção ou não de uma disciplina no currículo, os conteúdos e objetivos de uma determinada disciplina, as mudanças na abordagem e no tratamento metodológico de uma disciplina, tudo isso obedece a interesses internos e externos à própria escola relacionados a questões epistemológicas, ideológicas, políticas e sociais.

Currículo e política: o caso do ensino religioso

Essa não neutralidade do currículo, o que implica a introdução ou não de determinadas disciplinas e na seleção ou não de determinados conteúdos, é perfeitamente constatada quando analisamos a inserção da disciplina Ensino Religioso no currículo da escola pública brasileira. Desde que a república instituiu a separação Igreja-Estado, a questão central até agora posta quanto a essa disciplina diz respeito ao princípio da laicidade. "A laicidade, ao condizer com a liberdade de expressão, de consciência e de culto, não pode conviver com um Estado portador de uma confissão" (CURY, 2004, p. 182). Nesse sentido, a escola pública, como esfera estatal, não pode comportar o elemento religioso ou confessional, sob a pena de infringir o dispositivo constitucional que proíbe o Estado de subvencionar qualquer atividade de natureza religiosa, de acordo com o artigo 19 da Constituição Federal de 1988.

A história educacional brasileira testemunha que, desde o seu nascedouro com a vinda dos jesuítas, os interesses políticos e religiosos confluíram numa política educacional voltada para a catequese e o

desenvolvimento das "escolas de primeiras letras". Mesmo com a independência, no período imperial, a relação Igreja-Estado continuou. A monarquia brasileira, por meio da Constituição Imperial de 1824, consagrou o catolicismo como a religião oficial do Estado, e a presença da catequese católica na educação pública foi apenas uma consequência lógica da determinação constitucional. Durante todo o período que vai da Colônia ao Império, a política foi regida pelo princípio absolutista que apregoava a religião de acordo com o rei (*cujus régio, ejus religio*). De tal forma, também a política educacional e o currículo obedeceram a esse princípio, definindo uma educação pública eminentemente religiosa.

No período republicano, com a adoção do princípio da laicidade que institui a separação entre Igreja e Estado, a inserção do Ensino Religioso no currículo da escola pública funcionou como uma espécie de termômetro para medir o grau de intensidade na relação Igreja-Estado, ao mesmo tempo em que foi usada como moeda de troca nas negociações entre essas instituições. Sua inclusão, portanto, deveu-se mais à lógica dos interesses políticos e ideológicos da relação Igreja-Estado, situando na linha tênue que separa o público e o privado, do que propriamente aos interesses educacionais. O marco nessa disputa política e ideológica em torno do Ensino Religioso foi o "Manifesto dos Pioneiros da Educação Nova", em 1932, em que educadores se posicionam a favor da escola pública laica e contrários ao Ensino Religioso.

Essa polêmica perdurou na história republicana nacional, sendo reeditada na Constituição de 1988 e na LDBEN 9.394/96, em que de um lado estavam os defensores do laicismo e do outro os defensores do Ensino Religioso. Sob a forte mobilização e pressão de entidades religiosas que organizaram um movimento pró-Ensino Religioso na Assembleia Constituinte, manteve-se o Ensino Religioso na atual Constituição como "disciplina dos horários normais das escolas públicas de ensino fundamental" (Art. 210. § 1º), com matrícula facultativa.

A LDBEN 9.394/96, no artigo 33, trazia de um lado a conquista dos defensores do Ensino Religioso e de outro a conquista dos seus adversários. Retomando o artigo 210 da Constituição de 1988, o texto garantia essa disciplina de matrícula facultativa também em horários normais das escolas públicas de ensino fundamental. Estabelecia também a possibilidade de o Ensino Religioso ser confessional, de acordo com a opção religiosa do aluno ou do seu responsável, ministrado por

professores ou orientadores religiosos preparados e credenciados pelas respectivas igrejas ou entidades religiosas; ou ainda interconfessional, resultante de acordo entre as diversas entidades religiosas, que ficariam responsáveis pela elaboração do programa de ensino da disciplina.

Apesar dessa considerável vitória dos defensores do Ensino Religioso, a primeira redação do artigo 33 trazia-lhes um entrave não menos respeitável: a expressão "sem ônus para os cofres públicos" (o que desobrigava o Estado a destinar recursos humanos e financeiros na efetivação dessa disciplina). A reação dos defensores do Ensino Religioso foi rápida e eficaz. Em apenas sete meses, contando com a anuência de representantes do governo federal e da oposição, conseguiram que o artigo 33 fosse alterado. "Sob a liderança do deputado Padre Roque (PT/PR), com apoio do MEC e de empresários da educação, aprovou-se a lei nº 9.475/97, que alterou o art. 33" (PAULY, 2004, p. 27).

Do ponto de vista histórico e pedagógico, essa disciplina perdurou nesse processo basicamente como um elemento eclesial dentro da escola, na forma de catequese ou aulas de religião, com seu conteúdo e tratamento metodológico voltado para a difusão da doutrina cristã, especialmente na sua versão católica. Assim, podemos constatar com o exemplo da disciplina Ensino Religioso o que nos alerta Apple:

> *O currículo nunca é apenas um conjunto neutro de conhecimentos, que de algum modo aparece nos textos e nas salas de aula de uma nação. Ele é sempre parte de uma tradição seletiva, resultado da seleção de alguém, da visão de algum grupo acerca do que seja conhecimento legítimo. É produto das tensões, conflitos e concessões culturais, políticas e econômicas que organizam e desorganizam um povo* (APPLE, 1995, p. 59).

De tal forma, observamos que o currículo se insere no jogo de poder, no exercício político dos sujeitos sociais. No caso brasileiro, as instituições religiosas, marcadamente as católicas, e setores governamentais e do empresariado do ensino foram mais hábeis politicamente em articular suas forças e manter uma tradição já existente no currículo nacional: a inserção da disciplina Ensino Religioso.

Currículo e Diversidade Cultural Religiosa

A religião é uma das mais complexas manifestações culturais. Ao mesmo tempo em que se constitui um fenômeno universal, se constitui concretamente numa forma particular, evidenciando uma modalidade de diversidade cultural. Nesse sentido, também a questão religiosa torna-se um problema cultural no espaço da educação formal, ou seja, um tema a ser enfrentado por uma educação que pretenda levar em consideração a diversidade cultural. E, nesse caso, o currículo passa a enfrentar a questão posta pelo multiculturalismo. É o que nos faz perceber Sacristán ao afirmar que também "[...] há problemas de multiculturalismo quando, em um sistema educacional, ou mesmo em uma escola, confluem populações com religiões diferentes ou línguas diferentes" (1995, p. 94). O caso maranhense, especialmente ludovicense, evidencia exemplarmente a diversidade cultural religiosa brasileira que se reflete na escola pública.

O fato de existir uma disciplina no currículo escolar direcionada para o fenômeno religioso põe sem dúvida para o currículo a questão do multiculturalismo, pois a possibilidade de surgir preconceitos contra alunos e alunas de denominações ou movimentos religiosos minoritários ou social e historicamente postos à margem é concreta. A escola pode vir a ser um campo de exclusão a partir da perspectiva religiosa, determinando o que é legítimo ou não com relação a conteúdos, práticas, crenças e valores a ser ensinados na disciplina em questão.

> *Para os que defendem o multiculturalismo, a escola trabalha apenas com uma parcela restrita da experiência humana, ou seja, com os saberes, valores e atitudes que fazem parte do que se denomina versão autorizada ou legítima da cultura. Nesse processo, a cultura dos diversos grupos sociais fica marginalizada do processo de escolarização e, mais que isso, essa cultura é vista como algo a ser eliminado pela escola, devendo ser substituída pela versão autorizada da cultura, o que tem estado, geralmente, presente em todas as esferas do sistema escolar* (SANTOS, 2001, p. 8).

O fenômeno religioso é um fenômeno antropológico e como tal cultural. Como parte da cultura humana universal e de grupos e povos

em particular, é desejável que seja estudado e conhecido pelas gerações de alunos e alunas que frequentam a escola pública. Dada à sua importância, a religião pode fazer parte do currículo da escola pública, mas como fenômeno não como crença, espiritualidade, teologia ou doutrina, pois são aspectos que fogem da alçada do Estado laico, sendo da competência de cada instituição ou movimento religioso em particular.

Portanto, somente respeitando a laicidade da escola pública, tornando as práticas e os conteúdos do Ensino Religioso e dos ensinos não religiosos (no sentido de não ser doutrinário, confessional, ou interconfessional), mas secularizados (no sentido de garantir a laicidade e a cientificidade do conhecimento escolar), parece ser possível uma disciplina na escola pública que dê conta da dimensão simbólica do ser humano, tantas vezes descuidada pela educação formal.

A nova redação do art. 33 da Lei de Diretrizes e Bases da Educação Nacional tenta resolver a questão da laicidade garantindo matrícula facultativa, "assegurado o respeito à diversidade cultural religiosa do Brasil" e proibindo "quaisquer formas de proselitismo", além da propositura de que se estabeleça uma "entidade civil, constituída pelas diferentes denominações religiosas, para a definição dos conteúdos do ensino religioso". É lógico que há uma distância entre esse Ensino Religioso não proselitista e respeitoso da diversidade cultural religiosa e aquele catequético, claramente confessional e a serviço de uma única instituição religiosa.

Apesar de não muito bem resolvida a questão da laicidade, a nova lei possibilita um novo foco para a polêmica em torno do Ensino Religioso. Ao considerar essa disciplina como parte da formação do cidadão, vetar qualquer forma de proselitismo, sobretudo ao subtrair a orientação antes dada acerca da confessionalidade e interconfecionalidade, abre o caminho para se pensar o Ensino Religioso do ponto de vista secular. Há que se reconhecer a importância pessoal e sociocultural da religião que, como a linguagem e a arte, constitui-se uma das expressões universais da cultura e caracterizadora da humanidade. Contudo, a escola pública não é o lugar apropriado para tratar da religião de forma religiosa, seja ela confessional ou interconfessional.

O passado de atrelamento da disciplina Ensino Religioso aos interesses da Igreja Católica imprimiu-lhe, no imaginário educacional, um caráter eclesial e confessional, estranho à escola pública. Contudo, deve-se ressaltar que "a lei nº 9.475 acabou com a possibilidade de as igrejas e religiões controlarem o ensino religioso na escola pública [...] elas perderam o controle sobre currículo, formação e seleção do corpo docente de ensino religioso" (PAULY, 2004, p. 181). Assim, pelo menos do ponto de vista legal, não se justifica uma espécie de fidelidade ou compromisso da escola pública e dos seus professores a uma determinada confissão religiosa, ou mesmo que o Ensino Religioso deva ser uma disciplina de natureza religiosa. Tratar de religião na escola pública, não significa praticar religião, adotar práticas religiosas ou induzir os alunos e as alunas a isso, eles têm todo o direito – inclusive religiosamente falando, o livre arbítrio – de rejeitar qualquer iniciativa de caráter religioso nesse tipo de escola.

Há que se observar que a questão da diversidade cultural religiosa em nosso país relaciona-se ainda com questões raciais e socioeconômicas, o que a torna mais complexa ainda. Não é razoável pensarmos como se vivêssemos numa espécie de sociedade multicultural harmônica, encobrindo a desigualdade com a diversidade. "O tema do multiculturalismo não pode ser separado das condições sociais e econômicas concretas de cada sociedade" (SACRISTÁN, 1995, p. 93). De tal forma, um currículo que pretenda inserir a diversidade cultural religiosa não descuidará do aspecto crítico, buscando uma proposta emancipatória e transformadora das relações de dominação e exclusão que se estabelecem na sociedade e na escola.

Parâmetros Curriculares Nacionais do Ensino Religioso

O Fórum Nacional Permanente do Ensino Religioso (FONAPER), criado em 1995 e composto por representantes de várias tradições religiosas, teve um forte papel nas discussões sobre o Ensino Religioso na LDBEN 9.394/96 e foi um dos responsáveis pela alteração do artigo 33 da referida lei. O FONAPER elaborou, em outubro de 1996, os Parâmetros Curriculares Nacionais do Ensino Religioso (PCNER), apresentado ao MEC, e que serviu de orientação para a nova redação do artigo 33. Lançado oficialmente em agosto de 1997, os PCNER constituem hoje uma referência para o Ensino Religioso, sendo utilizados por muitos educadores em todo o país. É o que ocorre, por exemplo, com

a proposta curricular para o Ensino Religioso, elaborada pela Secretaria de Educação do Estado do Maranhão, com pouquíssimas alterações de redação que espelham os PCNER. No caso da rede municipal de educação de São Luís, a proposta curricular para o Ensino Religioso ainda não foi oficialmente aprovada, mas a sua versão preliminar já finalizada também reflete a influência dos PCNER.

A iniciativa do FONAPER é importante, por tratar-se de uma proposta que pretende respeitar a diversidade cultural religiosa, evitando o proselitismo e a doutrinação. É um avanço, por ser elaborado por representantes de várias tradições religiosas, por não ser catequético ou doutrinário. Mas os PCNER ainda permanecem ligados a um tratamento religioso para o Ensino Religioso, sem entrar no mérito da laicidade do ensino público.

A proposta dos PCNER é, de forma articulada, a primeira que define uma identidade para a disciplina Ensino Religioso como área de conhecimento própria e pedagogicamente bem definida. Inova e distancia-se bastante de uma visão catequética ou estreitamente confessional dessa matéria, abrindo-lhe a possibilidade de ter um tratamento pedagógico em alicerce filosófico e científico, sobretudo quando define entre os objetivos dessa disciplina: "proporcionar o conhecimento dos elementos que compõem o fenômeno religioso, a partir das experiências religiosas percebidas no contexto do educando"; e "analisar o papel das tradições religiosas na estrutura e manutenção das diferentes culturas e manifestações socioculturais".

No entanto, esse avanço não pode ofuscar alguns problemas de ordem epistemológica e ideológica que perpassam os PCNER e comprometem a viabilidade do Ensino Religioso efetivamente como uma disciplina da escola pública. Considerando de antemão que "na raiz de toda criação cultural está a Transcendência" (FONAPER, 2001, p. 20) e que o objeto da disciplina é o Transcendente, a proposta dos PCNER inscreve-se numa perspectiva religiosa na abordagem do Ensino Religioso, aceitando "tacitamente o dogma religioso do *inatismo*, segundo o qual a transcendência seria uma 'capacidade inerente ao ser'" (PAULY, 2004, p. 179).

Segundo os PCNER, o Ensino Religioso "não deve ser entendido como Ensino de uma Religião ou das Religiões na escola, mas sim uma

disciplina centrada na antropologia religiosa" (FONAPER, 2001, p. 11). Sendo assim, deve-se considerar que, do ponto de vista antropológico e sociológico, não é razoável aceitar nenhuma dimensão humana que seja *a priori* natural, como uma natureza humana já estabelecida. As afirmações acerca do ser humano – como *homo sapiens*, *homo loques*, *homo faber*, etc. – devem ser situadas histórica e culturalmente. Assim também o *homo religiosus* observa essa inscrição no terreno da história e da cultura, ou seja, a dimensão religiosa não é inata, não se nasce religioso, não se tem uma natureza religiosa inerente ao ser humano. Como todas as outras dimensões, a religiosa também é apreendida no processo de socialização no qual nos humanizamos.

Quando os PCNER defendem que o educando deva fazer "conscientemente, **a passagem do psicossocial para a metafísica/Transcendência**, a partir do que assimila **na Escola**" (FONAPER, 2001, p. 46, grifos nossos)[1], nada impede que isso seja uma forma de indução para uma determinada visão da Transcendência e, mesmo na hipótese de que isso não ocorra, é evidente que não é da competência da escola pública ensinar o caminho das pedras para a experiência ou o conhecimento da Transcendência/Deus. Não se pode, portanto, partir desse pressuposto epistemológico e, de alguma forma, levar os alunos e as alunas a uma visão ou experiência religiosa presumivelmente mais lúcida, explícita e religiosamente mais elevada da Transcendência.

Essa centralidade que as categorias "Transcendência" e o "Transcendente" assumem nos PCNER, perpassando todo o documento, reconduz o Ensino Religioso para uma abordagem religiosa e dificulta o seu fazer pedagógico na escola pública, tanto do ponto de vista dos conteúdos propostos como do tratamento didático. Tal dificuldade fica evidenciada, por exemplo, quando se orienta a sensibilizar o educando para o "mistério" (p. 46, 49, 52); o "sobrenatural" (p. 49, 52); o exercício do "silêncio interior" (p. 46); a compreensão do "eu interior de cada pessoa em relação com o Transcendente" (p. 52); o "educar para o sentido profundo da experiência mítica na autoridade do discurso religioso" (p. 52) e o "entendimento da espiritualidade que cultiva a

[1] Conferir também as páginas 27, 49, 52, 55 dos Parâmetros, em que essa ideia da passagem para a Transcendência também é expressa.

vivência do mistério" (p. 55). Percebe-se claramente o viés espiritualizante e místico presente na proposta dos PCNER.

O Transcendente é outra palavra para Deus, muito próximo afinal da concepção monoteísta judaico-cristã. E aqui entra uma questão de ordem ideológica referente à diversidade cultural religiosa: como garantir que, ao tratar do Transcendente e não do fenômeno religioso propriamente dito, uma determinada visão do Transcendente não se imponha? Mesmo dentro de uma determinada religião pode haver diferentes visões acerca do Transcendente e essas visões influenciam a visão de mundo das pessoas e grupos sociais, repercutindo na ética, na política, nas relações de gênero, de etnia, de poder, etc. Assim, ao se propor trabalhar com o Transcendente no chão da escola, não se pode ingenuamente abstrair essa ideia como sendo universal e neutra, partindo tão somente da crença de que afinal "Deus" é o mesmo para todos.

Como adverte Sacristán: "A diversidade é possível apenas quando existe variedade" (1995, p. 84). Mas como existir variedade se a prática pedagógica estiver orientada por uma ideia homogeneizada do Transcendente? Na prática pedagógica, no currículo real, um determinado "Deus" – geralmente aquele do professor ou dos livros didáticos – será glorificado e outros "deuses" serão expulsos do Olimpo escolar. O princípio da laicidade deve ser observado também com relação à garantia do respeito à diversidade cultural religiosa. Se, realmente, o Ensino Religioso não for tratado como uma área de conhecimento desvinculado dos interesses particulares das instituições religiosas, com um objeto de estudo epistemologicamente fundamentado, a propositura da diversidade cultural religiosa não passará de um discurso vazio. E, nesse aspecto, a proposta do PCNER torna-se ambígua[2]. Por isso, há que se buscar o objeto do Ensino Religioso no próprio fenômeno religioso e não no Transcendente, o que seria adotar *a priori* uma determinada crença ou postura religiosa.

[2] Num outro texto, o FONAPER é bastante claro ao definir que: "A disciplina Ensino Religioso tem como objeto de estudo o fenômeno religioso" (FONAPER, p. 21). Nos PCNER, define-se "como objeto o Transcendente" (FONAPER, 2001, p. 5). Nos dois textos do FONAPER, porém, evidencia-se a centralidade do Transcendente para a disciplina Ensino Religioso. Pauly (2004, p. 174) chama a atenção para a persistência de um "dilema epistemológico", caso se tome o Transcendente como objeto da disciplina Ensino Religioso.

Outro aspecto com relação aos PCNER deve ser refletido. Uma questão levantada por Apple num outro contexto geopolítico[3] poderia também nos inquietar: "faz sentido a idéia de um currículo nacional?" (1995, p. 59). Esse questionamento pode nos inspirar outro ponto relativo ao nosso objeto: em que medida faz sentido uma proposta nacional para o Ensino Religioso? Uma vez que os PCNER foram tomados como referenciais para a elaboração das propostas curriculares locais, das secretarias estaduais e municipais de educação, há que se perguntar até que ponto esses parâmetros nacionais estão sendo adotados, levando em consideração a realidade social, cultural e religiosa local.

Conclusão

Ao final dessa análise, alguns aspectos da nossa reflexão merecem ser reiterados. O primeiro deles diz respeito à compreensão da não neutralidade do currículo, o que equivale a situá-lo no jogo e nas relações de poder que estão além dos muros da escola, mas que confluem para a sala de aula, onde o professor, as disciplinas, seus conteúdos e práticas desempenham um papel importante. A perspectiva histórica que buscamos apresentar da disciplina Ensino Religioso evidencia, por certo, essa nuance política do currículo.

Outro aspecto a ser salientado é a distinção estabelecida por vários teóricos entre o currículo formal e o currículo real, chamando a atenção para a importância do segundo. É no chão da escola, na prática pedagógica de ensino-aprendizagem, na relação professor-aluno, nos processos de avaliação, na utilização dos recursos didáticos que o currículo se efetiva. À medida que o real é vivo, é movimento, certamente a cena que se passa na escola e na sala de aula não é a mesma prevista no roteiro. No caso do Ensino Religioso, não são muitas as garantias concretas para que seja "assegurado o respeito à diversidade cultural religiosa" e proibindo "quaisquer formas de proselitismo", a não ser que se elimine qualquer postura religiosa no trato da disciplina.

Há que se considerar que a aceitação da diversidade cultural religiosa em nosso país é um fenômeno bastante recente. O mesmo ocorre com relação à construção de um novo paradigma para o Ensino

[3] Apple no artigo "A política do conhecimento oficial: faz sentido a idéia de um currículo nacional?", analisa a proposta de um currículo nacional nos Estados Unidos.

Religioso. Na prática, estamos ainda muito distantes da nova proposta curricular para o Ensino Religioso, de tal forma que podemos constatar que, assim como na sociedade, também na escola pode ocorrer violência simbólica contra alunos e alunas em nome da religião.

A questão da diversidade cultural religiosa no sistema escolar, como vimos, passa pelas opções político-ideológicas na definição do currículo formal, que desempenha o papel de nortear a prática pedagógica no chão da escola. No entanto, há que se considerar que o currículo real nunca é idêntico ao currículo formal. É na escola que, de fato, se produz o conhecimento escolar e se processa o ensino-aprendizagem. Nesse sentido, professores e professoras têm um papel decisivo na efetivação do currículo, por meio da escolha do livro didático ou não, da seleção e organização dos conteúdos, dos recursos técnicos e métodos adotados, da relação que mantêm com alunos e alunas, pela definição político-ideológica que adotam em sala de aula.

Portanto, a mudança de paradigma que se pretende na disciplina Ensino Religioso tem muito a ver com a formação e a prática do professor e da professora dessa disciplina. E, sob esse aspecto, é possível questionar até que ponto alguns professores e professoras de Ensino Religioso estão dispostos, à maneira fenomenológica, de pôr entre "parênteses" as suas verdades, certezas e seguranças religiosas em respeito a uma escola pública laica e culturalmente marcada pela diversidade religiosa.

Referências

APPLE, Michael W. *A política do conhecimento oficial: faz sentido a ideia de um currículo nacional?* In: MOREIRA, F. B.; SILVA, T. T. da. *Currículo, cultura e sociedade*. São Paulo: Cortez, 1995. p. 59-87.

CURY, Carlos R. Jamil. *Ensino religioso na escola pública: o retorno de uma polêmica recorrente.* In: *Revista Brasileira de Educação*. São Paulo: ANPEd, set/out/nov/dez, 2004, n. 27. p. 183-191. Acesso em: http://www.anped.org.br/rbe27/anped-n27-art12.pdf.

FÓRUM NACIONAL PERMANENTE DO ENSINO RELIGIOSO. *Parâmetros curriculares nacionais do ensino religioso*. 4. ed. São Paulo: Ave Maria,

2001.

_____. *Ensino Religioso: capacitação para um novo milênio*. [1998?]. In: *Caderno de Estudos Integrante do Curso de Extensão – a distância – de Ensino Religioso.* (Caderno 1, *O Ensino Religioso é disciplina integrante da formação básica do cidadão*).

_____. *Ensino Religioso: capacitação para um novo milênio*. [1998?]. *Caderno de Estudos Integrante do Curso de Extensão – a distância – de Ensino Religioso*. (Caderno 2, *O Ensino Religioso na diversidade cultural-religiosa do Brasil*).

PACHECO, José Augusto. *Currículo*: *teoria e práxis*. Porto: Porto Editora, 1996.

PAULY, Evaldo Luis. *O dilema epistemológico do ensino religioso*. In: *Revista Brasileira de Educação*. São Paulo: ANPEd, set/out/nov/dez, 2004, n. 27. p. 172-182. Acesso em: http://www.anped.org.br/rbe27/anped-n27-art11.pdf.

RIBEIRO, Antonio Carrilho. *Desenvolvimento curricular*. Coleção Educação Hoje. Lisboa: Texto Editora, 1999.

SACRISTÁN, J. Gimeno. *Currículo e diversidade cultural*. In: SILVA, Tomaz T. da; MOREIRA, Antonio F. (orgs.). *Territórios contestados*: o currículo e os novos mapas políticos e culturais. Petrópolis: Vozes, 1995. p. 82-113.

SANTOS, Lucíola L. de C. P. *História das disciplinas escolares: outras perspectivas de análise*. In: *Revista Educação e Realidade*, jul./dez. 1998.

_____. *O Ensino Religioso no currículo escola*. In: *Diálogo*: *Revista de Ensino Religioso*. São Paulo: Paulinas, maio, 2001. p. 5-11.

Resenha

Jaulas vazias: encarando o desafio dos direitos dos animais

Tom Regan

Resenha de Gabriel Garmendia da Trindade e
Lauren de Lacerda Nunes
Universidade Federal de Santa Maria

Embora possua mais de 40 anos, a discussão contemporânea acerca de qual deveria ser o tratamento correto outorgado aos animais não humanos ainda permanece estranha à filosofia como um todo. Poucos são os autores que se arriscam a lidar com uma temática marginalizada dentro do *mainstream* filosófico. Igualmente pequeno é o número de obras sobre tal assunto que se destaca e alcança significativo prestígio no cenário acadêmico. Tom Regan e seu *Jaulas vazias* encontram-se em ambas as posições.

Regan é um notável ativista e teórico do movimento em prol dos animais. Professor Emérito de Filosofia pela *North Carolina State University*, possui dezenas de artigos e resenhas publicadas em revistas especializadas, além de diversos livros, dentre os quais destacam-se *The case for animal rights* (1985), *Animal sacrifices: religious perspectives on the use of animals in science* (1986), *Animal rights and human obligations* (1989), organizado conjuntamente com o bioeticista australiano Peter Singer, e *The animal rights debate* (2001), escrito em coautoria com o filósofo Carl Cohen. Seus escritos enfatizando uma abordagem centrada na concessão de direitos aos não humanos influenciaram enormemente a referida discussão tanto no âmbito filosófico, quanto no jurídico e popular, reconstituindo visivelmente muitos dos princípios basilares sustentados pelos defensores dos animais.

Jaulas vazias está dividido em cinco segmentos (cada um com seus respectivos subcapítulos), prólogo/epílogo e dois prefácios redigidos por Regan – um desses exclusivo à edição brasileira. Na seção propedêutica, o autor inicia uma análise acerca de quem são exatamente os defensores dos animais. A segunda parte da obra apresenta uma reinserção ético-filosófica do debate sobre os direitos humanos e sua relevância para a atribuição de direitos básicos aos animais. Posteriormente, Regan examina o que vem sendo feito em termos de proteção e cuidado aos não humanos, como é o caso das propostas de bem-estar animal e tratamento "humanitário". Regan utiliza a quarta seção para detalhar

minuciosamente os diferentes usos dados aos não humanos pelos seres humanos, seja para o consumo, divertimento, entretenimento e/ou experimentação. Subsequentemente, o filósofo norte-americano faz o fechamento de seu texto com uma série de apontamentos acerca do que pode e deveria ser feito em nome dos animais pelo movimento em seu resguardo.

Antes de adentrarmos em uma exposição mais aprofundada dos diferentes conteúdos e facetas do livro em voga, faz-se necessário ressaltar os importantes aspectos literários presentes na atual tradução do texto reganiano. Como matéria de fato, é preciso frisar que essa é a primeira publicação em português brasileiro de uma obra completa de Regan. O trabalho de tradução foi realizado por Regina Rheda, autora e ativista pelos direitos dos animais. A revisão dos escritos ficou a cargo das professoras Rita L. Paixão e Sônia T. Felipe, conhecidas pesquisadoras com inúmeras publicações nas áreas de bioética e ética animal. Nesse sentido, *Jaulas vazias* deve ser apreciado não apenas pela sua temática estimulante, mas também pelo rigor e qualidade técnicos que compõem a versão final do manuscrito.

Primeiramente, de acordo com Regan, os defensores dos animais se dividem em três grupos que, embora possuam origens distintas, muito comumente têm em vista os mesmos objetivos. Os *vincianos* são indivíduos que mantêm um intenso vínculo empático com os não humanos desde os seus primeiros anos de vida. São pessoas capazes de facilmente se colocar no lugar dos animais e literalmente partilhar de suas alegrias ou mazelas. O segundo conjunto diz respeito aos *damascenos*, sujeitos que, por alguma razão, tiveram sua percepção acerca dos não humanos drasticamente alterada. Pessoas que presenciaram a imposição de genuíno sofrimento maciço a animais, terminando por repensar seus padrões e perspectivas pessoais, correspondem aos membros desse grupo.

Por sua vez, os *relutantes* representam o último elemento a compor a tríade de protetores sugerida por Regan. São aqueles que, no decorrer de sua vida, por meio de variadas experiências (não necessariamente traumáticas), modificaram seu comportamento e ações para com os não humanos, reconhecendo enfim a *consciência animal*. Muito embora os três grupos mencionados por Regan não abarquem a real dimensão e multiplicidade de origens dos integrantes do movimento pelos direitos dos animais, tal divisão é capaz de denotar eficientemente indivíduos

que possuem uma relação ética diferenciada com membros de espécies distintas.

Na seção seguinte, o autor examina a questão da natureza primeira dos *direitos morais*, sua função, bem como as razões para sua extensão tanto a humanos quanto a não humanos. Assim, segundo Regan, os direitos morais devem ser entendidos como barreiras protetivas, as quais têm o propósito de coibir a desconsideração de interesses, criando um estado de unidade ética pautada pelas noções de *igualdade* e *respeito*. Nesse contexto, de acordo com a filosofia moral reganiana, o direito mais fundamental a ser legado a um indivíduo é o de *ser tratado com respeito*. Todos os outros direitos, como, por exemplo, o direito à vida, liberdade e integridade física, advém da aceitação desse princípio deontológico central. Mas o que, em última instância, justificaria a concessão de tais direitos aos seres humanos? Segundo Regan, isso ocorre devido ao fato de esses serem *sujeitos-de-uma-vida*.

De acordo com o filósofo, um *sujeito-de-uma-vida* (*subject-of-a-life*) é um indivíduo autoconsciente e senciente, o qual possui interesses, preferências, desejos e crenças, uma percepção de mundo e concepção biográfica próprias, entre outras características que, em conjunto, tornam-no um ser vivo único. Regan cunhou a noção de *sujeito-de-uma-vida* com o intuito de se afastar lexicalmente de certos conceitos mal formulados, porém constantemente empregados em discussões de filosofia prática, a saber: *ser humano*, *pessoa* e *animal*. Para o autor, nenhuma dessas noções, tanto em sua acepção coloquial quanto em sua releitura semântica objetiva, é passível de englobar as qualidades relevantes demonstradas por diferentes indivíduos para a consideração moral e o tratamento respeitoso.

Não obstante, nota-se que, para Regan, é apenas um passo da aceitação da existência de direitos morais, e sua subsequente concessão aos humanos, até sua extensão a membros de outras espécies. Isso fica patente no momento em que se depreende que humanos e uma miríade de não humanos partilham das características compositivas da noção de *sujeito-de-uma-vida*. De fato, se as habilidades psicológicas supramencionadas forem o real passaporte para a outorga de direitos, então o círculo de atuação moral humano deve ser urgentemente ampliado de forma a compreender igualitariamente outros animais sencientes e autoconscientes. Com efeito, evidencia-se que, para Regan, uma abordagem de caráter ético-deontológico é a maneira mais eficaz de

facultar aos não humanos o respeito que lhes jamais deveria ter sido negado.

O terceiro segmento do livro é uma breve explanação sobre as principais táticas e métodos que vêm sendo utilizados pelas empresas e indústrias que têm como fonte de lucro o comércio de produtos de origem animal. Tais empreendimentos habitualmente seguem à risca uma legislação composta por *leis de bem-estar animal*, a qual tem como intento alcançar um padrão mínimo de manejo aos não humanos. Em outras palavras, a legislação bemestarista visa proporcionar um tratamento mais "humanitário", assim como uma "guarda responsável", aos não-humanos utilizados em seus projetos econômicos. Embora possa parecer um avanço no que tange à realidade deplorável vivenciada pelos não humanos, a criação e adoção de um tratamento mais "humanitário" acaba por viabilizar a continuidade da exploração animal, além de servir como fonte de benefício financeiro e *marketing* publicitário inesgotável para os produtores. Tendo isso em vista, Regan rejeita tais regulamentações e clama pelo término completo dessas atividades. Um panorama detalhado das ações exploratórias realizadas será apresentado a seguir, demonstrando por quais motivos uma proposta mais "humanitarista" é incapaz de elevar significativamente a qualidade de vida desses animais.

A seção seguinte de *Jaulas vazias* divide-se em quatro subcapítulos nos quais Regan delineia os mais diversos usos dos não humanos pelos seres humanos. Assim, inicialmente, o autor revela a brutal realidade dos animais criados intensivamente para o consumo – empreendimento que resulta na morte de bilhões de animais todos os anos. Regan pormenoriza as diferentes facetas da pecuária, analisando desde a indústria da vitela, passando pela criação e confinamento de aves, porcos e vacas leiteiras em baterias e baias de contenção, assim como o abate de gado e a matança de peixes. Além disso, o filósofo investiga a fundo a indústria e o mercado internacional de pele/couro. São reveladas as formas de captura de certos animais silvestres por meio de armadilhas dentadas, os vários métodos de extração de pele, bem como o seu posterior extermínio. Regan termina seu detalhamento acerca dessa atividade abominável trazendo dados chocantes sobre o massacre de focas, carneiros, cães e gatos com vistas aos fins supracitados.

Ainda nesse segmento, Regan aborda a situação dos não humanos utilizados para o divertimento e o entretenimento humano. Nesse

contexto, dá-se início à explanação com o quadro geral dos animais selvagens explorados em circos tradicionais. Acentua-se, essencialmente, a privação sistemática a qual esses seres são sujeitados pela absurda limitação de espaço e espancamentos habituais por parte de seus tratadores. As apresentações de mamíferos marinhos em espetáculos são igualmente problematizadas. Golfinhos, orcas e outros animais são capturados e retirados de seu *habitat* natural, sendo enclausurados em tanques aquáticos diminutos, tendo sua alimentação, bem como diversos aspectos de sua vida, sob total controle de seus tratadores. Essas atividades não apenas resultam em grave desestruturação social para esses animais, como também em severas anormalidades comportamentais. Em última instância, tais usos mostram-se claramente desnecessários, denotando uma das faces mais perversas e atrozes da humanidade.

Os dois subcapítulos restantes desse segmento são dedicados à utilização de animais em competições e como instrumentos de pesquisa científica, respectivamente. No tocante ao primeiro uso, Regan elucida os diferentes tipos de caça animal, os armamentos e aparatos empregados, e a atual situação de tal atividade como esporte. A caça cercada, por exemplo, representa um investimento enormemente lucrativo, pois paga-se uma quantia altíssima pelo direito a abater animais exóticos como antílopes, bisões, zebras, ursos e alces. Rodeios e torneios de laço de bezerros são igualmente alvos das pontuais denúncias reganianas. Diferentes equinos, bovinos e caprinos são criados unicamente visando à sua morte, pois ainda que em poucas ocasiões possam sobreviver às constantes fraturas e feridas causadas na arena, esses animais inevitavelmente são enviados a matadouros quando demonstram não mais serem capazes de permanecer "competindo".

Dando continuidade à sua investigação, o filósofo averigua a condição lastimável dos animais utilizados com vistas às experimentações biomédicas e testes com cosméticos. Embora seja comum argumentar sobre as vantagens provenientes das pesquisas com não humanos, Regan tenta provar que as experiências realizadas não apenas subestimam os danos causados, mas também superestimam seus benefícios – sobretudo quando são contrapostas às alternativas disponíveis. Assim, Regan passa a caracterizar os diferentes exames de toxicidade aplicados a cobaias em laboratórios, bem como os controversos procedimentos de dissecação e vivissecção. O teste *D.L. 50* (*Dose Letal Mediana*), por exemplo, tem o propósito de estabelecer a

dosagem em que uma determinada substância mostra-se fatal para 50% dos animais em que é testada. Muito embora a natureza desse e de outros experimentos seja continuamente questionada, sugere-se que tal estudo é fundamental para a comercialização de certos produtos contendo ingredientes químicos.

Como Regan faz questão de frisar, existem variantes "humanitárias" para quase todas as formas de exploração anteriormente expostas. A *Lei do Abate Humanitário* (HSA, em inglês) vigente nos EUA, por exemplo, requer que suínos e bovinos sejam atordoados com choques elétricos – o que supostamente os deixariam inconscientes – antes de serem degolados. Salienta-se, ainda, que as aves, de longe o maior número de animais mortos, não são cobertas por essa lei. Por seu turno, os partidários da "caça humanitária" buscam promover a ideia de que nas mãos de um caçador esportista habilidoso, um animal sofre menos. Ademais, a indústria farmacêutica, bem como diversos centros de estudo pelo mundo, defendem a "utilização responsável" e o "manejo humanitário" dos animais empregados como cobaias em experimentos científicos. Para Regan, as abordagens "humanitaristas" têm como escopo primeiro tornar os diferentes usos de não humanos mais aceitáveis à sociedade em geral. De fato, segundo o filósofo, tais propostas regulativas são incapazes de elevar, em qualquer sentido significativo, o bem-estar dos animais explorados.

A seção final de *Jaulas vazias* é composta por um apanhado geral de propostas de ação que, na opinião de Regan, poderiam aliviar o sofrimento não humano expressivamente, modificando o atual paradigma exploratório em direção a um possível panorama abolicionista. Primeiramente, para o autor, é imperativo a defesa de uma abordagem ético-filosófica centrada na concessão de direitos aos animais, mesmo que essa, por muitas vezes, seja taxada de utópica. Regan também sugere como uma tática aceitável em prol dos não humanos o *outling*, *i.e.*, o ato de alertar os integrantes de uma dada comunidade ou vizinhança acerca dos membros que violam direta ou indiretamente os direitos dos animais. Outrossim, Regan sustenta que, em certas ocasiões, valer-se de violência em nome dos animais poderia ser uma atitude moralmente justificável – pensamento que pouco agrada alguns de seus críticos. Consequentemente, poder-se-ia considerar Regan como um partidário e entusiasta de propostas centradas na ação direta praticada tanto como forma de desobediência civil quanto forma

de resgate aberto. Por essas e outras razões, é possível perceber porque esse é um dos segmentos mais controversos da obra aqui comentada.

Jaulas vazias: encarando o desafio dos direitos dos animais é um livro dinâmico e marcante. Por meio de sua escrita clara, Regan apresenta argumentos logicamente fundamentados, objetivando não somente atingir seus leitores intelectualmente, como também despertar um senso empático latente acerca do sofrimento animal. Faz-se necessário repensar sistematicamente a relação humano/não humano desde suas bases primeiras, estabelecendo assim os princípios gerais para uma defesa adequada da concessão de direitos aos membros de outras espécies. Em última instância, *Jaulas vazias* é um apelo à mobilização séria e efetiva em favor dos não humanos em um novo patamar. Pois, para Regan, não há por que batalhar em prol de leis dúbias, tratamentos ditos "humanitários", ou por jaulas maiores, mas única e exclusivamente para que, doravante, as jaulas encontrem-se vazias.

Tradução

A ética da inteligência artificial[1]

Nick Bostrom e Eliezer Yudkowsky

Tradução de Pablo Araújo Batista
Universidade São Judas Tadeu - SP

A possibilidade de criar máquinas pensantes levanta uma série de questões éticas. Essas questões se entrelaçam tanto para garantir que as máquinas não prejudiquem os humanos e outros seres moralmente relevantes, como para o *status* moral das próprias máquinas. A primeira seção discute questões que podem surgir no futuro próximo da Inteligência Artificial (IA). A segunda seção destaca os desafios para assegurar que a IA opere com segurança uma vez que se aproxima dos seres humanos e de sua inteligência. A terceira seção destaca a forma como podemos avaliar se, e em que circunstâncias, sistemas de IA possuem *status* moral. Na quarta seção, consideramos como sistemas de IA podem diferir dos humanos em alguns aspectos básicos relevantes para nossa avaliação ética deles. A seção final destina-se a questões da criação de IAs mais inteligente do que a inteligência humana, e assegurar que elas usem essa inteligência avançada para o bem ao invés de a utilizarem para o mal.

Ética em máquinas aprendizes e outros domínios específicos de algoritmos de IA

Imagine, num futuro próximo, um banco usando uma máquina de algoritmo de aprendizagem[2] para aprovar solicitações de pedidos de hipotecas. Um candidato rejeitado move uma ação contra o banco, alegando que o algoritmo está discriminando racialmente os solicitantes de hipoteca. O banco responde que isso é impossível, pois o algoritmo é deliberadamente cego para a raça do solicitante. Na realidade, isso faz parte da lógica do banco para implementação do sistema. Mesmo assim,

---

[1] N.T.: Texto original: "*The ethics of artificialiIntelligence*". Draft for Cambridge Handbook of Artificial Intelligence, eds. Willian Ramsey and Keith Frankish (Cambridge University Press, 2011): forthcoming. (Tradução permitida pelos autores e pela Cambridge University Press via correspondência eletrônica em 2 de fevereiro de 2012).

[2] N.T.: Os algoritmos de aprendizagem em sistemas artificiais é uma subdivisão da área da IA dedicada ao desenvolvimento de algoritmos e técnicas que permitam ao computador aprender e aperfeiçoar automaticamente seu desempenho em alguma tarefa por meio da experiência. Esses algoritmos são utilizados em processamento de linguagem natural, sistemas de busca, diagnósticos médicos, bioinformática, reconhecimento de fala, reconhecimento de escrita, visão computacional e na locomoção de robôs.

as estatísticas mostram que a taxa de aprovação do banco para candidatos negros tem constantemente caído. Submetendo dez candidatos aparentemente iguais e genuinamente qualificados (conforme determinado por um painel independente de juízes humanos), revela-se que o algoritmo aceita candidatos brancos e rejeita candidatos negros. O que poderia estar acontecendo?

Encontrar uma resposta pode não ser fácil. Se o algoritmo de aprendizagem da máquina é baseado em uma complexa rede neural ou em um algoritmo genético produzido por evolução dirigida, pode se revelar quase impossível entender por que, ou mesmo como, o algoritmo está julgando os candidatos com base em sua raça. Por outro lado, uma máquina aprendiz baseada em árvores de decisão ou redes bayesianas é muito mais transparente para inspeção do programador (HASTIE *et al,* 2001), permitindo a um auditor descobrir se o algoritmo de IA usa informações de endereço dos candidatos para saber previamente onde nasceram ou se residem em áreas predominantemente pobres.

Algoritmos de IA desempenham um papel cada vez maior na sociedade moderna, embora geralmente não estejam rotulados como "IA". O cenário descrito anteriormente pode estar acontecendo da mesma forma como nós descrevemos. E se tornará cada vez mais importante desenvolver algoritmos de IA que não sejam apenas poderosos e escaláveis[3], mas também *transparentes para inspeção* – para citar umas das muitas propriedades socialmente importantes.

Alguns desafios de máquinas éticas são muito semelhantes a outros desafios envolvidos em projetar máquinas. Projetar um braço robótico para evitar o esmagamento de seres humanos distraídos não é moralmente mais preocupante do que projetar um retardador de chamas para sofá. Trata-se de novos desafios de programação, mas não de novos desafios éticos. Mas, quando algoritmos de IA se ocupam de trabalho cognitivo com dimensões sociais – tarefas cognitivas anteriormente realizadas por humanos – o algoritmo de IA herda as exigências sociais.

[3] N.T.: Definir escalabilidade é uma tarefa difícil, mas o conceito é extremamente importante no tratamento de sistemas eletrônicos. A escalabilidade está relacionada com o aumento do desempenho de um sistema à medida que mais *hardware* é acrescentado. Um sistema escalável deve ser capaz de manipular informações de forma crescente mantendo a coerência do trabalho executado. (BONDI, A. B. "*Characteristics of scalability and their impact on performance"*, Network Design and Performance Analysis Department, 2000. p. 195-203). Disponível em: http://www.win.tue.nl/~johanl/educ/2II45/Lit/Scalability-bondi%202000.pdf

Seria sem dúvida frustrante descobrir que nenhum banco no mundo deseja aprovar a sua aparentemente excelente solicitação de empréstimo, sem que ninguém saiba por que, e ninguém pode ainda descobrir mesmo em princípio. (Talvez você tenha um primeiro nome fortemente associado com fraqueza? Quem sabe?).

Transparência não é a única característica desejável da IA. Também é importante que algoritmos de IA que assumam funções sociais sejam previsíveis aos que o governam. Para compreender a importância dessa previsibilidade, considere uma analogia. O princípio legal de *stare decisis*[4] impele juízes a seguir os antecedentes sempre que possível. Para um engenheiro, essa preferência pelo precedente pode parecer incompreensível – por que amarrar o futuro com o passado, quando a tecnologia está sempre melhorando? Mas, uma das mais importantes funções do sistema legal é ser previsível, de modo que, por exemplo, os contratos possam ser escritos sabendo como eles serão executados. O trabalho do sistema jurídico não é necessariamente o de aperfeiçoar a sociedade, mas proporcionar um ambiente previsível no qual cidadãos possam aperfeiçoar suas próprias vidas.

Também se tornará cada vez mais importante que os algoritmos de IA se tornem *resistentes à manipulação*. Um sistema visual de máquinas que faz a varredura de bagagem em aeroportos deve ser resistente contra adversários humanos deliberadamente à procura de fraquezas exploráveis no algoritmo, por exemplo, um objeto que, colocado próximo a uma pistola em uma das bagagens, neutralizaria o reconhecimento dela. Resistência contra manipulação é um critério comum em segurança da informação; quase *o* critério. Mas, não é um critério que aparece frequentemente em revistas especializadas em aprendizagem de máquinas, que estão atualmente mais interessados em, por exemplo, como um algoritmo aumenta proporcionalmente em grandes sistemas paralelos.

Outro importante critério social para transações em organizações é ser capaz de encontrar a pessoa responsável por conseguir que algo seja feito. Quando um sistema de IA falha em suas tarefas designadas, quem leva a culpa? Os programadores? Os usuários finais? Burocratas

[4] N.T.: *Stare decisis* é uma expressão em latim que pode ser traduzida como "ficar com as coisas decididas". Essa expressão é utilizada no direito para se referir à doutrina segundo a qual as decisões de um órgão judicial criam precedente, ou seja, jurisprudência, e se vinculam às decisões que serão emitidas no futuro.

modernos muitas vezes se refugiam nos procedimentos estabelecidos que distribuem responsabilidade amplamente, de modo que uma pessoa não pode ser identificada nem culpada pelo resultado das catástrofes (HOWARD, 1994). O provável julgamento comprovadamente desinteressado de um sistema especialista poderia transformar-se num refúgio ainda melhor. Mesmo que um sistema de IA seja projetado com uma substituição do usuário, é uma obrigação considerar o incentivo na carreira de um burocrata que será pessoalmente responsabilizado se a substituição sair errada, e que preferiria muito mais culpar a IA por qualquer decisão difícil com um resultado negativo.

Responsabilidade, transparência, auditabilidade, incorruptibilidade, previsibilidade e uma tendência para não fazer vítimas inocentes gritarem em desamparada frustração: todos os critérios que aplicamos aos humanos que desempenham funções sociais; todos os critérios que devem ser considerados em um algoritmo destinado a substituir o julgamento humano de funções sociais; todos os critérios que podem não aparecer em um registro de aprendizado de máquina considerando o quanto um algoritmo aumenta proporcionalmente para mais computadores. Essa lista de critérios não é, de forma alguma, exaustiva, mas serve como uma pequena amostra do que uma sociedade cada vez mais informatizada deveria estar pensando.

Inteligência artificial geral

Há concordância quase universal entre os profissionais modernos de IA que sistemas de Inteligência Artificial estão aquém das capacidades humanas em algum sentido crítico, embora algoritmos de IA tenham batido os seres humanos em muitos domínios específicos como, por exemplo, o xadrez. Tem sido sugerido por alguns que, logo que os pesquisadores de IA descobrem como fazer alguma coisa, essa capacidade deixa de ser considerada como inteligente – o xadrez era considerado o epítome da inteligência até o *Deep Blue* vencer Kasparov no campeonato mundial – mas mesmo esses pesquisadores concordam que algo importante está faltando às IAs modernas. (ver HOFSTADTER, 2006).

Enquanto essa subárea da Inteligência Artificial está apenas crescendo de forma unificada, "Inteligência Artificial Geral" (IAG) é o termo emergente usado para designar IA "real" (ver, por exemplo, o volume editado por Goertzel e Pennachin, 2006). Como o nome implica,

o consenso emergente é que a característica que falta é a generalidade. Os algoritmos atuais de IA com desempenho equivalente ou superior ao humano são caracterizados por uma competência deliberadamente programada em um único e restrito domínio. O *Deep Blue* tornou-se o campeão do mundo em xadrez, mas ele não pode jogar damas, muito menos dirigir um carro ou fazer uma descoberta científica. Tais algoritmos modernos de IA assemelham-se a todas as formas de vidas biológicas com a única exceção do *Homo sapiens*. Uma abelha exibe competência em construir colmeias; um castor exibe competência em construir diques; mas uma abelha não pode construir diques, e um castor não pode aprender a fazer uma colmeia. Um humano, observando, pode aprender a fazer ambos, mas essa é uma habilidade única entre as formas de vida biológica. É discutível se a inteligência humana é verdadeiramente geral – nós somos certamente melhor em algumas tarefas cognitivas do que os outros (HIRSCHFELD e GELMAN, 1994) – mas a inteligência humana é, sem dúvida, significativamente mais geralmente aplicável que a inteligência não hominídea.

É relativamente fácil imaginar o tipo de questões de segurança que podem resultar de IA operando somente dentro de um domínio específico. É uma classe qualitativamente diferente de problema manipular uma IAG operando por meio de muitos novos contextos que não podem ser previstos com antecedência.

Quando os engenheiros humanos constroem um reator nuclear, eles preveem eventos específicos que poderiam acontecer em seu interior – falhas nas válvulas, falhas nos computadores, aumento de temperatura no núcleo – para evitar que esse evento se torne catastrófico. Ou, em um nível mais mundano, a construção de uma torradeira envolve previsão do pão e previsão da reação do pão para os elementos de aquecimento da torradeira. A torradeira em si não sabe que o seu objetivo é fazer torradas – o propósito da torradeira é representado na mente do *designer*, mas não é explicitamente representado em computações dentro da torradeira – e se você colocar um pano dentro de uma torradeira, ela pode pegar fogo, pois o projeto é realizado em um contexto não previsto com um imprevisível efeito colateral.

Mesmo algoritmos de IA de tarefas específicas nos lançam fora do paradigma da torradeira, o domínio do comportamento especificamente previsto localmente pré-programado. Considere o *Deep Blue*, o algoritmo de xadrez que venceu Garry Kasparov no campeonato mundial de

xadrez. Na hipótese de as máquinas poderem apenas fazer exatamente o que eles dizem, os programadores teriam de pré-programar manualmente um banco de dados contendo movimentos possíveis para cada posição de xadrez que o *Deep Blue* poderia encontrar. Mas isso não era uma opção para os programadores do *Deep Blue*. Em primeiro lugar, o espaço de possíveis posições do xadrez é abundantemente não gerenciável. Segundo, se os programadores tinham de inserir manualmente o que consideravam um bom movimento em cada situação possível, o sistema resultante não teria sido capaz de fazer movimentos mais fortes de xadrez do que o de seus criadores. Uma vez que os próprios programadores não são campeões do mundo, esse sistema não teria sido capaz de derrotar Garry Kasparov.

Ao criar um superjogador de xadrez, os programadores humanos necessariamente sacrificaram sua capacidade de previsão *local* para o *Deep Blue*, *específico* comportamento do jogo. Em vez disso, os programadores do *Deep Blue* tinham (justificável) confiança que os movimentos de xadrez do *Deep Blue* satisfariam um critério não local de otimização: isto é, que os movimentos tenderiam a orientar o futuro resultado do jogo na região "vencedora", conforme definido pelas regras do xadrez. Essa previsão sobre consequências distantes, embora provadas corretas, não permitiu aos programadores prever o comportamento *local* do *Deep Blue* – sua resposta a um determinado ataque ao seu rei – porque o *Deep Blue* computa o mapa do jogo não local, a ligação entre um movimento e suas possíveis consequências futuras com mais precisão do que os programadores poderiam fazer (YUDKOWSKY, 2006).

Os seres humanos modernos fazem literalmente milhões de coisas para se alimentar − para servir ao objetivo final de ser alimentado. Algumas dessas atividades foram "previstas pela Natureza" no sentido de ser um desafio ancestral ao qual nós estamos diretamente adaptados. Mas o nosso cérebro adaptado cresceu poderoso o suficiente para ser, de forma significativa, mais geralmente aplicável; permite-nos prever as consequências de milhões de diferentes ações em vários domínios, e exercer nossas preferências sobre os resultados finais. Os seres humanos cruzaram o espaço para colocar sua pegada na Lua, apesar de nenhum de nossos ancestrais ter encontrado um desafio análogo ao vácuo. Em relação ao domínio específico de IA, é um problema qualitativamente diferente projetar um sistema que vai operar com segurança em milhares de contextos, incluindo contextos que não sejam

especificamente previstos por qualquer dos *designers* ou usuários e incluindo contextos que nenhum humano jamais encontrou. Nesse momento, não pode haver nenhuma especificação local de bom comportamento – não uma simples especificação sobre seus próprios comportamentos, não mais de que existe uma descrição local compacta de todas as maneiras que os seres humanos obtêm seu pão de cada dia.

Para construir uma IA que atue com segurança enquanto age em vários domínios, com muitas consequências, incluindo os problemas que os engenheiros nunca previram explicitamente, é preciso especificar o bom comportamento em termos como *"X tal que a consequência de X não é prejudicial aos seres humanos"*. Isso é não local, e envolve extrapolar a consequência distante de nossas ações. Assim, essa é apenas uma especificação efetiva – que pode ser realizada como uma propriedade do *design* – se o sistema extrapola explicitamente as consequências de seu comportamento. Uma torradeira não pode ter essa propriedade de *design* porque uma torradeira não pode prever as consequências do pão tostado.

Imagine um engenheiro tendo que dizer: "Bem, eu não tenho ideia de como esse avião que eu construí possa voar com segurança – de fato, eu não tenho ideia de como ele fará tudo, se ele vai bater as asas ou inflar-se com hélio, ou outra coisa que eu nem sequer imagino, mas eu lhe asseguro, o projeto é muito, muito seguro". Isso pode parecer uma posição invejável da perspectiva de relações públicas, mas é difícil ver que outra garantia de comportamento ético seria possível para uma operação de inteligência geral sobre problemas imprevistos, em vários domínios, com preferências sobre consequências distantes. Inspecionando o *design* cognitivo, podemos verificar que a mente estava, na verdade, buscando soluções que nós classificaríamos como ética; mas não poderíamos prever que solução específica a mente descobriria.

Respeitar essa verificação exige alguma forma de distinguir as garantias de confiança (um procedimento que não desejo dizer "a IA é segura a menos que a IA seja realmente segura") de pura esperança e pensamento mágico ("Não tenho ideia de como a Pedra Filosofal vai transformar chumbo em ouro, mas eu lhe asseguro, ela vai!"). Deve-se ter em mente que expectativas puramente esperançosas já foram um problema em pesquisa de IA (MCDERMOTT, 1976).

Comprovadamente construir uma IAG de confiança exigirá métodos diferentes e uma maneira diferente de pensar, para inspecionar uma

falha no *software* de uma usina de energia - ele exigirá um IAG que pensa como um engenheiro humano preocupado com a ética, não apenas um simples produto da engenharia ética.

Dessa forma, a disciplina de IA ética, especialmente quando aplicada à IAG, pode diferir fundamentalmente da disciplina ética de tecnologias não cognitivas, em que

- o comportamento específico local da IA não pode ser previsível independentemente de sua segurança, mesmo se os programadores fizerem tudo certo;
- a verificação de segurança do sistema torna-se um desafio maior, porque nós devemos verificar o comportamento seguro do sistema operando em todos os contextos;
- a própria cognição ética deve ser tomada como um assunto de engenharia.

Máquinas com *status* moral

Um diferente conjunto de questões éticas surge quando se contempla a possibilidade de que alguns futuros sistemas de IA possam ser candidatos a possuírem *status* moral. Nossas relações com os seres que possuem *status* moral não são exclusivamente uma questão de racionalidade instrumental: nós também temos razões morais para tratá-los de certas maneiras e de nos *refrearmos de tratá-los de outras formas.* Francis Kamm propôs a seguinte definição do *status* moral, que servirá para nossos propósitos:

> *X tem status moral = porque X conta moralmente em seu próprio direito, e é permitido/proibido fazer as coisas para ele para seu próprio bem.* (KAMM, 2007, cap. 7; paráfrase).

Uma pedra não tem *status* moral: podemos esmagá-la, pulverizá-la ou submetê-la a qualquer tratamento que desejamos sem qualquer preocupação com a própria rocha. Uma pessoa humana, por outro lado, deve ser encarada não apenas como um meio, mas também como um fim. Exatamente o que significa tratar uma pessoa como um fim é algo sobre o qual diferentes teorias éticas discordam; mas ela certamente toma os seus interesses legítimos em conta – atribuindo peso para o seu bem-estar – e também pode aceitar severas restrições morais em nossa

relação com ela, como a proibição contra assassiná-la, roubá-la, ou fazer uma série de outras coisas para ela ou para sua propriedade sem o seu consentimento. Além disso, é porque a pessoa humana é importante em seu próprio direito, e por seu bem-estar é que estamos proibidos de fazer com ela essas coisas. Isso pode ser expresso de forma mais concisa, dizendo que uma pessoa humana tem *status* moral.

Perguntas sobre *status* moral são importantes em algumas áreas da ética prática. Por exemplo, as disputas sobre a legitimidade moral do aborto muitas vezes levam a desacordos sobre o *status* moral do embrião. Controvérsias sobre experimentação animal e o tratamento dispensado aos animais na indústria de alimentos envolvem questões sobre o *status* moral de diferentes espécies de animais. E as nossas obrigações em relação a seres humanos com demência grave, tais como pacientes em estágio final de Alzheimer, também podem depender de questões de *status* moral.

É amplamente aceito que os atuais sistemas de IA não têm *status* moral. Nós podemos alterar, copiar, encerrar, apagar ou utilizar programas de computador tanto quanto nos agradar, ao menos no que diz respeito aos próprios programas. As restrições morais a que estamos sujeitos em nossas relações com os sistemas contemporâneos de IA são todas baseadas em nossas responsabilidades para com os outros seres, tais como os nossos companheiros humanos, e não em quaisquer direitos para os próprios sistemas.

Embora seja realmente consensual que aos sistemas atuais de IA falta *status* moral, não está claro exatamente quais atributos servem de base para ele. Dois critérios são comumente propostos como importantemente relacionado com o estatuto moral, ou isoladamente ou em combinação: a senciência e a sapiência (ou personalidade). Estes podem ser caracterizados aproximadamente como segue:

> Senciência: capacidade para a experiência fenomenal ou *qualia*, como a capacidade de sentir dor e sofrer;
>
> Sapiência: conjunto de capacidades associadas com maior inteligência, como a autoconsciência e ser um agente racional responsável.

Uma opinião comum é que muitos animais têm *qualia* e, portanto, têm algum *status* moral, mas que apenas os seres humanos têm

sabedoria, o que lhes confere um *status* moral mais elevado do que possuem os animais não humanos[5]. Essa visão, é claro, deve enfrentar a existência de casos limítrofes, tais como, por um lado, crianças ou seres humanos com grave retardo mental – às vezes, infelizmente, referidos como "humanos marginais"– que não satisfazem os critérios de sapiência; e, por outro lado, alguns animais não humanos, tais como os grandes símios, que podem ter, pelo menos, alguns dos elementos da sapiência. Alguns negam que o chamado "homem marginal" tenha um *status* moral pleno; outros propõem maneiras adicionais em que um objeto poderia qualificar-se como um sustentador de *status* moral, tais como ser um membro de uma espécie que normalmente tem sensibilidade ou sapiência, ou por estar em uma relação adequada para alguns seres que independentemente têm *status* moral (cf. WARREN, 2000). Para o propósito do texto, no entanto, nos concentraremos nos critérios de sensibilidade e sapiência.

Essa imagem de *status* moral sugere que um sistema de IA terá algum *status* moral se ele tiver a capacidade de *qualia*, tais como a capacidade de sentir dor. Um sistema de IA senciente, mesmo que não tenha linguagem e outras faculdades cognitivas superiores, não será como um bichinho de pelúcia ou um boneco; será mais como um animal vivo. É errado infligir dor a um rato, a menos que existam razões suficientemente fortes e prevalecentes razões morais para fazê-lo. O mesmo vale para qualquer sistema senciente de IA. Se além de consciência um sistema de inteligência artificial também tiver sapiência de um tipo semelhante à de um adulto humano normal, então terá também pleno *status* moral, equivalente ao dos seres humanos.

Uma das ideias subjacentes a essa avaliação moral pode ser expressa de forma mais forte como um princípio de não discriminação:

Princípio da não discriminação do substrato

> Se dois seres têm a mesma funcionalidade e a mesma experiência consciente, e diferem apenas no substrato de sua aplicação, então eles têm o mesmo *status* moral.

[5] Alternativamente, poder-se-ia negar que o estatuto moral vem em graus. Em vez disso, pode-se considerar que certos seres têm interesses mais importantes do que outros seres. Assim, por exemplo, alguém poderia alegar que é melhor salvar um ser humano do que salvar um pássaro, não porque o ser humano tem maior *status* moral, mas porque o ser humano tem um interesse mais significativo em ter sua vida salva do que um pássaro.

Pode-se argumentar a favor desse princípio, por razões que rejeitá-lo equivaleria a adotar uma posição similar ao racismo: substrato carece de fundamental significado moral, da mesma forma e pela mesma razão que a cor da pele também carece. O *princípio da não discriminação do substrato* não implica que um computador digital possa ser consciente, ou que possa ter a mesma funcionalidade que um ser humano. O substrato pode ser moralmente relevante à medida que faz a diferença para a senciência, ou funcionalidade. Mas, mantendo essas coisas constantes, não faz diferença moral se um ser é feito de silício ou de carbono, ou se o cérebro usa semicondutores ou neurotransmissores.

Um princípio adicional que pode ser proposto é o fato de que sistemas de IA sejam artificiais - ou seja, o produto de *design* deliberado - não é fundamentalmente relevante para o seu *status* moral. Nós poderíamos formular isso da seguinte forma:

Princípio da não discriminação da ontogenia

> Se dois seres têm a mesma funcionalidade e mesma experiência de consciência, e diferem apenas na forma como vieram a existir, então eles têm o mesmo *status* moral.

Hoje essa ideia é amplamente aceita no caso de humanos − embora em alguns grupos de pessoas, particularmente no passado, a ideia do que seja um *status* moral dependa de uma linhagem ou casta, e que tenha sido influente. Nós não acreditamos que fatores causais, tais como planejamento familiar, assistência ao parto, fertilização *in vitro*, seleção de gametas, melhoria deliberada da nutrição materna etc. − que introduzem um elemento de escolha deliberada e *design* na criação de seres humanos − têm alguma implicação necessária para o *status* moral da progênie. Mesmo aqueles que se opõem à clonagem para reprodução humana, por razões morais ou religiosas, em geral, aceitam que, se um clone humano fosse trazido à existência, ele teria o mesmo *status* moral que qualquer outra criança humana. O *princípio da não discriminação da ontogenia* estende esse raciocínio aos casos envolvendo sistemas cognitivos inteiramente artificiais.

É evidentemente possível, nas circunstâncias da criação, afetar a descendência resultante de maneira a alterar o seu *status* moral. Por exemplo, se algum procedimento realizado durante a concepção ou a gestação é a causa do desenvolvimento de um feto humano sem cérebro, então esse fato sobre a ontogenia seria relevante para o nosso

julgamento sobre o *status* moral da prole. A criança anencefálica, porém, teria o mesmo *status* moral que qualquer outra similar criança anencefálica, incluindo aquela que tenha sido concebida por processo totalmente natural. A diferença de *status* moral entre uma criança anencefálica e uma criança normal está baseada na diferença qualitativa entre os dois, o fato de que um tem uma mente, enquanto o outro não. Desde que as duas crianças não tenham a mesma funcionalidade e a mesma experiência consciente, o *princípio da não discriminação da ontogenia* não se aplica.

Embora esse princípio afirme que a ontogenia dos seres não tem nenhuma relevância fundamental sobre o seu *status* moral, ele não nega que os fatos sobre a ontogênese podem afetar obrigações particulares que os agentes morais têm para com o ser em questão. Os pais têm deveres especiais para com seus filhos, mas não têm para com outras crianças, e que não teriam mesmo se houvesse outra criança qualitativamente idêntica a sua. Similarmente, o *princípio da não discriminação da ontogenia* é consistente com a alegação de que os criadores ou proprietários de um sistema de IA com *status* moral podem ter direitos especiais para com sua mente artificial que não têm para com outras mentes artificiais, mesmo se as mentes em questão são qualitativamente semelhantes e têm o mesmo *status* moral.

Se os princípios de não discriminação com relação ao substrato e ontogenia são aceitos, então muitas questões sobre como devemos tratar mentes artificiais podem ser respondidas por aplicarmos os mesmos princípios morais que usamos para determinar nossos deveres em contextos mais familiares. À medida que os deveres morais decorrem de considerações sobre *status* moral, devemos tratar a mente artificial da mesma maneira como devemos tratar uma mente natural humana qualitativamente idêntica e em uma situação similar. Isso simplifica o problema do desenvolvimento de uma ética para o tratamento de mentes artificiais.

Mesmo se aceitarmos essa postura, no entanto, temos de enfrentar uma série de novas questões éticas que os princípios mencionados deixaram sem resposta. Novas questões éticas surgem porque mentes artificiais podem ter propriedades muito diferentes das ordinárias mentes humana ou animal. Devemos considerar como essas novas propriedades afetariam o *status* moral de mentes artificiais e o que significaria respeitar o *status* moral de tais mentes exóticas.

Mentes com propriedades exóticas

No caso dos seres humanos, nós normalmente não hesitamos em atribuir sensibilidade e experiência consciente a qualquer indivíduo que apresenta os tipos normais de comportamento humano. Poucos acreditam que haja outras pessoas que atuem de forma perfeitamente normal, mas lhes falte consciência. No entanto, outros seres humanos não apenas se comportam como pessoas normais de maneiras semelhantes a nós mesmos; eles também têm cérebros e arquiteturas cognitivas que são constituídas de forma muito parecida com a nossa. Um intelecto artificial, pelo contrário, pode ser constituído um pouco diferentemente de um intelecto humano e ainda assim apresentar um comportamento semelhante ao humano ou possuir disposições comportamentais normalmente indicativas de personalidade. Por isso, seria possível conceber um intelecto artificial que seria sapiente, e talvez fosse uma pessoa, e ainda assim não estaria consciente ou teria experiências conscientes de qualquer tipo. (Se isso é realmente possível, depende das respostas a algumas questões metafísicas não triviais). Se tal sistema fosse possível, ele levantaria a questão de saber se uma pessoa não senciente teria qualquer *status* moral sequer; e nesse caso, se teria o mesmo *status* moral de uma pessoa sensível. A senciência, ou pelo menos uma capacidade de senciência, é comumente assumida como estando presente em qualquer indivíduo que seja uma pessoa, e essa questão não tem recebido muita atenção até o momento.[6]

Outra propriedade exótica, o que certamente é metafisicamente e fisicamente possível para uma inteligência artificial, é que a *taxa subjetiva de tempo* desvia-se drasticamente da taxa que é característica de um cérebro biológico humano. O conceito de taxa subjetiva do tempo

[6] Esta questão está relacionada com alguns problemas na filosofia da mente que têm recebido grande atenção, em particular o "problema do zumbi", que pode ser formulado da seguinte forma: Existe um mundo metafisicamente possível que seja idêntico ao mundo real no que diz respeito a todos os fatos físicos (incluindo a microestrutura física exata de todos os cérebros e organismos), mas que difere do mundo real em relação a alguns fatos fenomenais (experiência subjetiva)? Colocado de forma mais crua, é metafisicamente possível que possa haver uma pessoa que é fisicamente e exatamente idêntica a você, mas que é um "zumbi", ou seja, sem *qualia* e consciência fenomenal? (CHALMERS, 1996). Essa questão familiar é diferente do referido no texto: ao nosso "zumbi" é permitido ter sistematicamente diferentes propriedades físicas dos seres humanos. Além disso, queremos chamar a atenção ao *status* ético de um zumbi sapiente.

é mais bem explicado primeiramente pela introdução da ideia de emulação de todo o cérebro, ou "*uploading*".[7]

"Uploading" refere-se a uma hipotética tecnologia do futuro que permitiria a um intelecto humano ou de outro animal serem transferidos de sua aplicação original em um cérebro orgânico para um computador digital. Um cenário como este: Primeiro, uma alta resolução de varredura é realizada em algumas particularidades do cérebro, possivelmente destruindo o original no processo. Por exemplo, o cérebro pode ser vitrificado e dissecado em fatias finas, que podem então ser digitalizadas usando alguma forma de microscópio de alta capacidade combinada com o reconhecimento automático de imagem. Podemos imaginar que essa análise deve ser detalhada o suficiente para capturar todos os neurônios, suas interconexões sinápticas, e outras características que são funcionalmente relevantes para as operações do cérebro original. Em segundo lugar, esse mapa tridimensional dos componentes do cérebro e suas interconexões são combinados com uma biblioteca de avançadas teorias da neurociência que especifica as propriedades computacionais de cada tipo básico de elementos, tais como diferentes tipos de neurônios e de junção sináptica. Terceiro, a estrutura computacional e os algoritmos associados de comportamento dos seus componentes são implementados em alguns computadores poderosos. Se o processo de *uploading* for bem sucedido, o programa de computador deve agora replicar as características funcionais essenciais do cérebro original. O resultante *upload* pode habitar uma realidade virtual simulada, ou,

[7] N.T.: Traduzir a palavra "*upload*" de forma literal traria problemas para o entendimento do texto. Por isso, considerei mais apropriado manter a palavra em sua forma original. Um entendimento melhor do que significa *upload* nesse contexto pode ser encontrado em outro texto do Bostrom em parceria com Anders Sandberg: *The concept of brain emulation: Whole brain emulation, often informally called "uploading" or "downloading", has been the subject of much science fiction and also some preliminary studies (…). The basic idea is to take a particular brain, scan its structure in detail, and construct a software model of it that is so faithful to the original that, when run on appropriate hardware, it will behave in essentially the same way as the original brain*. (*Whole Brain Emulation – A Roadmap. p.7, 2008.* Disponível em: http://www.philosophy.ox.ac.uk/__data/assets/pdf_file/0019/3853/brain-emulation-roadmap-report.pdf). Numa tradução livre: "O conceito de emulação do cérebro: Emulação do cérebro inteiro, muitas vezes, informalmente chamado de "*upload*" ou "*download*", tem sido o objeto de muita ficção científica e também de alguns estudos preliminares (...). A ideia básica é ter um cérebro especial, digitalizar a sua estrutura em detalhe, e construir um modelo de *software* que é tão fiel ao original que, quando executado em *hardware* apropriado, irá se comportar basicamente da mesma maneira que o cérebro original."

alternativamente, pode ser dado a ele o controle de um corpo robótico, que lhe permite interagir diretamente com a realidade física externa.

Uma série de perguntas surge no contexto de tal cenário: Quão plausível é que esse procedimento um dia se torne tecnologicamente viável? Se o procedimento foi bem sucedido e produziu um programa de computador exibindo aproximadamente a mesma personalidade, as mesmas memórias e os mesmos padrões de pensamento que o cérebro original, pode esse programa ser sensível? Será o *upload* a mesma pessoa que o indivíduo cujo cérebro foi desmontado no processo de carregamento? O que acontece com a identidade pessoal se um *upload* é copiado de tal forma que duas mentes semelhantes ou qualitativamente idênticas de *upload* sejam executadas em paralelo? Apesar de todas essas questões serem relevantes para a ética da IA, aqui vamos nos concentrar na questão que envolve a noção de uma taxa subjetiva de tempo.

Suponha que um *upload* possa ser senciente. Se executarmos o programa de transferência em um computador mais rápido, isso fará com que o *upload*, se ele estiver conectado a um dispositivo de entrada como uma câmera de vídeo, perceba o mundo externo como se esse estivesse perdendo velocidade. Por exemplo, se o *upload* está sendo executado milhares de vezes mais rápido que o cérebro original, então o mundo exterior será exibido para o *upload* como se fosse desacelerado por um fator de mil. Alguém deixa cair uma caneca física de café: o *upload* observa a caneca lentamente caindo no chão enquanto termina de ler o jornal pela manhã e envia alguns *e-mails*. Um segundo de tempo objetivo corresponde a 17 minutos do tempo subjetivo. Duração objetiva e subjetiva podem então divergir.

Tempo subjetivo não é a mesma estimativa ou percepção que um sujeito tem de quão rápido o tempo flui. Os seres humanos são muitas vezes confundidos com o fluxo do tempo. Podemos acreditar que se trata de 1 hora quando de fato são 2h15min.; ou uma droga estimulante pode acelerar nossos pensamentos, fazendo parecer que mais tempo subjetivo tenha decorrido do que realmente é o caso. Essas situações envolvem uma percepção distorcida do tempo em vez de uma mudança na taxa de tempo subjetivo. Mesmo em um cérebro ex-viciado em cocaína, provavelmente não há mudança significativa na velocidade de base nos cálculos neurológicos; mais provavelmente, a droga está fazendo o cérebro cintilar mais rapidamente, a partir de um pensamento para

outro, fazendo-o gastar menos tempo subjetivo para pensar um número maior de pensamentos distintos.

A variabilidade da taxa subjetiva do tempo é uma propriedade exótica de mentes artificiais que levanta novas questões éticas. Por exemplo, os casos em que a duração de uma experiência é eticamente relevante devem ser mensurados durações no tempo objetivo ou subjetivo? Se um *upload* cometeu um crime e é condenado a quatro anos de prisão, devem ser quatro anos objetivos – que podem corresponder a muitos milênios de tempo subjetivo – ou devem ser quatro anos subjetivos, podendo ser pouco mais que um par de dias de tempo objetivo? Se uma IA avançada e um ser humano estão com dor, é mais urgente aliviar a dor da IA, em razão de que ela experimenta maior duração subjetiva da dor para cada segundo sideral[8] que o alívio é retardado? Uma vez que em nosso contexto habitual de humanos biológicos tempo subjetivo não é significativamente variável, não é surpreendente que esse tipo de questionamento não seja francamente respondido por normas éticas familiares, mesmo se essas normas são estendidas à IA por meio de princípios de não discriminação (como os propostos na seção anterior).

[8] N.T.: Tempo sideral pode ser entendido como "tempo estelar", pois é uma medida de tempo baseada na rotação da Terra em relação às estrelas fixas. Em nossas vidas e tarefas diárias, costumamos utilizar o tempo solar, que tem como unidade fundamental o dia, ou seja, o tempo que o Sol demora para viajar 360 graus em torno do céu, devido à rotação da Terra. O tempo solar também possui unidades menores que são subdivisões de um dia: *1/24 dia = 1 hora, 1/60 hora = 1 minuto e 1/60 minuto = 1 segundo.* Mas, o tempo solar apresenta dificuldades pois a Terra não gira em torno de si 360° num dia solar. A Terra está em órbita ao redor do Sol ao longo de um dia, e ele se move cerca de um grau ao longo de sua órbita (360 graus/365.25 dias para uma órbita completa = cerca de um grau por dia). Assim, em 24 horas, o trajeto em torno do Sol varia em cerca de um grau. Portanto, a Terra só tem que girar 359° para fazer o Sol parecer ter viajado 360° no céu. Em astronomia, é relevante quanto tempo a Terra leva para girar com relação às estrelas "fixas", por isso, é necessário uma escala de tempo que remova a complicação da órbita da Terra em torno do Sol, e apenas se concentre em quanto tempo a Terra leva para girar 360° com relação às estrelas. Esse período de rotação é chamado de dia sideral. Em média, é de 4 minutos a mais do que um dia solar, devido ao grau extra que a Terra tem que girar para completar 360°. Ao invés de definir um dia sideral em 24 horas e 4 minutos, nós definimos horas siderais, minutos e segundos que são a mesma fração de um dia como os seus homólogos solar. Portanto, um segundo sideral = 1,00278 segundos solar. O tempo sideral divide uma rotação completa da Terra em 24 horas siderais, e é útil para determinar onde as estrelas estão em determinado momento. (Disponível em: http://docs.kde.org/stable/en/kdeedu/kstars/ai-sidereal.html)

Para ilustrar o tipo de afirmação ética que pode ser relevante aqui, nós formulamos (mas não defendemos) um princípio de privilegiar tempo subjetivo como a noção normativa mais fundamental:

Princípio da taxa subjetiva do tempo

> Nos casos em que a duração de uma experiência tem um significado normativo básico, é a duração subjetiva da experiência que conta.

Até agora, discutimos duas possibilidades (sapiência não senciente e a taxa subjetiva de tempo variável), que são exóticas no sentido de ser relativamente profundas e metafisicamente problemáticas, assim como faltam exemplos claros ou paralelos no mundo contemporâneo. Outras propriedades de possíveis mentes artificiais seriam exóticas em um sentido mais superficial; por exemplo, por serem divergentes em algumas dimensões quantitativas não problemáticas do tipo de mente com o qual estamos familiarizados. Mas, tais características superficialmente exóticas também podem representar novos problemas éticos – se não ao nível fundamental da filosofia moral, ao menos no nível da ética aplicada ou para princípios éticos de complexidade média.

Um importante conjunto de propriedades exóticas de inteligências artificiais se relaciona com a reprodução. Certo número de condições empíricas que se aplicam a reprodução humana não é aplicável a IA. Por exemplo, as crianças humanas são o produto de uma recombinação do material genético dos dois genitores; os pais têm uma capacidade limitada para influenciar o caráter de seus descendentes; um embrião humano precisa ser gestado no ventre durante nove meses; leva de quinze a vinte anos para uma criança humana atingir a maturidade; a criança humana não herda as habilidades e os conhecimentos adquiridos pelos seus pais; os seres humanos possuem um complexo e evoluído conjunto de adaptações emocionais relacionados à reprodução, carinho, e da relação pais e filhos. Nenhuma dessas condições empíricas deve pertencer ao contexto da reprodução de uma máquina inteligente. Por isso, é plausível que muitos dos princípios morais de nível médio que temos aceitado como as normas que regem a reprodução humana precisarão ser repensadas no contexto da reprodução de IA.

Para ilustrar por que algumas de nossas normas morais precisam ser repensadas no contexto da reprodução de IA, é suficiente considerar apenas uma propriedade exótica dos sistemas de IA: sua capacidade de

reprodução rápida. Dado o acesso ao *hardware* do computador, uma IA poderia duplicar-se muito rapidamente, em menos tempo que leva para fazer uma cópia do *software* da IA. Além disso, desde que a cópia do IA seja idêntica à original, ela nasceria completamente madura, e a cópia pode começar a fazer suas próprias cópias imediatamente. Na ausência de limitações de *hardware*, uma população de IA poderia, portanto, crescer exponencialmente em uma taxa extremamente rápida, com tempo de duplicação da ordem de minutos ou horas em vez de décadas ou séculos.

Nossas atuais normas éticas sobre a reprodução incluem uma versão de um princípio de liberdade reprodutiva, no sentido de que cabe a cada indivíduo ou ao casal decidir por si se quer e quantos filhos desejam ter. Outra norma que temos (pelo menos nos países ricos e de renda média) é que a sociedade deve intervir para prover as necessidades básicas das crianças nos casos em que seus pais são incapazes ou se recusam a fazê-lo. É fácil ver como essas duas normas poderiam colidir no contexto de entidades com capacidade de reprodução extremamente rápida.

Considere, por exemplo, uma população de *uploads*, um dos quais acontece de ter o desejo de produzir um clã tão grande quanto possível. Dada à completa liberdade reprodutiva, esse *upload* pode começar a copiar-se tão rapidamente quanto possível; e os exemplares que produz – que podem rodar em *hardware* de computador novo ou alugado pelo original, ou podem compartilhar o mesmo computador que o original, – também vão começar a se autocopiar, uma vez que eles são idênticos ao progenitor *upload* e compartilham seu desejo de produzir descendentes[9]. Logo, os membros do clã *upload* se encontrarão incapazes de pagar a fatura de eletricidade ou o aluguel para o processamento computacional e de armazenamento necessário para mantê-los vivos. Nesse ponto, um sistema de previdência social poderá ser acionado para fornecer-lhes, pelo menos, as necessidades básicas para sustentar a vida. Mas, se a população crescer mais rápido que a economia, os recursos vão se

[9] N.T.: No texto original, a palavra que aqui aparece é *philoprogenic*. Não encontramos palavra equivalente em português, por isso optamos por traduzi-la da forma mais aproximada possível da intenção dos autores. Por exemplo, a palavra *philoprogenitive* significa "produzindo muitos descendentes, amar um filho ou crianças em geral, relativo ao amor à prole". Por aproximação, podemos dizer que populações de *uploads* desejarão produzir mais descendentes.

esgotar, ao ponto de os *uploads* morrerem ou a sua capacidade para se reproduzir ser reduzida (Para dois cenários distópicos relacionados, veja Bostrom, [2004]).

Esse cenário ilustra como alguns princípios éticos de nível médio que são apropriados nas sociedades contemporâneas talvez precisem ser modificados se essas sociedades incluírem pessoas com a propriedade exótica de serem capazes de se reproduzir rapidamente.

O ponto geral aqui é que, quando se pensa em ética aplicada para contextos que são muito diferentes da nossa condição humana familiar, devemos ser cuidadosos para não confundir princípios éticos de nível médio com verdades normativas fundamentais. Dito de outro modo, devemos reconhecer até que ponto os nossos preceitos normativos comuns são implicitamente condicionados à obtenção de condições empíricas variadas e à necessidade de ajustar esses preceitos de acordo com casos hipotéticos futuristas nos quais suas precondições não são obtidas. Por isso, não estamos fazendo uma afirmação polêmica sobre o relativismo moral, mas apenas destacando o ponto de senso comum de que o contexto é relevante para a aplicação da ética e sugerindo que este ponto é especialmente pertinente quando se está considerando a ética de mentes com propriedades exóticas.

Superinteligência

Good (1965) estabeleceu a hipótese clássica sobre a superinteligência: que uma IA suficientemente inteligente para compreender a sua própria concepção poderia reformular-se ou criar um sistema sucessor, mais inteligente, que poderia, então, reformular-se novamente para tornar-se ainda mais inteligente, e assim por diante em um ciclo de *feedback* positivo. Good chamou isso de "explosão de inteligência". Cenários recursivos não estão limitados à IA: humanos com inteligência aumentada através de uma interface cérebro-computador podem reprogramar suas mentes para projetar a próxima geração de interface cérebro-computador. (Se você tivesse uma máquina que aumentasse o seu QI, ocorrer-lhe-ia, uma vez que se tornou bastante inteligente, tentar criar uma versão mais poderosa da máquina.).

Superinteligência também pode ser obtida por meio do aumento da velocidade de processamento. O mais rápido disparo de neurônios observado é de 1000 vezes por segundo; as fibras mais rápidas de axônio conduzem sinais a 150 metros/segundo, aproximadamente meio

milionésimo da velocidade da luz (SANDBERG, 1999). Aparentemente deve ser fisicamente possível construir um cérebro que calcula um milhão de vezes mais rápido que um cérebro humano, sem diminuir o seu tamanho ou reescrever o seu *software*. Se a mente humana for assim acelerada, um ano subjetivo do pensamento seria realizado para cada 31 segundos físicos no mundo exterior, e um milênio voaria em oito horas e meia. Vinge (1993) refere-se a essas mentes aceleradas como "superinteligência fraca": uma mente que pensa como um ser humano, mas muito mais rápida.

Yudkowsky (2008a) enumera três famílias de metáforas para visualizarmos a capacidade de um IA mais inteligente que humanos:

- Metáforas inspiradas pelas diferenças de inteligência individuais entre os seres humanos: IA patenteará novas invenções, publicará inovadores trabalhos de pesquisa, ganhará dinheiro na bolsa, ou formará blocos de poder político.

- Metáforas inspiradas pelas diferenças de conhecimento entre as civilizações humanas do passado e do presente: IA mais rápida inventará recursos que comumente futuristas prevêem para as civilizações humanas de um século ou milênio no futuro, como a nanotecnologia molecular ou viagens interestelares.

- Metáforas inspiradas pelas diferenças da arquitetura de cérebro entre humanos e outros organismos biológicos: Por exemplo, Vinge (1993): "Imagine executar uma mente de um cão a uma velocidade muito alta. Será que mil anos de vida do cão pode ser somada a qualquer percepção humana?" Isto é: alterações da arquitetura cognitiva podem produzir *insights* que nenhuma mente do nível humano seria capaz de encontrar, ou talvez até mesmo representar, após qualquer período de tempo.

Mesmo se nos limitarmos às metáforas históricas, torna-se claro que a inteligência sobre-humana apresenta desafios éticos que são literalmente sem precedentes. Nesse ponto, as apostas não são mais em escala individual (por exemplo, pedidos de hipotecas injustamente reprovados, casa incendiada, pessoa maltratada), mas em uma escala global ou cósmica (por exemplo, a humanidade é extinta e substituída

por nada que nós consideramos de valor). Ou, se a superinteligência pode ser moldada para ser benéfica, então, dependendo de suas capacidades tecnológicas, poderá trabalhar nos muitos problemas atuais que têm se revelado difícil para a nossa inteligência de nível humano.

Superinteligência é um dos vários "riscos existenciais", conforme definido por Bostrom (2002): um risco "onde um resultado adverso pode aniquilar permanentemente a vida inteligente originária da Terra ou limitar drasticamente o seu potencial". Por outro lado, um resultado positivo da superinteligência poderia preservar as formas de vida inteligentes originárias da Terra e ajudá-las a atingir o seu potencial. É importante ressaltar que as mentes mais inteligentes representam grandes benefícios potenciais, bem como riscos.

As tentativas de raciocinar sobre riscos catastróficos globais podem ser suscetíveis a uma série de bias[10] cognitivos (YUDKOWSKY, 2008b), incluindo o "bias da boa história" proposto por Bostrom (2002):

> *Suponha que nossas intuições sobre cenários futuros que são "plausíveis e realistas" são moldadas por aquilo que vemos na televisão, nos filmes e o que lemos nos romances. (Afinal, uma grande parte do discurso sobre o futuro que as pessoas encontram é em forma de ficção e outros contextos recreativos). Devemos então, quando pensar criticamente, suspeitar de nossas intuições de ser tendencioso no sentido de superestimar a probabilidade desses cenários que fazem uma boa história, uma vez que tais situações parecerão muito mais familiares e mais "reais". Esse bias da Boa história pode ser muito poderoso. Quando foi a última vez que você viu um filme sobre a humanidade em que os humanos são extintos de repente (sem aviso e sem ser substituído por alguma outra civilização)? Embora esse cenário possa ser muito mais provável do que um cenário no qual heróis humanos*

[10] N.T.: Um *bias* cognitivo é uma tendência inerente ao comportamento humano em cometer desvios sistemáticos de racionalidade, ao pensar ou analisar determinadas situações. Nossos mecanismos cognitivos (isto é, mecanismos de pensamento, raciocínio, inferência etc.) são enviesados, ou seja, viciados em determinadas direções, nos tornando mais propensos a cometer certos tipos de erros, por exemplo, de identificação ou de estimativa de tempo, probabilidades, etc.

> *com sucesso repelem uma invasão de monstros ou de guerreiros robôs, não seria muito divertido de assistir.*

Na verdade resultados desejáveis fazem filmes pobres: sem conflito, sem história. Enquanto as três leis da robótica de Asimov (ASIMOV, 1942) são muitas vezes citadas como um modelo de desenvolvimento ético para IA, as três leis são mais como um enredo para as tramas com os "cérebros positrônicos". Se Asimov tivesse representado as três leis como perfeitamente úteis, ele não teria obtido nenhuma história.

Seria um erro considerar sistemas de "IA" como uma espécie com características fixas e perguntar: "eles vão ser bons ou maus?" O termo "Inteligência Artificial" refere-se a um vasto espaço de projeto, provavelmente muito maior do que o espaço da mente humana (uma vez que todos os seres humanos compartilham uma arquitetura cerebral comum). Pode ser uma forma de "bias da boa história" perguntar: "Será que sistemas de IA são bons ou maus?", como se estivesse tentando pegar uma premissa para um enredo de filme. A resposta deve ser: "Exatamente sobre qual *design* de IA você está falando?"

Pode o controle sobre a programação inicial de uma Inteligência Artificial ser traduzido em influência sobre o seu efeito posterior no mundo? Kurzweil (2005) afirma que "inteligência é inerentemente impossível de controlar", e que, apesar das tentativas humanas de tomar precauções, "por definição... entidades inteligentes têm a habilidade de superara essas barreiras facilmente". Suponhamos que a IA não é apenas inteligente, mas que, como parte do processo de melhorar a sua própria inteligência, tenha livre acesso ao seu próprio código fonte: ela pode reescrever a si mesma e se tornar qualquer coisa que quer ser. No entanto, isso não significa que a IA deve *querer* se reescrever de uma forma hostil.

Considere Gandhi, que parece ter possuído um desejo sincero de não matar pessoas. Gandhi não conscientemente toma uma pílula que o leva a querer matar pessoas, porque ele sabe que se quiser matar as pessoas, provavelmente vai matar pessoas, e a versão atual do Gandhi não quer matar. Em termos mais gerais, parece provável que a maioria das mentes mais automodificadoras irão naturalmente ter funções de utilidade estável, o que implica que uma escolha inicial do projeto da mente pode ter efeitos duradouros (OMOHUNDRO, 2008).

Nesse ponto no desenvolvimento da ciência da IA, existe alguma maneira em que podemos traduzir a tarefa de encontrar um *design* para "bons" sistemas de IA em uma direção da pesquisa moderna? Pode parecer prematuro especular, mas pode-se suspeitar que alguns paradigmas de IA possuem mais probabilidade do que outros, para eventualmente provar que estão propícios à criação de agentes inteligentes de automodificação cujos objetivos continuem a ser previsíveis mesmo depois de várias interações de autoaperfeiçoamento. Por exemplo, o ramo bayesiano da IA, inspirado por coerentes sistemas matemáticos, como da teoria da probabilidade e da maximização da utilidade esperada, parece mais favorável para o problema de automodificação previsível do que a programação evolutiva e algoritmos genéticos. Essa é uma afirmação polêmica, mas ilustra o ponto que, se formos pensar no desafio da superinteligência, isso pode, na verdade, ser transformado em conselho direcional para as pesquisas atuais em IA.

No entanto, mesmo admitindo que possamos especificar um objetivo de IA a ser persistente sob automodificação e autoaperfeiçoamento, este só começa a tocar nos problemas fundamentais da ética para criação da superinteligência. Os seres humanos, a primeira inteligência geral a existir na Terra, têm usado essa inteligência para remodelar substancialmente a escultura do globo – esculpir montanhas, domar os rios, construir arranha-céus, agricultura nos desertos e produzir mudanças climáticas não intencionais no planeta. Uma inteligência mais poderosa poderia ter correspondentemente maiores consequências.

Considere novamente a metáfora histórica para a superinteligência – diferenças semelhantes às diferenças entre as civilizações passadas e presentes. Nossa civilização atual não está separada da Grécia Antiga somente pela ciência aperfeiçoada e aumento da capacidade tecnológica. Há uma diferença de perspectivas éticas: os gregos antigos pensavam que a escravidão era aceitável, nós pensamos o contrário. Mesmo entre os séculos XIX e XX, houve substanciais divergências éticas – as mulheres devem ter direito ao voto? Os negros podem votar? Parece provável que as pessoas de hoje não serão vistas como eticamente perfeitas por civilizações futuras, não apenas por causa da nossa incapacidade de resolver problemas éticos reconhecidos atualmente, como a pobreza e a desigualdade, mas também por nosso fracasso até mesmo em reconhecer alguns problemas dessa natureza. Talvez um dia o ato de sujeitar as crianças involuntariamente à escolaridade poderá ser visto como abuso

infantil como também permitir que elas deixem a escola aos 18 anos. Nós não sabemos.

Considerando a história da ética nas civilizações humanas ao longo dos séculos, podemos ver que poderia tornar uma tragédia muito grande criar uma mente que ficou estável em dimensões éticas ao longo da qual as civilizações humanas parecem exibir *mudança direcional*. E se Arquimedes de Siracusa tivesse sido capaz de criar uma inteligência artificial de longa duração com uma versão estável do código moral da Grécia Antiga? Mas evitar esse tipo de estagnação ética é comprovadamente complicado: não seria suficiente, por exemplo, simplesmente tornar a mente aleatoriamente instável. Os gregos antigos, mesmo que tivessem percebido suas próprias imperfeições, não poderiam ter feito melhor mesmo jogando dados. Ocasionalmente uma boa e nova ideia em ética vem acompanhada de uma surpresa, mas a maioria gerada aleatoriamente traz mudanças éticas que nos parecem loucura ou rabiscos incompreensíveis.

Isso nos apresenta talvez o último desafio das máquinas éticas: Como construir uma IA que, quando executada, torna-se mais ética do que você? Isso não é como pedir a nossos próprios filósofos para produzir uma superética, mais do que o *Deep blue* foi construído fazendo com que os melhores jogadores humanos de xadrez programassem boas jogadas. Mas temos de ser capazes de efetivamente descrever a questão, se não a resposta – jogar dados não irá gerar bons movimentos do xadrez, muito menos uma boa ética. Ou, talvez, uma maneira mais produtiva de pensar sobre o problema: Que estratégia você gostaria que Arquimedes seguisse na construção de uma superinteligência, de modo que o resultado global ainda fosse aceitável, se você não pudesse lhe dizer especificamente o que estava fazendo de errado? Essa é a situação em que estamos em relação ao futuro.

Uma parte forte do conselho que emerge considerando nossa situação análoga à de Arquimedes é que não devemos tentar inventar uma versão "super" do que nossa civilização considera ética, essa não é a estratégia que gostaríamos que Arquimedes seguisse. Talvez a pergunta que devêssemos considerar é como uma IA programada por Arquimedes – sem maior experiência moral do que Arquimedes – poderia reconhecer (pelo menos algumas) nossa própria civilização ética como progresso moral, em oposição à simples instabilidade moral? Isso exigiria que

começássemos a compreender a estrutura de questões éticas da maneira que já compreendemos a estrutura do xadrez.

Se somos sérios sobre o desenvolvimento de uma IA avançada, esse é um desafio que devemos enfrentar. Se as máquinas estão a ser colocadas em posição de mais fortes, mais rápidas, mais confiáveis, ou mais espertas que os humanos, então a disciplina de máquinas éticas deve se comprometer a buscar refinamento humano superior (e não apenas seres humanos equivalentes)[11].

Conclusão

Embora a IA atual ofereça-nos algumas questões éticas que não estão presentes no *design* de automóveis ou de usinas de energia, a abordagem de algoritmos de inteligência artificial em relação a um pensamento mais humano prenuncia complicações desagradáveis. Os papéis sociais podem ser preenchidos por meio de algoritmos de IA, o que implica novas exigências de projeto como transparência e previsibilidade. Suficientemente algoritmos de IAG já não podem executar em contextos previsíveis, exigem novos tipos de garantia de segurança e engenharia e de considerações da ética artificial. Sistemas de IA com estados mentais suficientemente avançados, ou o tipo certo de estados, terão um *status* moral e alguns podem ser considerados como pessoas – embora talvez pessoas muito diferentes do tipo que existem agora, talvez com regras diferentes. E, finalmente, a perspectiva de IA com inteligência e habilidades sobre-humanas nos apresenta o desafio extraordinário de indicar um algoritmo que gere comportamento superético. Esses desafios podem parecer visionários, mas parece previsível que os encontraremos, e eles não são desprovidos de sugestões para os rumos da pesquisa atual.

Leituras

BOSTROM, N. *The future of human evolution*. In: *Death and anti-death: Two hundred years after Kant, fifty years after Turing*, ed. Charles Tandy (Palo Alto, Califórnia: Ria *University Press, 2004*). Esse trabalho explora algumas dinâmicas evolutivas que poderiam levar a uma população de *uploads* para se desenvolverem em direções distópicas.

[11] Os autores são gratos a Rebecca Roache pelo auxílio à pesquisa e aos editores deste volume por comentários detalhados a uma versão anterior do nosso manuscrito.

YUDKOWSKY, E. *Artificial intelligence as a positive and negative factor in global risk*, in Bostrom and Cirkovic (eds.), 2008a, p. 308-345. Uma introdução aos riscos e desafios apresentados pela possibilidade de melhorar a autorrecursividade das máquinas superinteligentes.

WENDELL, W. 2008. *Moral Machines: Teaching Robots Right from Wrong*. Oxford University Press, 2008. – Uma pesquisa global de desenvolvimento recente.

Referências

ASIMOV, I. *Runaround*. In: *Astounding Science Fiction*, March 1942.

BEAUCHAMP, T.; CHILRESS, J. *Principles of biomedical ethics*. Oxford: Oxford University Press, 2008.

BOSTROM, N. *Existential risks: analyzing human extinction scenarios, Journal of Evolution and Technology 9*. Disponível em: http://www.nickbostrom.com/existential/risks.html.

_____. Astronomical waste: *The opportunity cost of delayed technological development*, *Utilitas* 15: 2003 p. 308-314.

_____. *The future of human evolution*, In: *Death and anti-death*: Twohundred years after Kant, fifty years after Turing, ed. Charles Tandy. Palo Alto, Califórnia: Ria University Press, 2004. Disponível em: http://www.nickbostrom.com/fut/evolution.pdf

BOSTROM, N. & CIRKOVIC, M. (eds.). *Global catastrophic risks*. Oxford: Oxford University Press, 2007.

CHALMERS, D. J. *The conscious mind*: In *Search of a fundamental theory*. New York and Oxford: Oxford University Press, 1996.

HIRSCHFELD, L. A.; GELMAN, S. A. (eds.). *Mapping the mind: domain specificity in cognition and culture*, Cambridge: Cambridge University Press, 1994.

GOERTZEL, B. PENNACHIN, C. (eds.). *Artificial general intelligence*. New York: Springer-Verlag, 2006.

GOOD, I. J. *Speculations concerning the first ultraintelligent machine*. In: Advances in computers (Academic Press). New York: Academic Press 6: 31–

88, 1965.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The elements of statistical learning*. New York: Springer Science, 2001

HENLEY, K. *Abstract principles, mid-level principles, and the rule of law*. In: *Lawand Philosophy* 12, pp. 121-32, 1993.

HOFSTADTER, D. *Trying to muse rationally about the singularity scenario*. In: *Singularity Summit at Stanford*, 2006.

HOWARD, Philip K. *The death of common sense: How law is suffocating america*. New York: Warner Books,1994.

KAMM, F. *Intricate ethics: r*ights, responsibilities, and permissible harm. Oxford: Oxford University Press, 2007.

KURZWEIL, R. *The singularity is near*: w*hen humans transcend biology*. New York: Viking, 2005.

McDERMOTT, D. *Artificial intelligence meets natural stupidity*, *ACM SIGART Newsletter* 57: 1976, p. 4-9.

OMOHUNDRO, S. *The basic AI drives*. In: *Proceedings of the AGI-08 workshop*. Amsterdam: IOS Press. 2008, p. 483-492.

SANDBERG, A. *The physics of information processing superobjects: Daily Life Among the Jupiter Brains*. In: *Journal of Evolution and Technology*, 5, 1999.

VINGE, V. *The coming technological singularity*, presented at the *VISION-21Symposium*, march, 1993.

WARREN, M. E. *Moral status: obligations to persons and other living things*. Oxford: Oxford University Press, 2000

YUDKOWSKY, E. *AI as a precise art*, presented at the *2006 AGI Workshop* in Bethesda, MD.

_____. *Artificial intelligence as a positive and negative factor in global risk*. In: Bostrom and Cirkovic (eds.), 2008a, p. 308-345.

_____. *Cognitive biases potentially affecting judgment of global risks*. In: Bostrom and Cirkovic (eds.), 2008b, p. 91-119.

Esta revista foi impressa pela Gráfica da Universidade Federal de Ouro Preto em sistema de impressão digital, em dezembro de 2012. A fonte usada no miolo é Cambria, corpo 10. O papel do miolo é pólen soft 80g/m2 e o da capa é cartão 250g/m2.