

Reflexões sobre alguns problemas filosóficos

Gerson Albuquerque de Araujo Neto
Org.



Editora Fundação Fênix

O presente livro, composto por textos de diversos filósofos com formação e pesquisas distintas é um amplo tesouro de estudos e reflexões sobre diversas questões atuais na filosofia. Os textos abordam questões que vão da epistemologia, lógica, filosofia da técnica, filosofia da linguagem, da ética, da filosofia política e da filosofia da mente. Portanto, é uma obra que interessa a leitores diversos.



Editora Fundação Fênix



Reflexões sobre alguns problemas filosóficos

Série Filosofia

Conselho Editorial

Editor

Agemir Bavaresco

Conselho Científico

Agemir Bavaresco – Evandro Pontel

Jair Inácio Tauchen – Nuno Pereira Castanheira

Conselho Editorial

Augusto Jobim do Amaral

Cleide Calgaro

Draiton Gonzaga de Souza

Evandro Pontel

Everton Miguel Maciel

Fabián Ludueña Romandini

Fabio Caprio Leite de Castro

Fabio Caires Coreia

Gabriela Lafetá

Ingo Wolfgang Sarlet

Isis Hochmann de Freitas

Jardel de Carvalho Costa

Jair Inácio Tauchen

Jozivan Guedes

Lenno Francisco Danner

Lucio Alvaro Marques

Nelson Costa Fossatti

Norman Roland Madarasz

Nuno Pereira Castanheira

Nythamar de Oliveira

Orci Paulino Bretanha Teixeira

Oneide Perius

Raimundo Rajobac

Renata Guadagnin

Ricardo Timm de Souza

Rosana Pizzatto

Rosalvo Schütz

Rosemary Sadami Arai Shinkai

Sandro Chignola

Thadeu Weber

Gerson Albuquerque de Araujo Neto
Organizador

Reflexões sobre alguns problemas filosóficos



Editora Fundação Fênix

Porto Alegre, 2023

Direção editorial: Agemir Bavaresco
Diagramação: Editora Fundação Fênix
Capa: Editora Fundação Fênix

O padrão ortográfico, o sistema de citações, as referências bibliográficas, o conteúdo e a revisão de cada capítulo são de inteira responsabilidade de seu respectivo autor.

Todas as obras publicadas pela Editora Fundação Fênix estão sob os direitos da Creative Commons 4.0 –

[Http://creativecommons.org/licenses/by/4.0/deed.pt_BR](http://creativecommons.org/licenses/by/4.0/deed.pt_BR)



Série Filosofia – 124

Catálogo na Fonte

R332 Reflexões sobre alguns problemas filosóficos [recurso eletrônico] / Gerson Albuquerque de Araújo Neto Organizador. – Porto Alegre : Editora Fundação Fênix, 2023.
115 p. (Série Filosofia ; 124)

Disponível em: <<http://www.fundarfenix.com.br>>

ISBN 978-65-5460-106-1

DOI <https://doi.org/10.36592/9786554601061>

1. Filosofia. 2. Igreja. I. Araújo Neto, Gerson Albuquerque de (org.).

CDD: 100

Responsável pela catalogação: Lidiane Corrêa Souza Morschel CRB10/1721

SUMÁRIO

INTRODUÇÃO	9
1. O CONCEITO DE SUPERINTELIGÊNCIA DE NICK BOSTROM	11
<i>Gerson Albuquerque de Araujo Neto</i>	
2. O PROBLEMA DOS OBJETOS ARBITRÁRIOS: REFLEXÕES SOBRE A REGRA DA INTRODUÇÃO DO QUANTIFICADOR UNIVERSAL E DA ELIMINAÇÃO DO QUANTIFICADOR EXISTENCIAL	23
<i>Martim Mottloch</i>	
3. UM NAVIO RECONSTRUÍDO EM ALTO MAR: CONSIDERAÇÕES SOBRE O SUPOSTO PRAGMATISMO DE QUINE EM 'DOIS DOGMAS DO EMPIRISMO'	33
<i>Maria Cristina de Távora Sparano</i> <i>Leonardo Bruno Vieira Santos</i>	
4. LIBERDADE E MORALIDADE EM KANT	51
<i>Fernando Facó de Assis Fonseca</i>	
5. CARL SCHMITT, UM HEGELIANO? A LEITURA CRÍTICA DE JEAN-FRANÇOIS KERVÉGAN SOBRE O POLÍTICO, A GUERRA E O ESTADO TOTAL	71
<i>Deyvison Rodrigues Lima</i> <i>José Maria Arruda</i>	
6. SOBRE A POSSIBILIDADE DE UMA CONSCIÊNCIA ARTIFICIALMENTE MORAL	89
<i>Dayvide Magalhães de Oliveira</i>	

INTRODUÇÃO

O presente livro é composto de uma série de estudos e reflexões sobre alguns problemas filosóficos que no momento preocupam a humanidade. A variedade da formação dos autores enriquece bastante as discussões.

O primeiro artigo trata de uma reflexão sobre o conceito de superinteligência do filósofo Nick Bostrom. No momento em que a Inteligência Artificial toma espaço nas discussões este texto vem contribuir para esclarecer e levar os leitores novos questionamentos sobre este assunto.

O capítulo dois, escrito pelo professor Martin Mottloch, com doutorado na Alemanha, trata do Problema dos Objetos Arbitrários: reflexões sobre a regra da introdução do quantificador universal e da eliminação do quantificador existencial.

No capítulo três a professora aposentada da Universidade Federal do Paraná com doutorado sanduíche no Canadá e pós doutorado na Suíça escreve junto com seu orientando de mestrado Leonardo Bruno Vieira Santos o texto Um Navio Reconstruído em Alto Mar: considerações sobre pragmatismo de Quine em Dois Dogmas do Empirismo.

No capítulo seguinte Fernando Facó de Assis Fonseca, doutor em Filosofia da Educação pela Universidade Federal do Ceará, nos apresenta uma reflexão sobre a Liberdade e Moralidade em Kant. Tema bastante atual no nosso momento político e social.

Na mesma linha da reflexão social e política, o professor Deyvison Rodrigues Rodrigues, doutor em Filosofia pela UFRJ e atualmente Chefe do Departamento de Filosofia da UFPI escreve em coautoria com José Maria Arruda professor da Universidade Federal Fluminense o texto Carl Schmitt, um hegeliano? A leitura crítica de Jean-François Kervégan sobre o político, a guerra e o Estado Atual.

O último texto é uma produção do professor da UFPI, Dayvide Magalhães de Oliveira e trata Sobre a Possibilidade de Uma Consciência Artificialmente Moral.

Acreditamos que este livro leve ao leitor a uma gama de informações, bem como matéria para uma série de belíssima e ricas reflexões.

Prof. Gerson Albuquerque de Araujo Neto.

1. O CONCEITO DE SUPERINTELIGÊNCIA DE NICK BOSTROM



<https://doi.org/10.36592/9786554601061-01>

Gerson Albuquerque de Araujo Neto¹

O objetivo deste texto é fazer alguns questionamentos sobre o conceito de superinteligência apresentado pelo filósofo Nick Bostrom. Além das ideias de Bostrom serão apresentadas algumas críticas de Roger Penrose, Miguel Nicolelis, João de Fernando Teixeira e algumas considerações minhas.

Nick Bostrom, filósofo sueco, atualmente professor na Universidade de Oxford. Nasceu Helsinboug, na Suécia, no dia 10 de março de 1973.

Bostrom tem se preocupado em refletir sobre os avanços tecnológicos e suas implicações. No seu livro Superinteligência (p. 16) ele nos diz que naquele "livro, procuro entender o desafio que surge com a perspectiva da superinteligência e de que maneiras poderíamos melhor responder a ele". Diz mais Bostrom sobre a superinteligência "esse é, muito provavelmente, o desafio mais importante e mais assustador que a humanidade já encarou. E, independentemente do nosso sucesso ou fracasso, promete ser o último desafio que encararemos". (p. 16).

O presente texto tem o objetivo de discutir as implicações filosóficas do conceito de Superinteligência proposto por Bostrom.

INÍCIO DO PROBLEMA

Um ponto importante na história da computação é o questionamento apresentado pelo matemático britânico Alan Turing. Turing nasceu a 23 de junho de 1912 e faleceu aa 7 de junho de 1954.

¹ Pós-Doutor em Filosofia pela UERJ e professor da UFPI.

12 | Reflexões sobre alguns problemas filosóficos

Ele elaborou um teste que ficou conhecido como teste de Turing. Neste teste consiste em uma máquina conseguir fazer tudo que um cérebro humano é capaz de fazer.

O questionamento que Nick Bostrom faz é: Será que depois de termos uma máquina que passa no teste de Turing, ou seja, uma máquina com a mesma inteligência do ser humano, poderemos ter uma máquina com uma inteligência superior a nossa. Uma superinteligência.

No livro Superinteligência, na página 51, Bostrom nos define o que é uma superinteligência como “qualquer intelecto que exceda em muito o desempenho cognitivo dos seres humanos em, virtualmente, todos os domínios de interesse”. Ele alerta que esta definição de superinteligência é provisória e que ele ainda vai, no decorrer do seu livro, dar outras definições de superinteligência e que esta é provisória.

Diz Bostrom, na página 127 do mesmo livro “Considerando que as máquinas, *em algum momento no futuro*, ultrapassarão consideravelmente a biologia em termos de inteligência geral”. Ele prossegue e afirma que neste momento a máquina superará a inteligência humana.

SUPERINTELIGÊNCIA

Como será esta superinteligência? Um computador? Um robô? Vamos fazer uma distinção entre robô, androide e cyborg. Um robô é um equipamento com a capacidade de executar tarefas humanas. Não necessariamente com forma humana. Etimologicamente vem da língua checa que significa servidão. Foi utilizada pela primeira vez pelo escritor checo Karel Capek.

Androide é um autônomo que tem forma de ser humano.

Cyborg é um ser dotado de partes humanas e cibernéticas. É resultado das palavras inglesas cyber(netics) com org(anism). Foram utilizadas pela primeira vez por Manfred Clynes e Nathan Kline.

Ao iniciar o capítulo quatro de seu livro Bostrom questiona “Uma vez que as máquinas tenham alcançado alguma forma de equivalência humana na habilidade

de raciocínio, quanto tempo levaria, então até que elas atingissem uma superinteligência radical?". (Bostrom, 2018, p 127).

Após esta pergunta Bostrom se questiona em quantos anos isto ocorrerá. Bostrom faz uma apresentação dos avanços da IA sobre diversos jogos. Relata a data em que a máquina venceu os seres humanos em diversos tipos de jogos. Apresenta, talvez o caso mais divulgado na mídia que é o computador Deep Blue que venceu, em 1993, o campeão mundial de xadrez Gary Kasparov. Mas relata outros casos, como o do programa de computador elaborado por Arthur Samuel, em 1952 e aperfeiçoado em 1955 que foi o primeiro programa em que a máquina conseguia vencer seu criador num jogo. No caso o jogo de damas.

Bostrom apresenta três formas de superinteligência: a superinteligência rápida; a superinteligência coletiva e a superinteligência de qualidade.

Define Bostrom que a superinteligência rápida (p.108) "um sistema que pode fazer tudo o que o intelecto humano é capaz de fazer, porém muito mais rapidamente.

A superinteligência coletiva, segundo Bostrom (p.111), é "um sistema composto de um grande número de intelectos menores, de forma que o seu desempenho total supere significativamente, em diversas áreas gerais do conhecimento, qualquer sistema cognitivo atual." Esta definição nos faz lembrar Pierre Lévy e Miguel Nicolelis. O primeiro em sua obra 'Inteligência Coletiva' e o segundo com alguns livros sobre o cérebro humano. Este tipo de superinteligência nos leva a inferir que ao tratar de problemas que podem ser decomposto em partes.

Já a superinteligência de qualidade é definida por Bostrom (p.115) como "um sistema que é no mínimo tão rápido quanto uma mente humana e qualitativamente muito mais inteligente".

QUESTIONAMENTO DE ROGER PENROSE

Um dos mais importantes pensadores da atualidade, Roger Penrose nasceu em 8 de agosto de 1931 na cidade de Colchester, na Inglaterra. Seus trabalhos são importantes para a física, para a matemática, para a cosmologia, para a filosofia da ciência e para os estudos da inteligência artificial.

14 | Reflexões sobre alguns problemas filosóficos

Penrose tem estudado e esclarecido vários pontos na física. Tem participado de debates sobre diversos pontos da ciência. Suas ideias e discussões tem implicações epistemológicas, sobretudo na filosofia da ciência e na filosofia da mente.

Um dos assuntos de grande interesse e importância para a filosofia é a questão da possibilidade de uma máquina que consiga reproduzir todas as atividades e ações de um cérebro humano. O lógico e matemático inglês, professor da Universidade de Oxford, Alan Turing propôs um desafio chamado Teste de Turing que consiste do desafio de uma máquina que reproduza todas as funções de um cérebro humano.

Até os dias atuais não foi possível a construção de uma máquina que passasse completamente no teste. Alguns defendem que é questão de tempo a construção de uma máquina que passe totalmente no teste. Estes são os chamados defensores do programa chamado Inteligência Artificial Forte. Contudo, há outros pensadores que defendem que isto nunca será possível. Um deles é Roger Penrose.

Uma das questões bastante discutida, mas não resolvida na filosofia é a questão da mente humana. Existe até uma área da filosofia hoje chamada filosofia da mente e diversas são as correntes nesta área como o dualismo, monismo e interacionismo.

A questão da mente humana ocupa a atenção e a preocupação de vários filósofos no Brasil e no mundo. No Brasil a comunidade filosófica criou um Grupo de Trabalho (GT) de Filosofia da Mente na Associação Nacional dos Programas de Pós-Graduação em Filosofia (ANPOF). Há várias publicações na comunidade filosófica no Brasil discutindo as diversas questões ligadas a mente humana.

Roger Penrose tem três grandes livros abordando a mente humana. São eles "Shadow of Mind", "O Grande, o pequeno e a mente humana" e "A Mente nova do rei: computadores, mentes e as leis da física".

Na obra "A Mente Nova do Rei: computadores, mentes e as leis da física", Penrose nega a possibilidade de uma máquina reproduzir todas as atividades do cérebro humano. Ocorre que os avanços da neurologia não conseguem explicar completamente como o cérebro humano funciona, então talvez esta conclusão, ainda,

não possa ser alcançada. A resposta ao teste de Turing exige um diálogo com os avanços da neurologia, bem como com os avanços da computação e da própria física.

No prefácio do livro *A Nova mente do rei* escreve Martin Gardner sobre as ideias de Penrose desenvolvidas no livro: "Embora os capítulos de Penrose abranjam um amplo campo, que passa pela teoria da relatividade, a mecânica quântica e a cosmologia, seu interesse central é o que os filósofos chamam de 'problema corpóreo'". (PENROSE, 1991)

Não podemos negar que houve avanço muito grande na computação. Além dos já relatados por Bostrom, a computação avançou bastante com as machine learn. Estas são máquinas que possuem programas que podem aprender e melhorar sem necessidade de um programador externo.

Também o desenvolvimento de programas com rede neural. Estes tentam imitar o cérebro. O programa IA chamado ChatGPT é consequência deste tipo de programa neural.

Contudo os avanços da neurociência por mais progressos tenham sido feitos nos últimos 150 anos e com o avanço dos equipamentos de neuro imagem ainda não conseguem explicar completamente como o cérebro é. Por exemplo a área de Wernick e área de Broca que são regiões do cérebro onde há o entendimento da linguagem e da fala estão reproduzidos pelas máquinas.

O físico inglês ganhador do prêmio Nobel de 2021, Roger Penrose no seu livro *A Nova Mente do Rei* levanta uma série de questionamentos sobre a IA Forte que é uma Inteligência Artificial que supere a mente humana.

Não é a toa que Penrose escreveu um outro livro após aquele chamado *Shadows of the Mind*, traduzido no Brasil como *sombras da mente*.

Neste livro Penrose faz uma série de questionamentos sobre o que é a mente. Ele chega dizer sobre o alcance que a ciência pode ter sobre a mente. No livro ele escreve "Qual é o objetivo final da ciência? É somente entender os atributos materiais do nosso universo que são suscetíveis aos seus métodos, enquanto a nossa existência mental permanece eternamente além do seu alcance? Ou será que um dia teremos um bom entendimento científico do obscuro mistério da mente? Será que o fenômeno da consciência humana é algo que está além do escopo da investigação

16 | Reflexões sobre alguns problemas filosóficos

científica ou será que o poder do método científico um dia será capaz de resolver o problema da própria existência de nossas experiências conscientes?". (p. 31).

Vai além Penrose "eu diria que ainda não há teoria física, biológica ou computacional que chegue perto de explicar nossa consciência e conseqüentemente inteligência; mas que isso não deve nos impedir de procurar por uma.". (p. 32).

QUESTIONAMENTOS

Porque comportamentos mentais diferentes? Tais como toc, apego, caridade, auto-eliminação. As máquinas superinteligentes também terão estes problemas.? Se não sabemos o que é mente humana e como realmente esta funciona, como poderemos dizer que ela faz ou é melhor que a mente humana? Se eu não sei o que é B, como posso dizer que $A > B$

Mais ao término do seu livro Bostrom afirma: "Desde a publicação da primeira edição deste livro, ocorreram algumas mudanças de atitude. Ficou mais fácil tratar a superinteligência como um tópico importante – levar a sério a visão de que a transição para inteligência de máquina pode acontecer neste século, que tal transição pode entrar os eventos mais importantes da história da humanidade, que ela pode vir acompanhada de uma certa dose de risco existencial, embora também possa ser extremamente positiva, e que seria sensato adiantar um pouco do trabalho de buscar entender se há algo que deveríamos estar fazendo para aumentar as chances de um resultado favorável."

No livro Superinteligência, Nick Bostrom (2018, p. 18) nos diz "máquinas com inteligência geral comparável à dos humanos – ou seja, dotada de bom senso e capacidade real de aprender, raciocinar e planejar a superação de desafios complexos de processamento de informação em uma vasta gama de domínios naturais e abstratos – tem sido esperadas desde a invenção dos computadores, na década de 1940. Naquele tempo, o advento de tais máquinas era frequentemente esperado para os vinte anos seguintes". Contudo isto não ocorreu.

Bostrom continua "desde então, a data estimada para o seu surgimento tem recuado numa razão de um ano a cada ano, fazendo com que ainda hoje futuristas

interessados na possibilidade de uma inteligência artificial geral acreditem que máquinas inteligentes surgirão dentro de duas décadas".

Resta-nos aguardar os próximos passos dos avanços na área da computação e mais reflexões na filosofia surgirão.

QUESTIONAMENTOS DE MIGUEL NICOLELIS

"Do ponto de vista científico, digo isso há anos, e agora Noam Chomsky usa a mesma frase, a inteligência artificial não é nem inteligente e nem artificial. Não é artificial por nós, é natural. E não é inteligente porque a inteligência é uma propriedade emergente de organismos interagindo com o ambiente e com outros organismos. É um produto do processo darwiniano de seleção natural. O algoritmo pode andar e fazer coisas, mas não inteligentes por definição. Se estivesse vivo Charles Darwin teria um infarto com isso". Este é um questionamento Miguel Nicolelis, neurocientista, que trabalha há 30 anos com redes neurais mecanismo por trás dos algoritmos de aprendizagem de máquina. Para ele, a inteligência é o resultado de milhões de anos de evolução, que não podem ser computados em código binário. Folha de São Paulo, 09-07-2023

Para Nicolelis a inteligência humana não é binária. E continua Nicolelis "O aprendizado de máquina, deep learning, machine learning, são grandes nomes que usam palavras que nós nos acostumamos coloquialmente a usar, relacionadas ao cérebro humano ou de qualquer animal, para definir coisas que nós fazemos com lógica binária. A inteligência humana não é binária. Por isso, é um nome impróprio". Para ele, a inteligência é o resultado de um processo que demorou milhões de anos de evolução, que não podem ser computados em código binário - Folha de São Paulo, 09-07-2023.

Nicolelis fazendo crítica a Inteligência Artificial levanta certos questionamentos como: "A mente humana é repleta de fenômenos não computáveis: inteligência, intuição, criatividade, senso estético, definições de beleza, de criatividade, tudo isso não é computável. Qual a fórmula para a beleza?"

No prefácio do livro Neurociências escrito por Dale Purves e outros encontramos "O sistema nervoso humano, seja em termos moleculares, celulares,

18 | Reflexões sobre alguns problemas filosóficos

sistêmicos, comportamentais ou cognitivos, é um estupendo espécie de maquinaria biológica. Levando-se em conta suas conquistas – todos os artefatos da cultura humana, por exemplo –, estamos repletos de motivos para querer compreender como funciona o encéfalo e o restante do sistema nervoso. Os efeitos debilitadores e de alto custo das doenças neurológicas e psiquiátricas reforçam o sentido de urgência desse conhecimento”.

Segundo ainda o prefácio do mesmo livro na neurociência estão “... os desafios intelectuais e o fascínio – bem como as incertezas – daquilo que muitos veem como sendo a última grande fronteira das ciências biológicas”.

É dito também no prefácio do citado livro: “Como qualquer grande outro desafio, as neurociências deveriam ser – e são! – cheias de polêmicas, divergências e muita diversão”.

No primeiro capítulo do livro é dito “...muitas questões ainda permanecem não respondidas”.

A neurociência é uma ciência que para ser trabalhada com eficiência precisa ser trabalhada de forma inter(transdisciplinar). Seu estudo envolve biologia, psicologia, psiquiátrica, fisiologia, química ou bioquímica, computação, física, matemática e outras. Portanto, não é uma área fácil de se trabalhar.

Os estudos sobre o cérebro humano sofreram um impulso formidável nos anos 90 do século XX, isto porque entraram em ação equipamentos de neuroimagens. Muito se aprendeu e se descobriu sobre o funcionamento com estes avanços. O filósofo João de Fernandes Teixeira afirma que estes anos pode ser chamado de “a década do cérebro”.

As PET sigla de Positron Emission Tomography e fMRI sigla para Functional Magnetic Resonance Imaging possibilitaram um avanço formidável no estudo do cérebro.

Conforme Purves “já no início do século XIX, a célula foi reconhecida como unidade fundamental de todos os organismos vivos. No entanto, foi apenas mais recentemente – durante o século XX – que os neurocientistas chegaram a um consenso de que o tecido nervoso, como os demais órgãos, também é constituído por essas unidades fundamentais”.

Pelas dificuldades e limitações dos equipamentos microscópicos da época "alguns biólogos da época concluíram que cada célula nervosa estava conectada a suas vizinhas por uniões protoplasmáticas, formando uma malha contínua de neurônios, o 'retículo' (do latim reticulum). Foi o patologista italiano Camilo Golgi quem articulou e defendeu essa 'teoria reticular' da comunicação de células nervosas". Embora Golgi tenha produzido e contribuído para o conhecimento médico esta sua teoria estava equivocada.

Uma teoria rival apareceu para substituir a teoria de Golgi. Ela foi proposta pelo espanhol Santiago Ramón y Cajal e pelo inglês Charles Sherrington e foi chamada de doutrina neuronal.

Nesta teoria é defendido "que as células nervosas são entidades distintas e que se comunicam por meio de contatos especializados, que Sherrington chamou de sinapses".

As células do sistema nervoso podem ser classificadas em células nervosas, também chamadas neurônios e células neurogliais, também chamadas glia.

O sistema nervoso é estruturado em dois: autônomo e somático. E o sistema nervoso é dividido em: sistema nervoso central e sistema nervoso periférico. O sistema nervoso central é formado pelo encéfalo e medula espinhal. O encéfalo é composto por cérebro, cerebelo, tálamo, hipotálamo e bulbo.

O sistema nervoso autônomo é composto pelo sistema simpático e parassimpático.

A complexidade do cérebro leva a uma dificuldade de explicar como ele é e como funciona.

Segundo Nicolelis "... o que em geral é apresentado, sobretudo na última década, é um argumento quase pueril, como o que abriu este capítulo, defendendo a ideia de que, para reproduzir as capacidades mentais altamente complexas do cérebro, basta conectar bilhões de transistores criados a semelhança dos neurônios energeticamente eficientes e depois apertar o botão 'on'". (Nicolelis, 1989, p. 139)

Nicolelis responde "Eu certamente discordo desta simplificação".

E Nicolelis acrescenta "A noção que o funcionamento intrínseco do cérebro humano pode ser reduzido a um algoritmo computacional para ser reproduzido em lógica digital pode ser considerada mais um dos mitos do mundo pós-moderno; uma

espécie de lenda urbana ou mesmo um exemplo típico da era da pós-verdade, tempo em que uma declaração falsa – repetida inúmeras vezes e disseminada amplamente na sociedade pela mídia ou pelas redes sociais – passa a ser aceita como verdade. A hipótese de que o grau de complexidade inerente ao cérebro pode ser recriado simplesmente ao se conectar um número extraordinariamente grande de elementos eletrônicos eficientes não só está muito longe de ser uma realidade científica, como, quando examinada em detalhes, não revela nenhuma chance crível de ser bem-sucedida”. (1989, p. 139).

Segundo Nicoletis, ele “introduz uma noção fundamental para a minha tese central: uma nova definição operacional para o conceito de informação, que chamei de ‘informação gödeliana’, podendo ser manipulada por tecidos orgânicos e cérebros animais como o nosso”. Continua Nicoletis que no seu livro “essa discussão é seguida pela descrição de uma série de dez princípios fisiológicos fundamentais que regem a operação do cérebro humano. Foi com base nesses dez princípios, descobertos ao longo dos meus trinta e sete anos de pesquisas neurofisiológicas, que foi criada a Teoria do Cérebro Relativístico. Desses princípios, portanto, deriva a principal predição feita pela TCR: de que o cérebro humano sempre opera como um todo, de forma contínua; ou seja, em vez de usar uma localização espacial restrita do tecido neural para executar cada uma das atividades mentais, o Verdadeiro Criador de Tudo se vale do trabalho altamente sincronizado de múltiplas regiões cerebrais, distribuídas por todo o seu volume, para produzir cada uma das suas funções neurológicas e cada um dos nossos comportamentos”.

Teixeira assume uma postura negativa a IA forte, pois ele diz “Será que, em algum momento poderemos suplantar a incomensurabilidade do cérebro humano?” E ele responde: “É bem provável que não.” Mas ele diz “Isso não quer dizer que não possamos, algum dia, simular artificialmente a inteligência humana através das máquinas.”

CONSIDERAÇÕES FINAIS

Pela discussão deste texto pode-se concluir que a questão de uma máquina que execute todas as funções de um cérebro humano ainda está em aberto. Muito, ainda, tem-se a aprender sobre o cérebro humano.

A rapidez com que o setor ligado a computação avança é assombrosa. Esta nos leva a inferir que novidades aparecerão nos próximos anos.

Estes avanços trarão desafios enormes para a reflexão. Mas surgirá uma máquina de Turing inquestionável. E os avanços da tecnologia nos levarão a uma máquina superinteligente como fala Nick Bostrom.

E quem regulará esta superinteligência? E como será as relações desta superinteligência com os seres humanos? Elas podem alcançar o que chamamos consciência?

E estas máquinas ocuparão e substituirão os homens em diversos campos de trabalho? E se isto ocorrer com que os homens se ocuparão? Haverá o ócio criativo como disse Domenico Massa?

E se por acaso estas máquinas se rebelarem contra os humanos? Quem vai estabelecer os limites delas? E mais, quem terá o poder de fazê-las cumprir estes limites?

Todas estas interrogações nos levam a uma reflexão filosóficas. E o que podemos esperar que nos próximos anos haverá muito progresso e novidades. Estas nos levarão a novas preocupações e reflexões.

BIBLIOGRAFIA

BRANQUINHO, João; MURCHO, Desidério; GOMES, Nelson Gonçalves. *Enciclopédia de termos lógicos-filosóficos*. São Paulo: Martins Fontes, 2006.

BOSTROM, Nick. *Superinteligência: caminhos, perigos, estratégias*. Rio de Janeiro: DarkSide Books, 2018.

LECLERC, André. *Uma introdução à filosofia da mente*. Curitiba: Appris, 2018.

NICOLELIS, Miguel. *O verdadeiro criador de tudo: como o cérebro humano esculpiu o universo como nós o conhecemos*. São Paulo: Pioneira, 2020.

22 | Reflexões sobre alguns problemas filosóficos

PENROSE, Roger. Sombras da mente: uma busca pela ciência perdida da consciência. São Paulo: Unesp, 2021.

PENROSE, Roger. A nova mente do rei: computadores, mentes e as leis da física. Rio de Janeiro: Campus, 1991.

PUTNAM, Hilary. *Corda Tripla: mente, corpo e mundo*. Aparecida: Letras e Letras, 2008.

SEARLE, John. *The construction of social reality*. New York/London: The Free Press, 1995.

SEARLE, John. *Mente, linguagem e sociedade*. Rio de Janeiro: Rocco, 2000.

SIGMAN, Mariano. A vida secreta da mente. Rio de Janeiro: Objetiva, 2017.

TEIXEIRA, João de Fernandes. O cérebro e o robô: uma inteligência, biotecnologia e a nova ética. São Paulo: Paulus, 2015.

TEIXEIRA, João de Fernandes. Filosofia do cérebro: inteligência artificial, biotecnologia e a nova ética. São Paulo: Paulus, 2018.

TURING, Dermot. A história da computação: do ábaco à inteligência artificial. São Paulo: M. Books do Brasil, 2019.

2. O PROBLEMA DOS OBJETOS ARBITRÁRIOS: REFLEXÕES SOBRE A REGRA DA INTRODUÇÃO DO QUANTIFICADOR UNIVERSAL E DA ELIMINAÇÃO DO QUANTIFICADOR EXISTENCIAL



<https://doi.org/10.36592/9786554601061-02>

Martim Mottloch¹

1) Introdução

No seu seminário de 1926, em Varsóvia, Łukasiewicz definiu o objetivo de modelar os procedimentos informais dos matemáticos de uma maneira mais fiel do que acontece nas teorias axiomáticas. Um dos seus alunos Jaśkowski (1934) realizou a tarefa em 1934 simultaneamente com Gentzen (1934), apresentando assim um sistema de dedução natural. A dedução natural é caracterizada pelo procedimento pressuposicional. Ela tem a vantagem diante os sistemas axiomáticos de as provas nela serem mais econômicas. Na filosofia o uso da dedução natural se tornou uma ferramenta indispensável para se obter argumentos válidas. Por essa razão a relação com a linguagem natural e a ontologia requerer o seu devido lugar ao das considerações puramente sintáticas e semânticas no sentido de teoria de modelos. A lógica expressa os aspectos mais gerais da realidade (CHATEAUBRIAND, 2001, p. 29). As regras dos quantificadores dão assim as características das propriedades lógicas de *existência* e *universalidade*. A dedução é chamada de "natural" por

- a) ser próxima dos procedimentos informais;
- b) por as regras darem uma interpretação dos símbolos lógicos.

Ora a regra da introdução do quantificador universal e a regra da eliminação do quantificador existencial que fazem referência a um tipo de objetos *esquisitos* chamados objetos "arbitrários", parecem em parte inaturais. As variáveis que os designam são chamadas de variáveis próprias ou parâmetros próprios, ou derivado do alemão *eigenvarável*. Claramente essa noção abrange duas classes diferentes. Por um lado, no caso da introdução da quantificação universal o parâmetro próprio significa algo como "não importa qual" e por outro lado no caso da quantificação existencial algo como "não sei qual."

¹ Professor da Universidade Federal do Piauí.

24 | Reflexões sobre alguns problemas filosóficos

(PELLETIER, p. 199, p. 14). Estes objetos arbitrários incentivam a reflexão filosófica. Pretendo refletir a natureza desses objetos tomando como pontos de partida os sistemas de Gentzen-Prawitz e Fitch (1952) da dedução natural do cálculo dos predicados de primeira ordem clássico.

2) Regras com restrições

Para justificar as regras de eliminação e introdução dos operadores será requerido mostrar que as restrições são necessárias, isto se demonstra através de contraexemplos.

2.1) Introdução do quantificador universal

Começaremos pela introdução do quantificador universal (segundo Gentzen-Prawitz)

$$\frac{Fa}{\forall x Fx}$$

A *eigenvariável* (o parâmetro próprio) a 1. não deve ocorrer em $\forall x Fx$ 2. nem em qualquer hipótese de qual Fa depende. (GENTZEN, 1934, p.186). Alguns exemplos simples revelam que a violação dessas restrições leva a deduções inválidas, por conseguinte que elas são necessárias.

(1) Quando a ocorre em $\forall x Fx$, derivação de $\forall x Fxx \rightarrow \forall x Fxa$:

$$\frac{\frac{\forall x Fxx}{Faa}}{\forall x Fxa}$$

Neste caso a ocorre na fórmula obtida através da regra da introdução na conclusão do argumento. Introduzindo o condicional obtemos $\forall x Fxx \rightarrow \forall x Fxa$. Intuitivamente a condicional diz que se todos os x estão numa relação reflexiva consigo mesmo, então todos os x estão nesta relação com a . Uma interpretação de F poderia ser a propriedade de *amar*, então se todos os objetos amam a si mesmo, eles amam o objeto a ; um argumento claramente inválido.

(2) Quando a já ocorre numa hipótese de qual Fa depende, derivação de $(Fa \rightarrow \forall x Fx)$

$$\frac{Fa \quad Ga}{\quad}$$

$$\frac{Fa}{\quad}$$

$$\forall x Fx$$

Neste caso Fa foi obtido da conjunção, então o a não é arbitrário.

Introdução do quantificador universal (\forall) no sistema de Fitch

$$\frac{\quad}{\boxed{\quad}}$$

$$\quad | Fa$$

$$\forall x Fx$$

A variável a não deve ocorrer fora da subprova onde foi introduzido. Abrir uma caixa (*box*), a subprova corresponde a supor algo (BARWISE; ETCHEMENDY, 2003, p. 343). Para isso se utiliza o meio gráfico de enquadramento (*boxing*)² que lembra o *flagging* de Quine (1950). A variável a somente pode ser utilizada dentro da subprova (a subprova, contudo, também pode conter outras subprovas dela): por conseguinte a nunca mais pode ser reutilizada nessa dedução. A condição de a não poder ocorrer fora da subprova correspondem às restrições segundo Gentzen-Prawitz, pois as hipóteses de quais Fa poderia depender e $\forall x Fx$ ficam fora da subprova. No sistema de Prawitz-Gentzen, todavia, o parâmetro a pode ser reutilizado na mesma dedução.

2.2) Quantificação existencial

Eliminação do quantificador existencial (\exists segundo Gentzen-Prawitz)

$$\frac{\quad [Fa] \quad \frac{\exists x Fx \quad B}{\quad}}{\quad} B$$

A variável própria a não deve ocorrer 1. nem em $\exists x Fx$, 2. nem em B 3. nem em qualquer hipótese da qual essa fórmula B depende com exceção de Fa . (GENTZEN, 1934, p. 186). Prawitz (1965, p. 20) modifica a 3. condição, dizendo que a não deve ocorrer em nenhuma hipótese de qual só a ocorrência superior de B depende.

As regras tanto da introdução do quantificador universal tanto da eliminação do quantificador existencial facilitam aplicar as regras para juntores lógicos dentro de fórmulas quantificadas. No caso de a trata-se meramente de uma constante intermediária

² No texto será o símbolo ' \wedge '.

26 | Reflexões sobre alguns problemas filosóficos

que não designa nenhum indivíduo do domínio em particular. E que tem que desaparecer na dedução.

Novamente apresentarei alguns contraexemplos que levam a deduções inválidas.

(3) Na derivação, a ocorre em $\exists x Fx$, derivação de $(\forall x \exists y Fxy) \rightarrow \exists x Fxx$

$$\frac{\frac{\frac{\quad \forall x \exists y Fxy \quad \quad \quad}{\quad \exists y Fay}}{\quad \quad \quad} \quad \frac{\quad \quad \quad}{\quad \quad \quad} Faa}{\quad \quad \quad} \exists x Fxx$$

(4) Caso onde a ocorre em B , derivação de $(\forall x \exists y Fxy) \rightarrow (\exists y \forall x Fxy)$

$$\frac{\frac{\frac{\quad \forall x \exists y Fxy \quad \quad \quad}{\quad \exists y Fby}}{\quad \quad \quad} \quad \frac{\quad \quad \quad}{\quad \quad \quad} Fba}{\quad \quad \quad} Fba$$

$$\frac{\quad \quad \quad}{\quad \quad \quad} \forall x Fxa$$

$$\frac{\quad \quad \quad}{\quad \quad \quad} \exists y \forall x Fxy$$

3) Essa vez a ocorre na hipótese de qual B depende, derivação da conclusão $\exists x (Gx \wedge Fx)$ das premissas Fa e Ga

$$\frac{\frac{\quad \quad \quad}{\quad \quad \quad} Fa \quad \frac{\frac{\quad \quad \quad}{\quad \quad \quad} Ga \quad \frac{\quad \quad \quad}{\quad \quad \quad} Fa}{\quad \quad \quad} \exists x (Gx \wedge Fx)}{\quad \quad \quad} \exists x (Gx \wedge Fx)$$

Eliminação do quantificador existencial (\exists segundo Fitch)

| $\exists x Fx$

|| $\hat{a} Fa$

|| B

| B

A regra requer a abertura de uma subprova onde a variável própria a novamente não deve ocorrer fora da subprova, para caracterizar isso usa-se o *boxing* (no texto “^”). A variável a denota o objeto indefinido que satisfaz a fórmula Fx . Na subprova, deve se deduzir uma fórmula B que não contém a (BARWISE; ETCHEMENDY, 2003, p. 348). No caso da eliminação do existencial se dá um nome temporário para a testemunha indefinida, não identificada; a única informação sobre ela é que o predicado se aplica a ela, por isso o nome tem que ser novo, não carregado de informações. Ele é um tipo de descrição indefinida. Problema do sistema ao modelo de Fitch é que se precisa manobrar muitas subprovas.

2.3) Parâmetros puros

Com a noção de parâmetros próprios puros Prawitz restringe as ocorrências de parâmetros próprios. Para isso ele introduz a noção *conexão*.

2.3.1) Conexão lateral (*side-connection*)

Seja B uma ocorrência de uma fórmula na dedução π e seja $(\pi_1, \pi_2, \dots, \pi_n/B)$ a subárvore de π (determinada por B) que tem B como fórmula final e sejam A_1, A_2, \dots, A_n fórmulas finais das árvores π_1, \dots, π_n respectivamente. Se A_1, \dots, A_n ocorrem imediatamente em cima de B segundo a ordem deles de esquerda para direita, A_i é *conectado lateralmente* com A_j ($i, j < n$). (PRAWITZ, 1965, p. 25-26). Assim na formulação da regra \exists de cima) (seriam: $A_1 = \exists x Fx, A_2 = B$, a ocorrência inferior de B a ocorrência superior de B seria conectada lateralmente com $\exists x Fx$.

Usando a noção de conexão lateral, Prawitz define a noção de conexão. As fórmulas A e B numa dedução π são conectadas se π seja uma sequência, A_1, \dots, A_n , tal que $A_1 = A$ e $A_n = B$ e para todos $i < n$ uma das seguintes condições obtêm:

1. A_i não é premissa maior de uma aplicação de \exists -E e \neg -E, e A_{i+1} fica imediatamente embaixo de A_i ou vice versa.
2. A_i é premissa de \neg -E e A_{i+1} é conectado lateralmente com A_i
3. A_i é premissa maior de uma aplicação de \exists -E ou \neg -E e A_{i+1} é uma hipótese descartada por essa aplicação ou vice versa. No caso da aplicação de \exists -E A_i seria $\exists x Fx$ e A_{i+1} seria Fa .
4. A_i é uma consequência da aplicação de \neg -I ou da regra do *falsum* clássico e A_{i+1} é uma hipótese descartada por essa aplicação ou vice versa. (PRAWITZ, 1965, p. 28).

Com a noção de *conexão* Prawitz definiu a ligação por um parâmetro. Uma fórmula A é *ligada* numa dedução π a uma fórmula B pelo parâmetro a se há uma conexão tal que a ocorre em todas as ocorrências das fórmulas nessa conexão. Um parâmetro de uma aplicação de \exists -E numa dedução π é puro in π se cada ocorrência de fórmula in π na qual a ocorre é ligada através de a a premissa $A(a)$ e a é o único parâmetro próprio da aplicação. (PRAWITZ, 1965, p. 28-29). O lema dos parâmetros estabelece o mesmo efeito que a restrição do sistema de Fitch do que o parâmetro próprio. Não pode ocorrer fora da

28 | Reflexões sobre alguns problemas filosóficos

subprova nas duas regras consideradas. A desvantagem, porém, é que se tem que ter bastantes parâmetros, pois nenhum parâmetro individual pode ser utilizado de novo como parâmetro.

O operador de Hilbert capta o sentido da *eigenvariável*. $\exists xF$ designa algum x que satisfaz F se tiver algum. Ele é um operador de escolha. Assim xFx é equivalente a $F(\exists xF)$. A *eigenvariável* tem exatamente a mesma função. A introdução de Fa corresponde a uma introdução de $F(\exists xF)$ e como xFx e $F(\exists xF)$ são equivalentes, ela é e justificada. A *eigenvariável* significa aqui na linguagem natural algo como "objeto indefinido". A designação exata do objeto é desconhecida, é um objeto não-identificado.

3) A Interpretação da noção de objeto arbitrário

Van Dalen (1994, p. 90, tradução minha) percebe a fraqueza de uma explicação intuitiva: " tem a seguinte explicação: se um objeto arbitrário x tem a propriedade Φ , então cada objeto tem a propriedade Φ . O problema é que nenhum dos objetos que nós conhecemos na matemática pode ser considerado 'arbitrário'. Assim em vez de procurar o 'objeto arbitrário' no mundo real (quanto à matemática), procuremos um critério sintático." Esse critério é para van Dalen o de Gentzen.

3.1) Interpretação de Locke segundo Celucci

Será possível deduzir de uma predicação sobre um indivíduo uma conclusão geral. Isso se não parecer mágico, parece absurdo. Um exemplo clássico deste tipo de raciocínio se encontra na prova da proposição 1.32 de Euclides (1944, p.21) segundo Celucci (2009). Euclides prova que a soma dos ângulos interiores num triângulo é igual a soma de dois ângulos retos. Euclides demonstra o teorema num triângulo individual. Mas como justificar a generalização? Uma proposta para justificar essa prova de Euclides vem do Locke (1975). A sua explicação é que o triângulo considerado de fato é um triângulo geral. Isso quer dizer a ideia geral de um triângulo. As propriedades dele são comuns a todos os triângulos. A maneira de obter esta ideia geral consiste em abstrair dos triângulos particulares todas as propriedades que todos eles possuem. Um argumento, conhecido contra a existência dos

objetos gerais é o seguinte:

Primeiro a seguinte equivalência tem que obter:

$F(a)$ é verdadeiro se e somente se $\exists x F(x)$ é verdadeiro.

Ora, segundo o *tertium non datur* $F(a)$ ou $\neg F(a)$, $F(x)$ simbolizando, por exemplo, "é vermelho". (CELUCCI, 2009, p. 3-4).

Se $F(a)$ então $\exists x F(x)$, por outro lado, se $\exists x F(x)$ então $F(a)$.

A solução desse impasse poderia consistir no argumento de que o indivíduo ideal não é maximal consistente com respeito às suas propriedades. Poderia se considerar o triângulo da prova de Euclides como um universal não-predicativo, universal individual, como foram concebidas as ideias platônicas. Todas as instâncias exemplificam o objeto abstrato a . O grande problema tanto na dedução natural como na vida cotidiana que se descarta da diferença entre eles e os objetos concretos. Numa versão mais moderna de Zalta (1983) o indivíduo abstrato, nesse caso o triângulo, codifica (*encodes*) as propriedades de *ter a soma de ângulos de 180°, possuir três ângulos etc.* As propriedades que ele codifica são exatamente essas por quais ele é constituído. Elas são, são as propriedades essenciais dele. O medo nominalista de uma superpopulação do universo ressurgue aqui. Será que quando utilizamos a introdução universal para descrever o mundo, estamos de fato fazendo referência aos indivíduos abstratos com propriedades essenciais?

3.2) Solução de Gentzen e Crítica

Voltaremos para a posição de Gentzen. Informalmente as restrições são do que se Fa for provado de um "a arbitrário", $\exists x Fx$ é válido. O requisito de a ser completamente arbitrário expresso de uma maneira mais exata é do que Fa não deve depender de nenhuma hipótese que contém a . Obviamente, todas as ocorrências de a devem ser substituídas por x (GENTZEN, p. 187). Assim na derivação não é feito nenhum uso de uma propriedade particular de a (além de F). Isso corresponderia ao abstrair das outras propriedades de a . Tennant (1983) chama a concepção de Gentzen de "*ontologically bloating*". Como variáveis são símbolos específicos numa linguagem, quando Gentzen diz que a designa um valor arbitrário da variável a , a designa um objeto arbitrário. Mas Celucci (2009, p.8) argumenta com Rescher (1958) que não faz sentido de falar de um indivíduo arbitrário, nisso se cometeria o que Whitehead chama de *fallacy of misplaced correctness*. A arbitrariedade não

reside nos indivíduos, mas assim na ambiguidade intencionada da noção. O que seria este indivíduo no domínio. Posso percorrer o domínio do discurso inteiro sem encontrá-lo. Ele não estará lá. Este problema é obviamente interligado com o problema da indução. Em vez de justificar uma proposição geral por várias proposições singulares, (o que não se deixa alcançar), os lógicos parecem escolher um representante típico da classe que satisfaz o predicado. O problema é que nenhum indivíduo concreto do domínio é tão completamente típico. E a predicação de um indivíduo particular nunca poderia justificar a generalização. O que poderia, enfim, significar “designar uma variável arbitrária”?

3.3) *Generalidade como repetibilidade :solução de Celucci*

Por essas razões Celucci (2009, p. 12-13) propõe uma solução diferente. Segundo ele a introdução do quantificador universal é possível porque a derivação pode ser repetida para outros objetos. Quer dizer substituindo a por outra *eigenvariável* b obteremos a mesma derivação. O que se prova com a vai ser válido para qualquer outro b . Esta versão e a do Gentzen são extensionalmente equivalente, porém não intensionalmente. Se substituirmos a por outra *eigenvariável* b na derivação, chegaremos à derivação de Fb e daí de $\forall xFx$. As duas interpretações da regra não são intensionalmente equivalente por elas referem a diferentes características das derivações, para Gentzen (1934) ao fato que a não ocorre em nenhuma hipótese de qual Aa depende, para Celucci (2009) que a substituição de a por um b diferente que não ocorre nessa derivação obtemos a derivação de Fb das mesmas hipóteses. Celucci avista a superioridade da sua proposta diante da de Gentzen por a de Gentzen depender da noção de abstração e a derivação acontece com respeito a um objeto individual sem poder dizer qual é. Celucci (2009, p. 9) alega que o intuicionista Herbrand (1968, p. 225, tradução minha) explica o significado da regra da introdução do quantificador universal do modo semelhante: “Quando se diz que um argumento (ou teorema) é verdadeiro para todos esses x , isso significa, para todo x tomado em particular, é possível repetir o argumento geral em questão.” Celucci (2009, p 13-14) vê a sua solução confirmada por formulações de Fitch (1952) que a subprova para introdução do quantificador universal seria do mesmo jeito válida se se substituísse a por qualquer outro b , que ela é, por conseguinte geral e na matemática grega onde tem algumas interpretações de

generalidade como repetibilidade.

4) Reflexões adicionais de Shapiro

Shapiro (2004, p. 566-567) tenta explicar a natureza dos objetos arbitrários utilizando a ideias de Kit Fine (1985). Segundo Fine um objeto arbitrário tem todas as propriedades comuns aos objetos do seu domínio. Outra expressão que é utilizada é do "membro típico". Ora, Fine distingue duas classes de objetos arbitrários que correspondem *grosso modo* as duas classes de variáveis: as universalmente quantificadas e as existencialmente quantificadas. A primeira classe de objetos *arbitrários independentes* corresponderia assim as variáveis universalmente quantificadas, os objetos arbitrários dependentes às variáveis da quantificação existencial. Os objetos arbitrários independentes são somente caracterizados pelo seu domínio de valores. Seja a um número real arbitrário, o seu domínio de valores seriam todos os números reais individuais. Os objetos arbitrários dependentes, por sua parte, dependem além do seu domínio de valores também de outros objetos arbitrários.

5) Conclusão

A regra da eliminação do quantificador existencial gera do ponto de vista filosófico menos problemas. O sentido do parâmetro próprio e bem captado pelo operador ε de Hilbert. A generalização, todavia, apresenta tanto na filosofia como na lógica um problema perpetuo. Seja a solução considerar o objeto arbitrário como um objeto abstrato individual (envolvendo-se numa espécie de platonismo) ou em justificar a generalidade pela possível repetibilidade. Do ponto de vista sintático é mais adequado restringir o uso de um parâmetro próprio a uma única aplicação da regra. Nunca se pode ter cuidado demais com generalizações!

REFERÊNCIAS BIBLIOGRÁFICAS

BARWISE, J; ETCEMENDY, J. (2003). *Language, Proof, and Logic*. Stanford: CSLI Publications.

32 | Reflexões sobre alguns problemas filosóficos

CHATEAUBRIAND, O. (2001) *Logical Forms. Part I: Truth and Description*. Campinas: UNICAMP, Centro de Lógica, Epistemologia e História da Ciência.

CELUCCI, C. (2009). The Universal Generalization Problem. *Logique et Analyse, Nouvelle Série*, vol. 52, no. 205 (março), p. 3-20.

EUCLIDES. (1944). *Elementos de Geometria* São Paulo: Edições Cultura.

FINE, K. (1985). *Reasoning with Arbitrary Objects*. New York: Blackwell.

FITCH, F. (1952). *Symbolic Logic: An Introduction*. New York: Ronald Press.

GENTZEN, G. (1934) Untersuchungen über das Logische Schließen. *Mathematische Zeitschrift*, vol. 39, p. 76-210.

HERBRAND, J. (1968). *Écrits logiques*. 3. ed. Van HEIJENOORT, J. (Ed.). Paris: Presses Universitaires de France.

JASKOWSKI, S. (1934). On the Rules of Suppositions in Formal Logic. *Studia Logica*, vol. 1, p. 5-32.

LOCKE, J. (1975). *An Essay Concerning Human Understanding*. NIDDITCH P H. (Ed.). Oxford: Oxford University Press. 1975.

PELLETIER, F.J. (1999) A Brief History of Natural Deduction. *Journal of History and Philosophy of Logic*, vol. 20, p.1-31.

PRAWITZ, D. (1965) *Natural Deduction: A Proof-Theoretic Study*. Stockholm: Almqvist & Wiksell.

QUINE, W.V. (1950). On Natural Deduction. *Journal of Symbolic Logic*, vol 12, no. 2, junho p. 93-102.

RESCHER, N. (1958). Can there be any random individuals? *Analysis*, vol. 18, p. 114-117.

SHAPIRO, S. T. (2004). A Logic of Arbitrary and Indefinite Objects. In: *KR*. 2004. p. 565-575.

TENNANT, N. (1983). A defence of arbitrary objects II. *Proceedings of the the Aristotelian Society*, supplementary vol. 57, p. 78-89.

VAN DALEN, D. (1994). *Logic and Structure*. 3ª ed. Berlin: Springer.

ZALTA, E. (1983). *Abstract objects: An Introduction to Axiomatic Metaphysics*, Dordrecht: D. Reidel

3. UM NAVIO RECONSTRUÍDO EM ALTO MAR: CONSIDERAÇÕES SOBRE O SUPOSTO PRAGMATISMO DE QUINE EM 'DOIS DOGMAS DO EMPIRISMO'



<https://doi.org/10.36592/9786554601061-03>

Maria Cristina de Távora Sparano¹

Leonardo Bruno Vieira Santos²

Introdução

Em *Dois dogmas do empirismo* Quine afirma que a superação dos dois dogmas do empirismo promove uma mudança de direção na Filosofia em direção ao pragmatismo. A superação dos dogmas empiristas também implica para Quine que se deva adotar uma perspectiva holista (holismo semântico), dada a impossibilidade de se distinguir entre enunciados que seriam puramente linguísticos e enunciados que possuem elementos linguísticos, mas também possuem elementos extralinguísticos, ou seja, dada a impossibilidade de se distinguir entre enunciados analíticos e sintéticos.

As questões que surgem, então, são as seguintes: qual a relação entre pragmatismo e holismo? Qual a concepção de pragmatismo Quine está fazendo referência? No presente trabalho optou-se por discutir tais questões comparando as ideias que Quine defendeu em *Dois dogmas do empirismo* com as ideias de Charles Peirce. A escolha das ideias de Peirce dentre os autores clássicos do pragmatismo americano se justifica pelo fato de que o autor, dentre os pioneiros do pragmatismo, ter ideias que se alinham mais com a tradição analítica, principalmente sua afinidade com questões da lógica.

¹ Doutora em Filosofia pela PUC-RS e é professora aposentada da UFPR

² Mestre em Filosofia e professor de Filosofia.

1. Pragmatismo

Uma definição exata de pragmatismo talvez seja uma tarefa complicada, pois, assim como a Fenomenologia, o Positivismo Lógico etc., não há consenso entre os defensores da corrente. Os diferentes modos como todos os filósofos da corrente conceberam o pragmatismo levou ao descontentamento de seu criador. Peirce expressa essa divergência quando afirma que:

I shouldn't wonder if they were ashamed of me. What could be more humiliating than to confess that one has learned anything of a logician? But for my part I am delighted to find myself sharing the opinions of so brilliant a company. The new pragmatists seem to be distinguished for their terse, vivid, and concrete style of expression together with a certain buoyancy of tone as if they were conscious of carrying about them the master key to all the secrets of metaphysics.

Every metaphysician is supposed to have some radical fault to find with every other, and I cannot find any direr fault with the new pragmatists than that they are lively. In order to be deep it is requisite to be dull.

On their side, one of the faults that I think they might find with me is that I make pragmatism to be a mere maxim of logic instead of a sublime principle of speculative philosophy. (PEIRCE, 1998, p. 134).

A divergência expressa por Peirce em relação ao que ele chama de “novos pragmatistas” já existia em relação a seu amigo e colega de Clube Metafísico, William James. Cornelis De Waal (2007) apresenta em seu livro sobre o pragmatismo, uma passagem que ilustra bem o ponto de divergência entre Peirce e James e que reforça as considerações daquele na citação que antecedeu o presente parágrafo.

James, então professor de Harvard, havia convidado Peirce para uma série de conferências em Cambridge. A proposta para as conferências que Peirce enviou para James o desagradou bastante, uma vez que se tratava de um conteúdo voltado inteiramente para a lógica. A resposta de James foi imediata:

Sinto que você se apegue tanto à lógica formal. [...] Você mal pode imaginar o pouco interesse que existe nos aspectos puramente formais da lógica. Coisas sobre esse assunto deveriam ser impressas para poucas pessoas. Você está cheio de novas idéias – e as conferências não precisam de modo algum formar um todo contínuo. Tópicos separados de um caráter vitalmente importante cairiam perfeitamente bem. (JAMES *apud* DE WAAL, 2007, p. 132).

Peirce não recebe bem as observações de James e sua resposta deixava claro seu descontentamento com a tentativa do psicólogo de lhe desencorajar e falar sobre temas ligados à lógica formal. “A resposta de Peirce não é desprovida de sarcasmo. Como uma referência direta à carta de James, ele propôs dar às conferências o nome de “Detached Ideas on Vitally Important Topics” [Idéias destacadas sobre tópicos vitalmente importantes]”. (DE WAAL, 2007, p. 132).

Essas divergências não só com James, mas com os demais autores que se declaravam pragmatistas levou Peirce a abandonar o termo pragmatismo e adotar o termo pragmaticismo, [...] “which is ugly enough to be safe from kidnappers”. (PEIRCE, 1998, p. 335). Para Peirce esse procedimento é necessário, pois:

Much as the writer has gained from the perusal of what other pragmatists have written, he still thinks there is a decisive advantage in his original conception of the doctrine. From this original form every truth that follows from any of the other forms can be deduced, while some errors can be avoided into which other pragmatists have fallen. (PEIRCE, 1998, p. 335).

A posição de Peirce em relação aos demais pragmatistas é um exemplo da dificuldade de se tentar apresentar o pragmatismo como uma corrente filosófica com abordagens aceitas por todos os seus membros. Essas divergências não existiam somente entre os pioneiros da corrente, os supostos neopragmatistas também herdaram esse desacordo, dependendo das filiações que estabeleciam com as ideias dos filósofos fundadores do pragmatismo. Um exemplo que pode ser citado é a oposição intelectual que pode ser estabelecida entre Susan Haack, que segue o pragmatismo peirceano e Richard Rorty, que está mais próximo do pragmatismo schilleriano (DE WAAL, 2007).

Essas considerações sobre as divergências entre os pragmatistas não impedem a apresentação de alguns pontos gerais sobre o pragmatismo e que possam servir de base para o objetivo proposto nesse trabalho. O pragmatismo é caracterizado de dois modos distintos: (1) "Há estudiosos que definem o pragmatismo como uma teoria do conhecimento"; (SHOOK, 2002, p. 11); (2) "Outros defendem que o pragmatismo é um método com o qual se faz filosofia, e não uma teoria filosófica". (DE WALL, 2007, p. 22).

Essa dificuldade inicial em determinar se o pragmatismo é um método ou uma teoria filosófica não é privilégio somente dos estudiosos do pragmatismo, como visto anteriormente, Peirce já havia constatado esse problema, no que se refere às inúmeras ramificações que se proliferaram no seio do pragmatismo, a ponto de declarar que o pragmatismo era "uma casa em guerra contra si mesma" (DE WALL, 2007, p. 15).

Do ponto de vista da interpretação (1) o pragmatismo é uma teoria do conhecimento que visa dar uma resposta adequada para a questão "como se dá o conhecimento?". A hipótese pragmatista é que o homem dispõe de apenas uma metodologia complexa de conhecimento, que apresenta vários níveis de habilidade. (SHOOK, 2002, p. 11-12). Há, entretanto, uma forma básica de investigação inteligente que garante a continuidade desses níveis.

Por mais que divergissem em vários pontos, Peirce, James e Dewey concordavam que:

1. embora a experiência seja o fundamento do conhecimento, a mente transforma a experiência em objeto de conhecimento;
2. a transformação da experiência visa apaziguar a dúvida, etapa preparativa de uma ação com vista a um fim específico;
3. como a mente visa a uma crença prática, a transformação que ela faz da experiência é guiada pela atividade experimental;
4. o processo experimental de criar crenças sólidas pode ser logicamente avaliado com base em sua função de nos possibilitar prever confiavelmente e controlar nosso ambiente. (SHOOK, 2002, p. 12).

Esses quatro pontos de concordância entre os pioneiros do pragmatismo revela as filiações epistemológicas que esses filósofos aderiram ao mesmo tempo em que revela que essas filiações não são passivas. Os pragmatistas clássicos são empiristas, mas não professam um empirismo aos moldes do empirismo britânico, principalmente se levarmos em conta a filosofia de David Hume.

Essa tendência empirista do pragmatismo nasce de uma rejeição contundente das teses racionalistas de cunho cartesiano. Peirce, por exemplo, dedicou dois textos importantes para lidar com essa questão, são eles *Questions concerning certain faculties claimed for man* e *Some consequences of four incapacities claimed for man* ambos publicados em 1878.

No segundo texto encontram-se uma das primeiras formulações do empirismo de Peirce.

1. We have no power of Introspection, but all knowledge of the internal world is derived by hypothetical reasoning from our knowledge of external facts.
2. We have no power of Intuition, but every cognition is determined logically by previous cognitions.
3. We have no power of thinking without signs.
4. We have no conception of the absolutely incognizable. (PEIRCE, 1992, p. 30).

O primeiro princípio apresentado por Peirce é o que permite caracterizar sua teoria de empirista, pois é uma defesa clara da ideia de que nosso conhecimento é derivado da nossa experiência imediata e não na nossa mente, como defenderam alguns racionalistas.

O segundo princípio do empirismo peirceano é interessante para o presente trabalho, principalmente quando se está discutindo as conclusões de *Dois Dogmas do Empirismo* de Quine. Peirce está defendendo a ideia de que:

[...] todo conhecimento adquirido precisa da contribuição de outros tipos de conhecimento já adquiridos. Essa tese não resulta em uma regressão ao infinito, porque a filosofia de Peirce afirma a existência de graus e níveis de cognição. O conhecimento surge gradualmente a partir de hábitos mais simples de ação em nossa infância. Quando crianças, nossa capacidade de adquirir conhecimento

faz parte do mesmo processo de socialização que nos ensina um idioma. (SHOOK, 2002, p. 49-50).

Essa tese de Peirce coloca em xeque as pretensões fundacionistas de fundamentar o conhecimento em bases epistemológicas que são independentes da experiência sensível. Essas teses também colocam em xeque o empirismo dualista dos filósofos modernos e de seus adeptos contemporâneos. "Juntos, esses quatro princípios implicam que todo pensamento é um processo de interações entre signos e deve estar relacionado à ação humana no mundo". (SHOOK, 2002, p. 50).

Peirce, consciente da importância da experiência humana, promoveu uma redução radical na lista de categorias legadas pela tradição filosófica, Aristóteles e Kant, por exemplo, tinham apresentado uma lista de categorias extensa. Peirce desenvolveu uma fenomenologia com apenas três categorias, a saber, primeiridade, secundidade e terceiridade. Poder-se-ia até mesmo defender que sua fenomenologia é o ponto de partida para a compreensão de suas ideias, uma vez que essas categorias tem papel decisivo no desenvolvimento das teorias de Peirce.

Em se tratando de pragmatismo, o elemento essencial para se compreender as ideias de Peirce é sua máxima que foi apresentada pela primeira vez em um texto de 1878, intitulado *How to make our ideas clear*: "Consider what effects which might conceivably have practical bearing, we conceive the object of our conception to have. Then, our conception of these effects is the whole of our conception of the object". (PEIRCE, 1992, p. 132).

Essa máxima é o terceiro grau de clareza proposto por Peirce em seu texto, o primeiro grau se dá quando uma ideia é apreendida de um modo que possa ser possível distingui-la claramente de outras ideias quando for necessário, o segundo grau de clareza se estabelece quando por meio de critérios abstratos que determinam sem ambiguidade o que se relaciona ou não ao conceito. (DE WAAL, 2007). A máxima pragmatista de Peirce revela a natureza de seu pragmatismo, a saber, "um método para determinar o significado de conceitos, ideias, crenças, alegações, proposições etc., de qualquer coisa que pode servir como signo" (DE WAAL, 2007, p. 41).

Assim o pragmatismo promoveu duas contribuições significativas ao empirismo: (1) O ponto de vista naturalista: surge da tentativa de naturalizar nossa

compreensão de como o sujeito conhece e o que é o conhecimento; (2) O ponto de vista historicista: os pragmatistas defenderam que, diferentemente do que acreditavam os empiristas e racionalistas clássicos, o modo em que a mente conhece é dinâmico, está em evolução.

Outro ponto interessante, que deve ser ressaltado é que a filosofia de Peirce e Dewey era antidualista, ou seja, eles se voltavam contra todos os dualismos da filosofia. Assim, o empirismo defendido pelos pragmatistas não era dualista, pois eles não acreditavam que a experiência humana ocorresse fora do mundo natural. (SHOOK, 2002, p. 33).

O suposto pragmatismo de Quine

Na introdução de *Dois dogmas do empirismo* Quine nos diz que uma das consequências de se abandonar os dois dogmas cultivados pelo empirismo é “uma mudança de direção rumo ao pragmatismo” (QUINE, 2011, p. 37). No entanto, essa “mudança de direção” é controversa, o próprio Quine chamava a atenção para o fato de que classificá-lo como pragmatista era interpretar de modo equivocado os últimos parágrafos de *Dois dogmas do empirismo*. (DE WALL, 2007, p. 203).

Uma leitura atenta do texto de Quine é o suficiente para extrair algumas considerações importantes para a questão que nos propomos tratar neste trabalho. Mas um passo de cada vez. Antes de tudo é preciso apresentar, mesmo que de modo resumido os argumentos de Quine contra os tais dois dogmas do empirismo (do Empirismo Lógico, vale ressaltar).

Os dois dogmas são a crença na distinção analítico/sintético, que serviria como o alicerce para uma proposta fundacionista, e o reducionismo. No caso do primeiro dogma o problema está na noção de analiticidade. Se Kant definia um juízo analítico como aquele em que o sujeito já contém o predicado, de tal modo que esse nada lhe acrescenta, os Positivistas (ou Empiristas Lógicos), no âmbito da virada linguística, definiram um enunciado analítico como aquele que é verdadeiro em virtude do significado dos termos que compõem esse enunciado.

O ataque de Quine à noção de analiticidade tem início exatamente contra seu fundamento: a noção de significado.

Para a teoria do significado, uma questão flagrante é a da natureza de seus objetos: que tipo de coisas são os significados? Uma necessidade perceptível de que haja entidades significadas pode derivar de uma falha anterior em perceber que significado e referência são distintos. Uma vez que a teoria do significado esteja nitidamente separada da teoria da referência, é necessário apenas um pequeno passo para reconhecer como primeira ocupação da teoria do significado tão somente a sinonímia de formas linguísticas e a analiticidade dos enunciados; os próprios significados, como entidades intermediárias obscuras, podem muito bem ser abandonados. (QUINE, 2011, p. 40).

Em seguida o seu alvo é a analiticidade. O argumento de Quine pode ser resumido nos seguintes termos: uma vez provada a impossibilidade de gerar a analiticidade nos contextos em que ela deveria ocorrer, pois em alguns casos a própria noção é pré-requisito para o processo, então, a distinção analítico/sintético é ilusória e insustentável (QUINE, 2011). Pode-se dizer que o dogma do reducionismo desmorona junto com o dogma da analiticidade, uma vez que necessitava dele para existir.

Os argumentos de Quine contra os dois dogmas são os seguintes: uma vez que existem duas classes de enunciados analíticos, sendo que a primeira classe é das verdades lógicas, o problema da analiticidade residiria na segunda classe que depende da noção de sinonímia para se tornar uma verdade lógica.

Quine apresenta três tipos de definição:

1. Definição que consiste em fazer uso da formação dos lexicógrafos, presentes nos dicionários;
2. Explicação, onde o objetivo é aperfeiçoar o *definiendum* (o objeto) aprimorando, assim, seu significado;
3. A introdução explicitamente convencional de novas notações para fins de pura abreviação – neste tipo de definição o *definiendum* se torna sinônimo do *definiens* (as características do objeto) porque foi criado para tal propósito.

Segundo Quine, a palavra “definição”, devido a seu uso frequente em escritos lógicos e matemáticos acabou assumindo um “tom perigosamente tranquilizante”. De qualquer modo, tanto no trabalho formal quanto no informal a definição está na

dependência de relações de sinonímia anteriores. Assim sendo, a definição não pode nos ajudar na questão da sinonímia e da analiticidade, segundo Quine.

Tendo examinado a relação da sinonímia e a definição e descoberto que esta não possui a chave daquela, Quine passa a examinar se a permutabilidade pode nos dar acesso ao tipo de sinonímia necessária para transformar enunciados analíticos da segunda classe em verdades lógicas.

Mas a permutabilidade que Quine leva em consideração é a permutabilidade *salva veritate*, ou seja, aquela que conserva o valor de verdade. Assim, de acordo com Quine, permanece a questão de saber se a permutabilidade *salva veritate* (excetuando as ocorrências no interior das palavras) é condição suficientemente rigorosa de sinonímia, ou se, ao contrário, algumas expressões heterônimas podem ser, deste modo, permutáveis.

Porém, esclarecemos que não se trata aqui de sinonímia no sentido da completa identidade nas associações psicológicas ou da qualidade poética; não existem verdadeiramente duas expressões sinônimas neste sentido. Apenas interessa o que pode ser chamado sinonímia cognitiva.

Sinonímia cognitiva é aquela que vale para o conhecimento. Assim, o que Quine pretende é tentar derivar a analiticidade da sinonímia cognitiva. Entretanto, tal forma de permutabilidade não tem qualquer significado até que seja relativizada a uma linguagem cuja amplitude esteja especificada em aspectos relevantes.

E mesmo aplicada a uma linguagem do tipo extencional a permutabilidade *salva veritate* não é possibilidade de uma sinonímia do tipo que Quine deseja, já que não se pode, também, derivar a analiticidade. Assim, Quine conclui que o esforço para explicar primeiramente a sinonímia cognitiva, para poder dela posteriormente derivar a analiticidade é uma forma inadequada de abordar o problema. O que deve ser feito é tentar de alguma forma explicar a analiticidade sem apelar à sinonímia cognitiva.

Deixando de lado a questão da sinonímia, Quine passa a examinar a possibilidade da analiticidade com base em regras semânticas. Quine nos diz que o problema em distinguir um enunciado analítico de um enunciado sintético na linguagem comum se deve a vagueza desta linguagem, sendo que tal distinção seria mais clara se possuíssemos uma linguagem artificial precisa com regras semânticas explícitas.

42 | Reflexões sobre alguns problemas filosóficos

Entretanto, do ponto de vista da analiticidade uma linguagem artificial com regras semânticas, da forma que a permutabilidade *salva veritate*, não nos ajuda a obter uma compreensão da analiticidade, visto que, para que tal linguagem tenha algum interesse, temos que já possuir uma compreensão de analiticidade.

Portanto, se nem a noção de analiticidade nem a de sinonímia podem ser definidas claramente e, além disso, fazem referência ao uso ordinário que não é explicado pela noção de significado que parecem requerer, assim, pode-se concluir que não havendo um enunciado que é verdadeiro por si mesmo, então a analiticidade não pode ser estabelecida.

Como foi anunciado no primeiro capítulo do presente trabalho, uma das quatro concepções de analiticidade criticadas por Quine em *Dois dogmas do empirismo* se relaciona com o que ele chamou de segundo dogma do empirismo, ou seja, o reducionismo.

O Positivismo Lógico almejava fazer uma filosofia científica e para alcançar tal fim este grupo apresentou teses segundo a qual visavam mostrar porque a filosofia tinha falhado em ser científica. Uma destas teses foi a teoria do significado, que é composta de dois princípios:

(1) O princípio de verificação: O significado de uma sentença é seu método de verificação ou confirmação;

(2) Os enunciados da lógica e da matemática, juntamente com enunciados que explicam as relações de significados, são analíticos no sentido de que eles são verdadeiros unicamente em virtude de significados e não fornecem nenhuma informação sobre o mundo.

O princípio de verificação tinha o objetivo explícito de mostrar porque a Filosofia, em particular, a metafísica, foi mal sucedida. Nas palavras de Quine:

O dogma do reducionismo, mesmo em sua forma mais atenuada, está intimamente ligada a outro dogma: o de que há uma separação entre analítico e o sintético. Com efeito, fomos levados deste problema ao primeiro por meio da teoria verificacionista do significado. De modo mais direto, um dogma claramente apoia o outro da seguinte forma: enquanto se considerar que em geral há sentido em falar de confirmação e invalidação de um enunciado, parece ter

sentido falar também de um tipo-limite de enunciado que é confirmado vacuamente ipso facto, aconteça o que acontecer, e tal enunciado é analítico. (QUINE, 2011, p. 65).

Para Quine a teoria verificacionista do significado é um método da qual se valiam os Positivistas Lógicos para confirmar ou invalidar um enunciado. Dentro da perspectiva desta teoria, os enunciados analíticos se apresentam como aqueles que podem ser confirmados em todos os casos.

Refinando o conceito, tendo em mente a noção de sinonímia requerida anteriormente para a analiticidade, teremos o seguinte conceito: um enunciado analítico pode ser considerado sinônimo de outro enunciado analítico se, e somente se, são semelhantes no que se refere ao método de confirmação ou invalidação empírica. (QUINE, 2011, p. 60). O dogma do reducionismo surge, então, do equívoco de se supor que todo enunciado significativo pode ser traduzido em outro enunciado sobre a experiência imediata.

Esse reducionismo foi esposado por Carnap em seu *Der Logische Aufbau der Welt*, mas suas raízes remontam ao empirismo de Locke e Hume. Entretanto, Carnap não chegou nem perto de efetivar a redução dos dados empíricos em construções lógicas.

Segundo Quine, o dogma do reducionismo persiste na medida em que ainda se cultiva a ideia de se que pode confirmar ou invalidar um enunciado isoladamente. Os enunciados analíticos, portanto, seriam aqueles aos quais estão associados um "domínio único de eventos sensoriais possíveis tais que a ocorrência de qualquer um deles aumentam a probabilidade da verdade do enunciado". (QUINE, 2011, p. 64).

A contraproposta de Quine a esse reducionismo é o holismo: a ideia de que os enunciados não podem ser confirmados ou invalidados um a um, de modo isolado; estes enunciados enfrentam o "tribunal da experiência" como um "corpo organizado".

Por outro lado, o holismo defendido por Quine, como uma alternativa ao reducionismo radical do Positivismo Lógico, não é uma consequência imediata da rejeição aos dois dogmas do empirismo, o próprio Quine admite que na prática o

44 | Reflexões sobre alguns problemas filosóficos

holismo acaba por ser uma questão de grau (gradualismo). Dito de outro modo, o suposto triunfo de Quine sobre o Positivismo Lógico deve ser posto em xeque.

Uma passagem de Dois dogmas do empirismo resume o argumento de Quine contra os dogmas cultivados pelos empiristas lógicos:

É óbvio que a verdade em geral depende tanto da linguagem como de fatos extralinguísticos. O enunciado "Brutus matou César" seria falso se o mundo tivesse sido diferente sob certos aspectos, mas também seria falso se "matou" tivesse o sentido de "gerou". Assim, é-se tentado a supor, em geral, que a verdade de um enunciado é de alguma forma decomponível em um componente linguístico e um componente factual. Dada essa suposição, parece em seguida razoável que, em alguns enunciados, o componente factual deva ser nulo; e estes são os enunciados analíticos. Mas, apesar de razoável a priori, simplesmente não foi traçada uma fronteira entre enunciados analíticos e sintéticos. Que tal distinção deva ser feita é um dogma não empírico dos empiristas, um metafísico artigo de fé. (QUINE, 2011, p. 59)

Desferido esse golpe fatal nas pretensões fundacionistas dos empiristas lógicos, Quine conclui que a consequência imediata da refutação dos dogmas é uma perspectiva antifundacionista que tem no holismo uma de suas principais características. As palavras do próprio Quine tornam-se necessárias para uma melhor compreensão de sua proposta.

A totalidade de nossos assim chamados conhecimento e crenças das mais casuais questões de Geografia e História até as mais profundas leis da Física atômica ou mesmo da Matemática pura e da Lógica, é um tecido feito pelo homem, que encontra a experiência apenas nas extremidades. Ou, mudando a imagem, a totalidade da ciência é como um campo de força, cujas condições limítrofes são a experiência. Um conflito com a experiência na periferia ocasiona reajustes no interior do campo. Os valores de verdade têm de ser redistribuídos em alguns de nossos enunciados. A reavaliação de alguns enunciados acarreta a reavaliação de outros, em função de suas interconexões lógicas, sendo as leis

da lógica, por sua vez, simplesmente certos enunciados adicionais do sistema, certos elementos adicionais do campo. [...]. (QUINE, 2011, p. 66-67).

Essa é a consequência da rejeição da divisão entre enunciados analíticos e sintéticos e da concepção verificacionista. O holismo semântico de Quine é o que ele definiu como sendo o pragmatismo para o qual a rejeição dos dogmas empiristas conduz.

A proposta holística de Quine é a de que a ciência (ou nossas teorias) se assemelha a um campo de força que está em contato com a experiência apenas em suas extremidades. Se há qualquer desacordo nessas extremidades podemos fazer ajustes de modo a tentar resolver o conflito.

Se reavaliarmos um enunciado, os outros também devem ser revisados, pois todos os enunciados estão ligados logicamente, não havendo espaço para enunciados privilegiados que podem ser usados para refutar a teoria.

Essa tese centro-periferia de Quine pode ser melhor explicitada como se segue... Diferentemente do que acreditavam os defensores da distinção analítico/sintético, que do ponto de vista metodológico tem como uma característica importante a bifurcação de nossos padrões de conhecimento, a saber, os padrões de aceitação, justificação e revisão de enunciados analíticos diferem dos padrões para a aceitação, justificação e revisão de enunciados sintéticos, Quine rejeita esta bifurcação em nossos padrões de conhecimento, alegando que nosso conhecimento está sujeito a ambos, pois estão sujeitos ao princípio de revisibilidade universal.

De acordo com tal princípio, não há enunciados que não possam ser revisados, mesmo as leis lógicas estão a mercê desse princípio. Desse modo, somos confrontados com outra característica da tese centro-periferia (o holismo) de Quine: seu caráter falibilista.

Essa tese ficou conhecida como tese Duhem-Quine, pois ela já havia sido proposta em 1906 por Pierre Duhem no livro *La Théorie Physique: Son Objet, As Structure*. É questionável o fato de se referirem desse modo a essa proposta, uma vez que como veremos, Quine simplesmente assimilou as teses de Duhem, o que não justificaria tal designação.

Na tentativa de deixar mais clara nossa crítica, deixemos Quine de lado e nos voltemos para Duhem no intuito de apresentar, mesmo que resumidamente, a proposta do cientista francês.

Um ponto importante é que da suposta rejeição dos dogmas do empirismo não pode se concluir que entrará em cena uma perspectiva holística. Principalmente porque a distinção analítico/sintético sai ilesa dos ataques de Quine, incluindo a versão linguística da analiticidade dos Positivistas Lógicos. Do fato de que em determinados casos não seja possível definir claramente uma noção, não se pode concluir que tal noção não exista que é o que Quine argumenta.

Muitas noções poderiam ser alvo das mesmas críticas que Quine avança contra a analiticidade, e, no entanto, essas noções nunca foram alvo do ceticismo dos filósofos, que em grande parte as mantiveram em uso. Um exemplo é a noção de substância, que sobreviveu desde Aristóteles até a Modernidade. A controversa noção metafísica de Deus sobrevive da Idade Média até a Modernidade, servido como fundamento para muitas teorias filosóficas, como é o caso de Descartes e Berkeley, dentre outros.

E no caso da noção de analiticidade, sua longevidade é ainda mais admirável. Logo, a rejeição da analiticidade e a suposta passagem para uma perspectiva holista, carecem de melhores argumentos. Após as considerações acima sobre o pragmatismo, é chegado o momento de se voltar para Quine e apresentar o motivo pelo qual é um equívoco classificá-lo como parte da tradição pragmatista que tem como seu fundador Charles S. Peirce.

As consequências que Quine deduz de seu ataque aos dogmas do empirismo, a saber, a analiticidade e o reducionismo, são tão controversas quanto os argumentos dos quais eles advêm. Os defensores da analiticidade se mostraram eficazes em mostrar que, diferentemente do que Quine alega, não há nada de errado com o termo "analítico", mas o problema reside na ambiguidade das palavras que compõem as linguagens naturais.

Mas as questões a que este trabalho se propõe responder permanece sem resposta: qual a relação entre pragmatismo e holismo? Qual a concepção de pragmatismo Quine está fazendo referência? No que diz respeito a primeira questão,

uma passagem de um livro de Durkheim sobre o pragmatismo pode ser um ponto de partida interessante para se formular uma resposta adequada.

Lembremos que a argumentação pragmatista é a seguinte: a verdade implica a distinção de elementos; o real é feito de indistinção; assim, a verdade não pode expressar o real sem apresentar como sendo distinto o que não é, em poucas palavras, sem desnaturar o real. A realidade forma uma massa unificada, onde tudo se liga sem separação radical. O que emana de uma parte repercute no todo. É, então, somente por abstração que separamos uma parte do todo. O conceito, ao contrário, é limitado, determinado, nitidamente circunscrito; o mundo dos conceitos é descontínuo e distinto. Assim, há heterogeneidade entre o conceito e o real. (DURKHEIM, 2004, p. 203).

Se a interpretação de Durkheim está correta, então, em certo sentido, os pragmatistas defendiam uma forma de holismo. A questão é saber se pode-se caracterizar essa ideia como um holismo semântico, ou seja, a [...] "tese metafísica acerca da natureza da representação na qual o significado de um símbolo é relativo ao sistema inteiro de representações que o contém" (AUDI, 2006, p. 472). A descrição de Durkheim dificilmente se refere a um holismo semântico.

A concepção de significado de Peirce, por exemplo, está ligada a sua teoria semiótica e, nesse sentido, não se pode falar em um holismo semântico peirceano. Por outro lado, se levarmos em conta que Peirce defendeu a ideia de uma semiose infinita (ação do signo) e que tudo no mundo era um signo, inclusive o ser humano, tem-se, assim, uma ideia de que para algo ser um signo há a necessidade desse suposto signo está ligado a um sistema de representações que lhe permita funcionar como tal.

O alegado pragmatismo ao qual Quine faz referência em seu texto talvez não seja propriamente o pragmatismo defendido por Peirce, James e Dewey, mas apenas a compreensão de que há critérios e consequências práticas envolvidas nas nossas teorizações em relação ao mundo, a tal ponto que não se pode sustentar verdades imunes ao tribunal da experiência, dado que as consequências práticas envolvidas podem variar caso haja alguma mudança em nossos esquemas conceituais.

Por outro lado, é preciso ressaltar que o holismo de Quine foi duramente criticado. Segundo Quine (esposando a tese de Duhem) as hipóteses científicas devem ser confrontadas com a experiência em sua totalidade, uma vez que não há enunciados privilegiados que são verdadeiros independentemente do modo como concebamos o mundo que tem o poder de condenar toda uma teoria.

Assim, até mesmo a lógica é passível de revisão. Se há qualquer desacordo da teoria com a experiência podemos fazer determinadas modificações a fim de promover a conciliação. O valor de verdade de um enunciado, desse modo, pode variar, dependendo das necessidades de ajuste que vão surgindo no confronto da teoria com a experiência.

Entretanto:

But the rejection of reductionism does not entail the acceptance of holism. The view that it does is plausible only if one begins by conceiving significance in terms of a "confirmation range" or something similar. If, however, we do not conceive of significance in that way, but think of it, as suggested earlier, in terms of logical role or synthetic consequences, then the fact that certain sentences do not have confirmation ranges would not show that they are non-significant as statements. Hence we would not be forced to look for significance only in larger groups of statements or ultimately in the whole of all accepted statements. Thus we need not adopt holism as a theory of significance simply because reductionism proves untenable. (HOFSTADTER, 1954, p. 412-413).

O holismo, desse modo, não pode ser uma consequência necessária da negação do reducionismo, é apenas uma alternativa ao qual nos permite outra abordagem das teorias. Em 1973, Dummett levantou algumas objeções contra a posição holística de Quine, e que assume uma estrutura de centro/periferia. Para Dummett a distinção analítico/sintético sobrevive nessa tese de Quine. O modelo centro/periferia de Quine nos fornece noções segundo as quais a distinção analítico/sintético pode ser bem definida. Os enunciados analíticos seriam aqueles enunciados que são imunes à experiência e os enunciados sintéticos seriam aqueles enunciados que estão sujeitos ao modo como a experiência se apresenta.

Dummett defende, assim, que esta versão da distinção analítico/sintético não corre o risco de ser acusada de ser vazia, pois não há enunciados analíticos nem enunciados sintéticos, segundo a noção acima, estando em conflito com o modelo centro/periferia de Quine.

Assim, a conclusão que se pode chegar é que a recusa da distinção analítico/sintético não se sustenta, pois acaba indo de encontro com seu holismo que tem a forma centro/periferia. O problema reside no fato de que o núcleo desse "campo de força" proposto por Quine é altamente teórico. Uma vez que somente a periferia está em contato com experiência, a distinção entre duas classes de enunciados pode ser mantida.

Considerações finais

A questão de determinar se Quine é ou não um pragmatista e qual a relação entre holismo e pragmatismo ainda permanece sem uma resposta definitiva. Neste trabalho buscou-se apenas esboçar as contradições internas do problema. Em todo caso, a resposta em si não obscurece a importância de Quine para a Filosofia e mesmo para a Lógica e a Matemática.

Por outro lado, as indefinições quanto aos rumos do pragmatismo entre seus adeptos é outro obstáculo para uma resposta satisfatória. A resposta está, em certa medida, condicionada a escolha de um modo de se conceber o pragmatismo, e essa escolha implica que deveríamos tomar partido em um debate, que como vimos, não era tranquilo.

Isso não significa que se deva ignorar o pragmatismo por causa dessa falta de consenso entre seus pioneiros. A adoção de tal postura significaria um radicalismo desnecessário que teria como consequência a rejeição da própria filosofia, que padece do mesmo mal. Não é esse o caso. A própria natureza do pensamento filosófico é o que dá origem a essas divergências e isso é o que há de mais atraente nessa forma de saber.

Referencias bibliográficas

DE WALL, C. **Sobre Pragmatismo**. Trad. Cassiano Terra Rodrigues. São Paulo: Edições Loyola, 2007.

DUMMETT, M. **Frege: Philosophy of Language**. 2nd. Cambridge, Mass.: Harvard University Press, 1981.

HOFSTADTER, A. The Myth of the Whole: an examination of Quine's view of knowledge. **Journal of Philosophy**, 51, 1954, p. 397-417.

LEPORE, E. Holismo semântico. *In*: AUDI, Robert (dir.). **Dicionário de Filosofia de Cambridge**. Trad. João Paixão Netto *et al.* São Paulo: Paulus, 2006, p. 472-473. (Coleção dicionários).

PEIRCE, C. S. **The Essential Peirce 1**. Nathan Houser e Christian Kloes. (eds.). The Peirce Edition Project. Bloomington, Indiana: Indiana University Press, 1992.

PEIRCE, C. S. **The Essential Peirce 2**. Nathan Houser e Christian Kloes. (eds.). The Peirce Edition Project. Bloomington, Indiana: Indiana University Press, 1998.

QUINE, W.V.O. Dois dogmas do empirismo. *In*: QUINE, W.V.O. **De um ponto de vista lógico**. São Paulo: Editora Unesp, 2011, p. 37-72.

SHOOK, J. **Os Pioneiros do Pragmatismo Americano**. Trad. Fábio Said. Rio de Janeiro: DP&A, 2002.

4. LIBERDADE E MORALIDADE EM KANT



<https://doi.org/10.36592/9786554601061-04>

*Fernando Facó de Assis Fonseca*¹

Resumo

O objetivo do trabalho é compreender como Kant fundamenta a liberdade humana a partir do princípio da razão prática. A ideia é demonstrar que o homem só pode ser efetivamente livre na medida em que segue a determinação da lei moral postulada pelo princípio de sua razão. Para isso, temos que: 1) compreender o conceito de razão pura prática, desenvolvido na "Fundamentação da Metafísica dos Costumes", 2) compreender o conceito de Liberdade como *factum* da Razão, desenvolvido na "Crítica da Razão Prática"; 3) Compreender o conceito de liberdade na perspectiva da *Aufklärung*, tal como Kant desenvolve no seu ensaio "O que é o Iluminismo?".

Palavras-chaves: Kant, Liberdade, Moral, Razão, Esclarecimento.

Abstract

The objective of this work is to understand how Kant grounds human freedom based on the principle of practical reason. The idea is to demonstrate that a person can only be effectively free to the extent that they follow the determination of the moral law postulated by the principle of their reason. To achieve this, we need to: 1) comprehend the concept of pure practical reason, developed in "Groundwork of the Metaphysics of Morals"; 2) understand the concept of Freedom as a factum of Reason, developed in "Critique of Practical Reason"; 3) grasp the concept of freedom from the perspective of Enlightenment, as Kant develops it in his essay "What is Enlightenment?".

Keywords: Kant, Freedom, Morality, Reason, Enlightenment.

¹ Professor temporário do curso de Filosofia da Universidade Estadual do Ceará. Doutor em Filosofia da Educação pela Universidade Federal do Ceará. Mestre em Filosofia pela Universidade de Toulouse/FR; Mestre em Filosofia pela Universidade Federal do Ceará; Graduado em filosofia pela Universidade Federal do Piauí. Graduado em Filosofia pela Universidade Federal do Piauí. Graduado em psicologia pelo Centro Universitário de Brasília.

Introdução

O tema da liberdade humana é, por assim dizer, o fio condutor que une, de uma ponta a outra, a história da filosofia, mas que nasce mesmo antes dela. A bem da verdade, a questão da liberdade é uma questão especialmente grega, todavia, ela não se reduz somente ao âmbito da filosofia. Desde a Grécia agrária, *pré-democrática*, podemos identificar o tema da liberdade a partir da preocupação dos gregos em relação ao ideal de formação (*Paideia*) do indivíduo mediante um horizonte cultural antropocêntrico. O que distingue, portanto, os gregos, na visão de Jaeger Werner (1994), das outras formas de organização social é justamente o princípio formativo que visava à emancipação do homem através da conquista plena de suas virtudes (*areté*) latentes. Ora, a partir disso, toda a cultura grega, e por extensão, a cultura ocidental, terá como *leitmotiv* o princípio da emancipação humana.

A filosofia não se esquivou do assunto. A diferença fundamental entre a Grécia pré-filosófica e a Grécia filosófica será a discussão sobre qual base (cultural, epistemológica, ontológica) o princípio de emancipação estará assentado. Se na Grécia arcaica, mitológica, essa base fundava-se nas virtudes das epopeias cavalheirescas dos heróis gregos, como a coragem e a honra, para a filosofia, esse princípio encontrar-se-á fundado na esfera da razão (*logos*). Sendo assim, o homem livre torna-se, com efeito, aquele que se utiliza de sua faculdade racional para estabelecer sua ética e sua moral. A questão fundamental para os antigos não era simplesmente a epistemologia enquanto categoria instrumentalizadora do mundo. A questão era primordialmente coincidir *logos* e *eudaimonía*, ou seja, razão e felicidade, ou bem-viver. Isso permitia uma plena aliança entre ser e dever-ser. Em resumo, o dever-ser, a base ética do homem civilizado, era, a rigor, já pré-estabelecido pela razão cósmica universal. Isso se revela desde Sócrates.

Na modernidade, essa relação harmônica entre ser e dever-ser é drasticamente rompida. Com o advento da ciência, a razão é deslocada para a esfera subjetiva, desvinculando-se, portanto, da esfera cósmica. O ser torna-se pura relação factual de causa e efeito, cujo estatuto não impõe mais ao homem qualquer inclinação de natureza ética. Para dizê-lo de outro modo, o que se constata na era

moderna é que a realidade nada mais tem a ver com ética, mas apenas com a correlação entre fatos. A ética deve, portanto, ser deslocada para o âmbito exclusivamente privado, enquanto a compreensão da realidade, postulada matematicamente, deve, por sua vez, garantir a objetividade factual. Isso se prolonga até o período que precede a filosofia de Kant.

A genialidade de Kant foi ter reestabelecido um fundamento moral para a humanidade sem negar a guinada radical que a ciência moderna propiciou ao conhecimento da realidade. Kant foi capaz de extrair, de onde parecia impossível, uma moral objetiva cuja matriz provinha do próprio modo como a ciência operava, ou seja, através dos juízos sintéticos a priori. Esse é o motivo da importância de Kant para a filosofia prática. Kant visava uma moral que fosse tão objetiva quanto o tipo de conhecimento que estava sendo produzido pelas ciências modernas, mais especificamente, pela mecânica newtoniana. Diante disso, nosso intuito é, no que se segue, compreender como se articula a relação entre liberdade e moralidade na filosofia de Kant, visando superar o impasse estabelecido com o advento das ciências modernas entre ser e dever-ser.

A motivação moral em Kant: uma crítica da razão prática pura

Em sua "Fundamentação da Metafísica dos Costumes", Kant (2011b) busca uma fundamentação para a ação moral no campo da razão pura. Para Kant, uma ação moral só pode ser fundamentada mediante uma razão estabelecida a priori enquanto causa primeira da ação prática. Segundo ele, somente a razão é capaz de fornecer o princípio incondicional e a priori da moralidade, cuja forma deve ser válida para todo ser racional. Por isso, podemos designar a "Fundamentação..." não propriamente como uma crítica da razão prática, mas, sim, como uma crítica da razão prática pura.

De acordo com o filósofo (KANT, 2011b), a boa vontade consiste no único bem incondicional de ajuizamento moral. Trata-se de um princípio comum a todo homem, e exprime-se como produto de sua consciência moral. A boa vontade é a expressão direta da consciência moral de todo e qualquer indivíduo dotado de racionalidade, atribuindo a este a capacidade inata de avaliar sua ação de acordo com um juízo

54 | Reflexões sobre alguns problemas filosóficos

moral autônomo e racional. Dito de forma mais simples, para que a vontade seja compreendida como “essencialmente” boa, ela deve estar sustentada pelo princípio da razão pura. Do contrário, a vontade tornar-se-ia limitada por causas patológicas ou exteriores a ela mesma, o que a converteria, por conseguinte, num princípio condicionado. Segundo Kant, para que uma vontade seja verdadeiramente autônoma, ela deve ser, antes de tudo, *causa sui*, causa de si mesma, ou seja, incondicionada.

Ocorre que, no homem, essa vontade é atravessada por sua natureza híbrida. O ser humano possui tanto uma natureza racional como uma natureza patológica. Nem somente divino nem somente animal, o ser humano está ontologicamente sujeito a esses dois domínios. Mas é somente por essa sua condição ontológica híbrida, que o homem apresenta-se como um ser moral.

Se o homem possuísse apenas uma natureza racional e transcendente, não haveria justificativa para se discutir sua condição moral, uma vez que seu querer corresponderia integralmente a sua razão. Por outro lado, se sua natureza fosse apenas de ordem animal, sua ação seria conduzida expressamente pelos seus desejos patológicos. Por isso, a boa vontade é a forma da vontade quando submetida ao princípio da razão pura que, como tal, exprime uma tensão constante entre as dimensões às quais o homem está submetido: racional e animal. Como assinala Paul Guyer:

“O imperativo categórico é a forma que o princípio fundamental da moralidade toma quando aplicado a criaturas racionais imperfeitos como nós: embora este princípio seja originado somente por nossa própria razão, e não externamente imposto sobre nós por nenhuma outra divindade ou lei humana, ele só pode aparecer como coação porque nós também possuímos inclinações que se não for controlada, nos levam a agir contra ele” (GUYER, P., 2006, p. 35).

Em outras palavras, Para Kant, somente o homem é um agente efetivamente moral, pois somente ele está sujeito à dor infligida pela disciplina do imperativo categórico.

Temos, todavia, que diferenciar, antes de seguir adiante, dois sentidos de vontade em Kant. Vontade no sentido *lato* da palavra, que pode ser compreendida como sendo a faculdade de desejar em geral; e vontade no sentido *stricto*, referindo-se à forma do imperativo da razão agindo sobre a faculdade de desejar. Vontade enquanto desejo pode ser compreendida como arbítrio, ao passo que vontade, como princípio formal e incondicional, é postulada somente pela razão.

A vontade vincula-se, pois, ao princípio moral de conduta, enquanto o arbítrio refere-se à ação determinada, ou não, por tal princípio. Assim, para Kant (2011b, p.67), a vontade consiste na "faculdade de determinar a si mesmo a agir em conformidade com a representação de certas leis". Isso significa que a vontade é uma faculdade da razão que tem o poder de impor à ação uma forma específica conforme a representação de uma lei moral.

Dito isto, podemos inferir que a boa vontade é o princípio formal da vontade pura, racional e a priori, que age sobre o arbítrio determinando a forma da ação a partir da representação da lei moral.

Além disso, é preciso afirmar também que a ética kantiana é uma ética deontológica, pois *deon* – palavra de origem grega – significa, literalmente, dever. Kant estabelece, portanto, uma ética do dever, e não uma ética dos fins.

Para compreender melhor esse princípio deontológico da ética kantiana, precisamos estabelecer o que Kant compreende como princípios práticos. Por princípios práticos, Kant (2011) compreende as determinações gerais da vontade. São, basicamente, regras gerais que desembocam em formas específicas de determinações. Assim, os princípios práticos são regras gerais prescritos pela determinação da razão, mas que podem ser decompostos em regras específicas e contextualizadas. Podemos dividir os princípios práticos em dois grupos: máximas e imperativos.

As máximas são princípios práticos voltados para o próprio sujeito que as profere. Logo, são subjetivas e não universalizáveis. Já os imperativos, são princípios práticos objetivos, que valem para todos os homens, e não apenas para o sujeito da ação. Eles são, assim, mandamentos ou leis que exprimem a necessidade objetiva da práxis, vinculando a vontade racional à forma específica da ação. Como tal, uma ação pautada no imperativo da vontade é uma ação que visa à objetividade universal.

Contudo, Kant (2011) ainda divide o imperativo em outros dois grupos: hipotéticos e categóricos.

Imperativos hipotéticos são aqueles cuja ação aspira um fim ou objetivo. Isso significa que, ainda que o imperativo hipotético valha de maneira objetiva, ele continua sendo condicionado em função de determinado objetivo. Sua fórmula pode ser colocada da seguinte maneira: é válido para todos que, *se queres* alcançar um determinado objetivo, *deves* agir de tal ou qual maneira. Ocorre que, para alcançar tal fim, ainda que posto de maneira objetiva, ele deve continuar, todavia, sendo uma determinação subjetiva, isto é, uma determinação do próprio querer do agente, e não uma necessidade universal. Isso torna, por conseguinte, o princípio hipotético do imperativo uma determinação particular. Os imperativos hipotéticos não são, portanto, uma lei da razão, mas uma prescrição da razão vinculada a um objetivo específico, particular.

Em contrapartida, o imperativo categórico se impõe como um dever sem fim, um dever não condicionado a nenhuma vontade específica, mas a uma determinação da vontade na sua forma pura, incondicional. Sua fórmula, diferentemente do imperativo hipotético, determinado por um fim (*se queres, debes*), pode ser postulada como uma necessidade sem fim, do tipo *deves, então debes*. O imperativo categórico desvincula, pois, a vontade de uma determinação particular do agente. Em vez disso, clama por uma vontade pura, racional, totalmente absoluta e incondicional. A partir do imperativo categórico, Kant (2011) atinge seu propósito, que consiste em pensar uma moral cuja forma do dever seja determinada de maneira completamente objetiva e universal. A lei moral restringe-se, pois, à forma do imperativo categórico, determinada por um dever incondicional à forma pura da vontade racional.

Importante salientar que o filósofo considerou o dever moral não como *müssen*, mas como *sollen*. Em alemão, *müssen* refere-se à necessidade natural de eventos que não podem, sob nenhum efeito, atuar de outra forma. *Sollen*, por sua vez, é compreendido, por Kant, como necessidade, mas que não se refere à uma necessidade restritiva, mas à possibilidade. Assim, o dever moral (*sollen*) apesar de se impor de maneira incondicional, não pode ser confundido com uma necessidade natural. Trata-se, portanto, de uma necessidade da razão que rompe com o princípio

da causalidade, sendo esse o motivo pelo qual, como veremos adiante, a liberdade torna-se um fundamento da ação moral.

Logo, a partir do entendimento sobre máximas e imperativos, podemos compreender como Kant formula a máxima da lei moral.

A fórmula do imperativo que Kant estabelece tanto na "Fundamentação" como na "Crítica da Razão Prática" é a seguinte: *Age de tal modo que a máxima de tua vontade possa valer sempre, ao mesmo tempo, como princípio de legislação universal*. Percebemos aqui, após o que foi dito acima, a passagem de uma máxima para um imperativo. A ideia de Kant não é impor ao sujeito uma lei de fora, heterônoma, e, sim, inferir que o sujeito é capaz de impor a si mesmo, através de sua máxima, ou seja, de uma autoimposição subjetiva, uma lei autônoma que sirva efetivamente como lei universal. Então, o que Kant propõe com a sua ideia de imperativo categórico, extraído de sua própria atividade racional, é a sobreposição entre máxima e imperativo, desde que este não seja condicionado hipoteticamente a um objetivo privado, mas que seja vinculado ao dever puro e simples da lei moral.

Liberdade como *Factum* da razão: a passagem da crítica da razão pura para a crítica da razão prática

Após o que foi dito acima, passemos agora a pensar a liberdade em Kant traçando um percurso que vai da sua primeira crítica, "A Crítica da Razão Pura", para a sua segunda crítica, "A Crítica da Razão Prática".

A reviravolta epistemológica em Kant, conhecida em sua primeira crítica como *revolução copernicana*, redimensiona completamente o lugar que o conhecimento ocupa na tradição. O conhecimento, para a filosofia, não é mais o meio pelo qual alcançamos a realidade exterior: ele torna-se um fim em si mesmo. Isto é, a tarefa da filosofia não é mais acessar o núcleo ontológico da realidade exterior, mas se perguntar sobre as condições de possibilidade de todo conhecimento possível. Para Kant, nem o racionalismo nem o empirismo foram capazes de compreender a verdadeira natureza do conhecimento, já que permaneceram enredados no discurso metafísico dogmático cujo teor pressupunha haver uma coincidência entre pensamento e realidade.

Sendo assim, o primeiro acreditava que, partindo dos axiomas necessários da razão, seria possível deduzir os predicados determinantes da realidade. Kant denominou esse procedimento de juízo analítico *a priori*. O segundo supunha que o conhecimento se dava de forma passiva, acedendo aos axiomas da razão por meio de sínteses da experiência. Kant denominou esse procedimento de juízo sintético *a posteriori*. A inovação kantiana corresponde a um terceiro tipo de juízo, um que articula tanto a forma ativa *a priori* da razão com a forma passiva *a posteriori* da experiência: o juízo sintético *a priori*. Para Kant, a ciência moderna se utiliza desse tipo de juízo, uma vez que ela não opera por uma mera coleta de dados caóticos disponíveis na natureza (o que a tornaria cega), nem através de uma mera formalidade, ignorando a experiência (o que a tornaria vazia). O procedimento científico consiste numa síntese *a priori*, cujo objetivo principal é fazer a natureza responder às perguntas colocadas pela ciência.

A preocupação de Kant na sua *Crítica da Razão Pura* é, pois, pensar a possibilidade de uma metafísica construída nos moldes da ciência moderna, isto é, segundo o modo do juízo sintético *a priori*. Para isso, Kant estabelece dois domínios de fundamentação do conhecimento: a Estética transcendental e a Lógica transcendental. A Estética seria a doutrina do conhecimento sensível e de suas formas *a priori*. A Lógica trata do conhecimento intelectual e suas formas, também, *a priori*. A lógica se divide em analítica transcendental e dialética transcendental. A primeira refere-se às formas puras do conhecimento que fundamenta a experiência sensível. A segunda investiga os três elementos essenciais de toda metafísica: alma, mundo, Deus, apontando para suas impossibilidades inerentes de serem conhecidas, dado sua natureza puramente transcendente, isto é, não empírica. A primeira parte da *Crítica da Razão Pura* é dividida em três partes menores. A primeira trata da estética transcendental, a segunda da analítica transcendental, e a terceira da dialética transcendental, onde ele trata da questão da liberdade. Façamos rapidamente de cada uma delas.

Kant serve-se da palavra estética referindo-se a sua raiz grega *aisthesis*, que significa sensação ou percepção sensorial. Assim, o termo "Estética Transcendental" corresponde às formas puras e *a priori* que captam diretamente – isto é, de forma

não mediada – o fenômeno dado pela experiência. Tal como em Aristóteles, Kant discerne no fenômeno tanto uma forma como uma matéria. Mas, diferentemente do filósofo grego, a forma do fenômeno não é ontológica, mas subjetiva, ou seja, consiste no modo como o sujeito confere uma determinada forma ao conteúdo recebido da realidade. Isto é, não são formas do objeto, mas formas do próprio sujeito da experiência. A partir disso, Kant discerne duas formas puras da intuição: espaço e tempo.

O espaço consiste na forma do sentido externo, ao passo que o tempo corresponde à forma do sentido interno. Assim, o espaço apreende todas as coisas que aparecem externamente, enquanto o tempo compreende todas as coisas que aparecem internamente. Kant discorda de racionalistas como Leibniz, que definia o espaço como uma entidade ontológica. Se assim o fosse, poderíamos situar o espaço ao lado de outros fenômenos e conhecê-lo de fora dele mesmo. Situação impossível, segundo Kant, posto que o espaço consiste na própria condição de possibilidade de conhecimento dos todos os fenômenos sensíveis. O mesmo vale para o tempo: impossível sair do tempo para apreender o tempo atemporalmente. Tempo e espaço são, portanto, formas *a priori* da intuição inerentes ao próprio sujeito.

Todavia, enquanto o tempo e o espaço fornecem o modo de apreensão *a priori* do objeto de maneira intuitiva, ou seja, como conhecimento sensível, Kant teve de prover também um quadro de categorias e conceitos puros *a priori*, responsáveis pelo conhecimento intelectual, discursivo. A função do conceito é, pois, unificar e ordenar um múltiplo sob uma representação comum. O intelecto consiste, portanto, na faculdade de julgar, uma vez que julgar significa exatamente aplicar uma regra do entendimento a um caso particular. Como tal, as formas puras do entendimento sintetizam a partir de conceitos puros os dados puros já elaborados pelas formas da intuição.

Kant desenvolve uma tabela com doze categorias (ou conceitos puros), derivadas da “tábua das categorias” de Aristóteles. Mas, enquanto para Aristóteles essas categorias eram caracterizadas por *leges sentis* (modos do ser), para Kant as categorias do entendimento são *leges mentis* (modos de funcionamento do pensamento). Portanto, as categorias kantianas são definidas como sendo as

condições de possibilidade de pensar o objeto da experiência, do mesmo modo que o tempo e o espaço são determinados como sendo as condições de possibilidade de captar sensivelmente o objeto da intuição.

A questão que se coloca é: ora, se temos de um lado a síntese a priori da sensibilidade na forma de espaço e tempo, e por outro, a síntese a priori do entendimento na forma das categorias da razão, o que então sintetizaria em um ponto único todas as sínteses já estabelecidas? Esse princípio, segundo Kant, é o sujeito transcendental. O sujeito seria, portanto, o ponto supremo da síntese, a síntese de todas as categorias do entendimento, a unidade originária e definitiva, denominada de Apercepção transcendental. O sujeito é um pressuposto necessário, posto que *deve poder acompanhar todas as minhas representações*. Sem pressupô-lo, não haveria como tomar o múltiplo da experiência como sendo pertencente a um único agente. Deve-se deixar claro que ele não possui conteúdo. Trata-se de um ponto puramente formal, mas ainda assim real e efetivo.

Na parte da dialética transcendental, Kant (2010) menciona certas ideias metafísicas como sendo "objetos" próprios do pensamento (*Gedankending*) e cuja negação (antítese) não pode ser nem lógica nem privativa, mas fundamentalmente antinômica. São eles: o universo em sua totalidade, a imortalidade da alma e Deus. Segundo Kant (2010), tais objetos, embora sejam *logicamente possíveis*, isto é, pensáveis, não podem, todavia, ser considerados objetos de nosso conhecimento. Isso pelo fato de não serem objetos de nossa experiência sensível, ou de não fazer parte daquilo que nós experimentamos enquanto realidade.

Em outras palavras, se, em parte, trata-se de objetos pensáveis, são, todavia, objetos que não podem ser enquadrados dentro da categoria de *possibilidade*, já que Kant considera como *possível* (aqui não no sentido lógico, mas no sentido fenomênico estrito) somente aquilo que faz parte de nosso horizonte de experiência. Desse modo, não podemos abordar esses "objetos" da razão pura nem como uma não-coisa (*Un- Ding*), ou seja, enquanto contradição lógica, nem como objetos de nosso conhecimento sensível. Eles devem estar sujeitos a um outro tipo de investigação.

Podemos dizer, portanto, que, se o efeito de uma contradição lógica produz um objeto sem conceito, dado a impossibilidade lógica de concebê-lo, as antinomias produzem um conceito sem objeto, já que não se trata de impossibilidade lógica, mas de uma abstração sem referência sensível capaz de preencher seus contornos objetivos. O limite transcendental da finitude humana produz, portanto, as antinomias da razão pura, cujo princípio do terceiro excluído não exerce mais sua função opositiva.

Na verdade, as antinomias tratam de um conflito insuperável entre, de um lado, o que posso efetivamente conhecer a partir das minhas condições finitas, submetidas à dimensão fenomênica, e, de outro, o excedente da razão que segue naturalmente seu curso em busca dos princípios dogmáticos enquanto fundamento de toda experiência possível – articulados, por sua vez, na dimensão propriamente numênica. Por isso, nas palavras de Jean-Marie Vaysse: “Antinomias são produzidas pelas Ideias cosmológicas que se originam num silogismo hipotético e que nos fornecem a ideia de uma primeira hipótese numa relação de causalidade” (2012, p.12).

Kant procura, então, conduzir dialeticamente os conflitos da razão pura a seu paroxismo, apresentando como tese a aspiração da razão em constituir uma totalidade, e enquanto antítese, as refutações de cunho empirista, que, por sua vez, desembocam no ceticismo. Nesse sentido, no segundo capítulo da seção sobre a *Dialética Transcendental*, Kant (2010) trabalha as antinomias de maneira sistemática, designando quatro fundamentais e dividindo-as em dois grupos: as duas primeiras são antinomias matemáticas (pois são consideradas no homogêneo) e as outras duas, dinâmicas (pois se situam no heterogêneo).

Nas duas primeiras antinomias, a ideia de mundo é abordada segundo a quantidade e a qualidade. “Segundo a quantidade, a tese propõe um começo do mundo no tempo e no espaço, ao passo que a antítese nega tal começo”; segundo a qualidade, a tese afirma que toda substância composta culmina numa realidade simples, enquanto a antítese afirma que a divisão prossegue até o infinito (VAYSSE, 2012, p. 12). Para Kant (2010), no que diz respeito às antinomias matemáticas, ambas as formulações são consideradas falsas, pois lidam com um conceito

ambíguo de mundo, que se apresenta tanto como coisa em si como fenômeno espaçotemporal.

No que tange à terceira antinomia, Kant investiga a ideia de liberdade – e esse é o ponto que nos interessa. De forma homóloga, enquanto a tese afirma uma causa livre, a antítese defende que toda causa é condicionada ao infinito. E, por fim, ao tematizar a possibilidade do incondicionado, ao passo que a tese propõe sua existência, a antítese a nega. Quanto às antinomias dinâmicas, diferentemente das antinomias matemáticas, Kant sugere que ambas as formulações (tese/antítese) sejam plausíveis quando embasadas na distinção transcendental entre fenômeno e coisa em si: a antítese é verdadeira para os fenômenos, enquanto submetidas ao tempo-espaço; a tese é verdadeira para a coisa-em-si, para além das apreensões da intuição sensível, mas ainda assim pensadas (HOLZHEY; MUDROCH, 2005, p. 47).

Assim, Kant salva a liberdade mediante um impasse ontológico, na medida em que é possível pensá-la na sua dimensão numênica, separada, portanto, da dimensão fenomênica. Nesse sentido, na sua primeira Crítica, a “Crítica da Razão Pura”, Kant está preocupado em compreender como é possível a liberdade no espaço da razão pura prática, ou seja, como que a questão da liberdade se articula quando pensada no campo especulativo da razão teórica.

Ocorre que, na sua segunda Crítica, a “Crítica da Razão Prática”, Kant não se limita mais a pensar a liberdade no âmbito da razão pura prática, mas entender como ela incide sobre a vontade e a ação. Em outras palavras, Kant procura explicar como a liberdade, pertencendo à razão pura ou numênica, pode atuar na esfera fenomênica. Ou ainda, como a razão pode determinar a vontade de modo que ela aja não movida por princípios patológicos determinados pela relação de causa e efeito, aplicados à esfera fenomênica, mas por princípios puros que rompa com tal processo de causalidade. Esse seria, portanto, o sentido de liberdade para Kant.

Nesse sentido, Kant se vê obrigado a reformular o lugar que ele havia atribuído à razão na sua primeira crítica. Kant encontra um lugar para a liberdade no campo da razão especulativa, mas esta permanece sendo uma ideia platônica, situada no reino numênico. Acontece que Kant se preocupa agora em pensar como a liberdade é possível no reino fenomênico. Para isso ele tem de inverter a ordem das relações

que conhecimento e razão ocupavam na primeira Crítica. Como vimos, para que o conhecimento fosse possível, a razão teria de ser acoplada ao campo da experiência por meio de categorias formais ou conceitos. Mas na "Crítica da Razão Prática", a fim de pensar uma liberdade efetiva, Kant define a razão em primeiro plano, agindo de cima para baixo, ou seja, da razão para a realidade fenomênica. Com isso, Kant foi capaz de dar à razão um valor objetivo e prático, postulada na realidade como *factum*, isto é, como um fato que, como um raio, toma o indivíduo com um ser livre e autônomo capaz de romper com a cadeia causal que rege o campo fenomênico.

Assim, Kant (2011a) define a liberdade como um *factum* da razão:

Pode-se denominar a consciência desta lei fundamental um *factum* da razão, porque não se pode sutilmente inferi-la de dados antecedentes da razão, por exemplo, da consciência da liberdade (pois esta consciência não nos é dada precisamente), mas porque ela se impõe por si mesma a nós como uma proposição sintética *a priori*, que não é fundada sobre nenhuma intuição, seja pura ou empírica, se bem que ela seria analítica se se pressupusesse a liberdade da vontade, para o que porém se requereria como conceito positivo uma intuição intelectual, que aqui de modo algum se pode admitir. Contudo, para considerar esta lei como inequivocamente **dada**, precisa-se observar que ela não é nenhum fato empírico mas o único *factum* da razão pura, que deste modo se proclama como originariamente legislativa (KANT, 2011a, p. 52-53: grifo nosso).

Isso possibilitou a Kant compreender a razão prática como sendo o **fecho de abóboda**, a ponte entre a razão pura e a realidade fenomênica. Mas, isso não deixa de conter algumas limitações no campo especulativo para a liberdade. Ou seja, Kant veta a possibilidade de a razão prática ser uma razão hipotética, ou seja, uma razão cujos princípios podem ser deduzidos logicamente. Contudo, é preciso destacar que é esse mesmo limite que garante à liberdade a sua legitimidade, uma vez que conhecer as causas que nos determinam no campo da razão especulativa não seria diferente de conhecer as causas que nos determinam no campo da esfera fenomênica.

Como Kant afirma acima, a lei moral consiste na irrupção da própria razão, isto é, não está sujeita a nenhuma intuição, nem pura nem empírica. Essa definição torna-

se crucial, pois isso elimina a hipótese de uma referência cognitivista da liberdade kantiana, a hipótese de que seria possível encontrar o fundamento da liberdade na própria atividade reflexiva do sujeito finito. Podemos compreender esse *factum* da razão como sendo, em última análise, inexplicável, irracional, inimputável, portanto, impossível de ser fundamentado enquanto princípio racional segundo uma atividade puramente cognitiva, por meio da qual poderíamos reflexivamente alcançar os fundamentos de nossa própria liberdade. A liberdade é, pois, racional, mas não hipotética. Conseqüentemente, não podemos jamais supor um fundamento cognitivo para a liberdade. A liberdade é *causa sui*, fundamento de si mesma.

Liberdade como emancipação

Até agora, compreendemos, em primeiro lugar, como Kant formula o princípio categórico da razão mediante uma Crítica da razão pura prática. Compreendemos também, em seguida, como que a liberdade é fundamentada na esfera fenomênica, a partir do *factum* da razão. Contudo, ainda precisamos compreender como que a razão pode ser concebida a partir da própria esfera fenomênica. Ou dito de outra forma, devemos entender agora como a razão pode ser pensada de baixo para cima, e não apenas de cima para baixo. Por isso, abordaremos a sua famosa resposta para a pergunta sobre o que é o Esclarecimento. Vejamos.

Em 1773, Kant (2005) respondeu a uma pergunta que circulava entre os intelectuais da época e cuja resposta, até então, ainda não havia sido satisfatória. A questão que exigia uma reflexão dos pensadores era nada menos que: "O que é o Esclarecimento (*Aufklärung*)?". A ideia de Esclarecimento circulava por toda a Europa como uma ideia compartilhada já fazia algum tempo, mas ainda não havia uma definição adequada para ela. A resposta de Kant ficou consagrada, portanto, como sendo a definição mais bem acabada para do significado de esclarecimento, e que nos serve ainda hoje de modelo para pensarmos as suas bases.

Kant define sua ideia logo na primeira linha do ensaio e que vale reproduzir novamente. Diz ele: "Esclarecimento é a saída do homem de sua menoridade, da qual ele próprio é culpado" (2005, p.63). Aqui fica nítida a diferença entre a ideia de

liberdade em Kant para a ideia de ética em Sócrates. Para este último, a moral é associada à sabedoria, de modo que agir de maneira ética é agir de maneira reflexiva. O mal, para Sócrates, só existe enquanto ignorância. Mas para Kant, a menoridade do homem, sua cegueira moral, é completamente associada a uma vontade prévia. Por isso ele diz que a menoridade é culpa daquele que quer manter-se situado nessa posição.

Logo em seguida ele afirma: "A menoridade é a incapacidade de se fazer uso de seu entendimento sem a direção de outro indivíduo" (KANT, 2005, p. 63). Kant apresenta a fórmula da menoridade: não utilizar da própria capacidade reflexiva sem se fiar em qualquer ponte de apoio externo. Como ele diz logo em seguida: "O homem é o próprio culpado dessa menoridade se a causa dela não se encontra na falta de entendimento, mas na falta de decisão e coragem de servir-se de si mesmo sem a direção de outrem". Por isso, o filósofo recupera a máxima de Horácio, em latim, *sapere aude*, que significa: ousar saber, ou seja, coragem para pensar. É, portanto, assim que Kant define a ideia de *Aufklärung*.

Mas isso implica em várias consequências morais, que nos fazem meditar sobre como essa ideia de Esclarecimento pode ser aplicado não somente no âmbito individual, mas também, e sobretudo, no campo social. Nesse momento, Kant encontrava-se preocupado em como pensar a liberdade não somente como um juízo determinante, mediante o qual um comando da razão opera de cima para baixo, como se a ordem das coisas tivessem que se adequar à ordem da razão. Kant se preocupa também em pensar o estado de coisas da liberdade a partir do que ele chamou, na sua "Crítica da Faculdade do Juízo" (KANT, 2008), de juízo reflexivo. Isso significa que devemos operar nossa capacidade de julgar não como se a realidade fosse submissa ao princípio da razão, mas como se a razão fosse processada de baixo para cima, transformando a realidade a partir das condições e das configurações que a esta, a realidade, apresenta.

Em uma palavra, o juízo determinante é aquele que opera do universal ao particular, ao passo que o juízo reflexivo é o que opera do particular para o universal. No segundo modelo, o sujeito procura uma forma, no campo da realidade fenomênica na qual ele se encontra, de alcançar a dimensão universal da razão. Assim, o sujeito faz uso de sua razão, servindo-se dos meios disponíveis para tal quando imerso na

própria realidade. Por isso, a ideia de *Aufklärung* tende a unificar o numênico e o fenomênico, a razão e a realidade, de maneira gradual e processual.

Esse é o sentido da diferença estabelecida por Kant entre o uso da razão pública e o uso da razão privada. Para o filósofo, o uso da razão pública significa a capacidade de o indivíduo usufruir da sua razão para além das limitações a que este, o indivíduo, pode estar submetido quando cumprindo alguma função social. O uso privado da razão seria, por sua vez, a obediência a certo sistema de normas e coordenadas no interior das instituições sociais. Assim, ao contrário do que poderíamos imaginar, o uso público não é o uso comunitário da razão, ou seja, formas intersubjetivas de articulação da razão pública. O uso público é assegurado pela capacidade individual de transcender às regras e deveres impostos pela sociedade e, assim, apontar críticas que conduzam uma reflexão sobre tais ou quais preceitos institucionais vigentes.

Consequentemente, o uso privado da razão não se refere ao uso individual e privativo da capacidade de raciocinar do sujeito quando apartado do senso comunitário. Pelo contrário, o uso privado da razão, para Kant reside justamente nos sistemas normativos que nos são impostos de cima para baixo através das instancias sociais.

Mas isso não significa, todavia, que uma forma de uso da razão necessariamente anule a outra. Longe disso, Kant deixa bastante claro que seu propósito não é a anulação do uso público em detrimento do privado, ou vice-versa, mas a manutenção da tensão entre essas duas formas de uso. Por isso, ele apresenta a enigmática fórmula "Obedeça e pense!". Devemos ser bastante cautelosos ao analisar essa injunção em Kant. A fórmula aparentemente induz uma forma de submissão e obediência às normas instituídas na e pela sociedade, ao passo que o pensamento seria apenas uma forma abstrata de indignação pessoal ao aparato burocrático, mas sem efeito efetivo diante da realidade. Mas não é isso que Kant tem em mente.

Primeiramente é preciso dizer que Kant não é um revolucionário, mas um reformador. Ele desconfia das causas revolucionárias que impetuosamente destituem um líder, mas que não sabem o que fazer, ou o que pôr, no lugar, uma vez

que o pensamento, a estrutura racional do sistema, não foi posto em causa. Por isso, o modelo de uma sociedade em processo de esclarecimento deve ser lento e gradual. O uso público da razão não pode ser pensado como um ato revolucionário. Não cabe simplesmente apelar para o ato de desobediência civil. De acordo com Kant, é preciso continuar agindo no interior das instituições. Qualquer ação extrainstitucional é considerada pelo filósofo como uma medida de manutenção da menoridade, uma vez que, de uma maneira ou de outra, continuaríamos a obedecer a dogmas, nos coibindo de pensar a partir de nossa própria capacidade reflexiva.

Assim, o regime político que melhor comporta a capacidade para o progresso com base no Esclarecimento, de acordo com Kant, é o liberalismo republicano, uma vez que este abre a capacidade de articular o pensamento público no interior da sociedade, ainda que esta seja considerada limitada e imperfeita. Kant não está preocupado com a forma final da *Aufklärung*, mas em instituir o seu dispositivo de ação. Uma vez que esse dispositivo esteja operando, o processo se põe automaticamente em andamento. Por isso, é preciso, também, obedecer.

Mas quando Kant fala em obediência, ele visa manter a estabilidade da sociedade para que se possa construir uma base crítica para a mesma. Assim, só é possível pensar de forma crítica se houver um mínimo de estabilidade institucional, se as coisas estiverem minimamente em funcionamento. Uma universidade pública, por exemplo, que é o espaço de formação do pensamento crítico por excelência, deve incorporar no seu cerne essa tensão imanente. Existem diretrizes protocolares que devem ser seguidas no interior da instituição, mas ao mesmo tempo é preciso que se permita a existência de um campo de pensamento livre para que essas mesmas questões protocolares e outros processos institucionais não sejam excessivamente burocratizados. Na Igreja, essa liberdade já é menor, mas, ainda assim, isso não significa dizer que um padre ou um pastor não possam expor seus pensamentos e críticas para o corpo de homens letrados da sociedade. Em suma: na instituição, cabe a obediência; no âmbito público do pensamento, cabe a crítica.

Desse modo, podemos compreender a razão pública como uma espécie de vírus que aos poucos vai corroendo as bases institucionais, mas sem demoli-las por inteiro. O pensamento de Kant sobre a *Aufklärung* não pode ser compreendida como um pensamento simples. A ideia é complexa e paradoxal, e se quisermos sustentar

a ideia de liberdade e de emancipação no âmbito individual e no âmbito social, devemos, acima de tudo, suportar essa contradição.

Considerações Finais

A guisa de conclusão, gostaríamos de tecer algumas palavras finais sobre o que trabalhamos aqui.

A ideia geral foi pensar como se dá a relação entre moralidade e liberdade em Kant. Aparentemente, tais conceitos são antagônicos, uma vez que moralidade pressupõe obediência a certo sistema de normas, enquanto liberdade pressupõe justamente o contrário, isto é, a ruptura com tais sistemas de normas. No entanto, um olhar mais profundo nos mostra que uma ruptura arbitrária pode caracterizar em um novo enquadramento no interior de um sistema de normas alienantes. Por isso, Kant se viu na necessidade de pensar uma forma de liberdade que conjugasse com a moral. Mas como articular esses dois polos?

Kant estabeleceu que a única forma de sermos livres efetivamente é obedecendo à nossa própria razão. Em outras palavras, seria obedecer as nossas próprias leis, formuladas pelos princípios da razão. Isso tornaria o homem um ser moral, uma vez que sua obediência estaria condicionada à forma da imposição da razão pura sobre a ação.

Kant fornece essa possibilidade analisando como a prática pode estar submetida à razão pura. Por isso, Kant estabelece o princípio de vontade boa, como sendo a vontade determinada pela razão. Ao pressupor uma vontade determinada pela razão, e não pelos desejos, Kant introduz uma ideia de vontade pura e formal, destituída de conteúdo patológico. Isso o obriga a determinar a vontade como dever à lei. Por isso sua ética é uma ética deontológica, já que o princípio de obediência moral não está assegurado por um fim ou horizonte último, mas pela simples obediência ao dever moral.

Mas até o período de sua primeira Crítica, Kant havia estabelecido a liberdade como uma antinomia dinâmica, abrindo a possibilidade de o numênico coexistir com o reino dos fenômenos. Para ele, seria a única possibilidade de abordar o tema da

liberdade, já que, na esfera da realidade fenomênica, nós somos completamente determinados pela relação de causa e efeito. Nesse reino, não há espaço para a liberdade. Porém, na sua segunda Crítica, o filósofo já abre a possibilidade de pensar a liberdade como um fecho de abóboda, que sela a lacuna antinômica entre númeno e fenômeno. Para isso, ele prioriza a razão sobre a realidade fenomênica, coisa que na primeira Crítica se dava justamente de modo inverso. Por isso, a razão age como *factum*, se impondo de cima para baixo, ou seja, do númeno ao fenômeno, e assim quebrando a consistência da cadeia de causa e efeito.

Mas ainda assim, temos de pensar a liberdade para além do campo puro da subjetividade e refletir como ela pode ser articulada com o mundo das coisas fenomênicas. Em outras palavras, é importante não só partir de uma lei moral que impõe uma ruptura com a realidade fenomênica – abrindo assim o campo da liberdade como sendo a forma numênica da razão agindo sobre a realidade – mas também pensar como que *in media res*, no meio do processo, podemos articular uma ideia de liberdade. Isso significa que passamos a utilizar não mais a ideia de juízo determinante, que parte do universal para o particular, mas de juízo reflexivo, que parte do particular em busca do universal.

No seu ensaio sobre “O que é Esclarecimento”, Kant trabalha a ideia de emancipação, como sendo a saída do homem de sua menoridade, da qual ele mesmo é culpado. Para sair da menoridade, Kant estabelece a relação entre obediência e pensar. Obediência aqui está relacionada com o que ele compreendia como o uso da razão privada, ou seja, o uso que fazemos das leis e normas instituídas socialmente, e que determinam o campo da realidade em seu aspecto coeso e sistemático. O pensar, por sua vez, tem a ver com o uso da razão pública, que vincula o sujeito à lei universal da razão. Nesse momento, Kant não está preocupado em determinar a liberdade categoricamente, nem como a priori racional nem como *factum* da razão. Sua preocupação é em pensar a relação tensional entre razão pública e razão privada. Em outras palavras, para Kant, a liberdade, enquanto emancipação, não se dá de uma vez, mas gradualmente, na medida em que o homem, aos poucos, sai da posição de menoridade a partir do uso de seu juízo reflexivo, construindo as bases para sua maioridade moral.

Referências

GUYER, P. **Kant**. London/New York: Routledge, 2006.

HELIO, Souza. **O problema da motivação moral em Kant**. Editora UNESP. Edição do Kindle

HÖFFE, O. **Kant: Crítica da razão pura** – Os fundamentos da filosofia moderna. São Paulo: Loyola, 2013.

HOLZHEY, H.; MUDROCH, V. **Historical Dictionary of Kant and Kantianism**. Lanham, Maryland: The Scarecrow Press, 2005.

JAEGER, W., W. **Paideia: a formação do homem grego** São Paulo, Martins Fontes, 1994.

KANT. I. **Textos Selectos**. Rio de Janeiro: Vozes. 2005.

_____. **Crítica da Faculdade do Juízo**. Rio de Janeiro, Forense Universitário, 2008.

_____. **Crítica da Razão Pura**. Lisboa: Fundação Calouste Gulbenkian, 2010.

_____. **Crítica da Razão Prática**. São Paulo: Martins fontes, 2011a.

_____. **Fundamentação da Metafísica dos Costumes**. Lisboa: Edições 70, 2011b.

PASCAL, G. **O pensamento de Kant**. Rio de Janeiro, Ed. Vozes, 1990.

VAYSSE, J. M. **Vocabulário de Immanuel Kant**. São Paulo: Martins Fontes, 2012.

5. CARL SCHMITT, UM HEGELIANO? A LEITURA CRÍTICA DE JEAN-FRANÇOIS KERVÉGAN SOBRE O POLÍTICO, A GUERRA E O ESTADO TOTAL



<https://doi.org/10.36592/9786554601061-05>

Deyvison Rodrigues Lima¹

José Maria Arruda²

Jean-François Kervégan realiza uma tarefa complexa em seu texto *Hegel, Carl Schmitt. Le politique entre spéculation et positivité (Hegel, Carl Schmitt. O político entre a especulação e a positividade)*: expõe nas entrelinhas uma filiação não assumida, privilegiando a compreensão de alguns temas schmittianos à luz de Hegel, em vez de Hobbes, autor associado com maior ênfase na interpretação entre os *scholars* do jurista alemão. Ao realizar uma "*lecture 'en miroir' de deux corpus*", Kervégan aposta que a relação entre Hegel e Schmitt seria útil para a compreensão de zonas obscuras dos pensadores: o decisionismo schmittiano serviria para compreender alguns aspectos da teoria hegeliana, enquanto esta, por sua vez, serviria para esclarecer alguns desvios e recusas da obra schmittiana. A tese do autor franco-argelino sobre o jurista pode ser apresentada a partir da hipótese de que "a obra de Schmitt visa ser a realização – no terreno da positividade e em condições ético-políticas profundamente modificadas – das posições fundamentais da filosofia hegeliana do direito e do Estado" (KERVÉGAN, 2006, p. 135). De qualquer forma, para Kervégan, "a referência à filosofia hegeliana desempenha um papel determinante na própria constituição da problemática de Carl Schmitt" (KERVÉGAN, 2006, p. 133) e o que se vê é uma tentativa sofisticada de compreender Schmitt a partir de Hegel. De maneira precisa, Kervégan refaz a argumentação schmittiana acerca do liberalismo, da *Rechtsstaatlichkeit*, das críticas ao legalismo e ao parlamentarismo, traz ao debate as ambiguidades do realismo político de Hegel, reabsorvido por Schmitt, por exemplo, quando trata do direito internacional e elabora um paralelismo entre

¹ Departamento de Filosofia da Universidade Federal do Piauí (UFPI)

² Departamento de Filosofia da Universidade Federal Fluminense (UFF)

Grundlinien der Philosophie des Rechts e o *Der Begriff des Politischen*, bem como desenvolve a recusa à submissão da política à moral abstrata, entre outros temas. Na leitura de Kervégan, Schmitt, apesar da inspiração marcadamente hegeliana, assume as consequências pós-hegelianas da dialética especulativa, permanecendo na positividade, ou melhor, como Kervégan afirma, numa metafísica da decisão. Dessa forma, a leitura de Kervégan é exposta em três momentos principais: (a) o percurso do decisionismo ao Estado total, (b) a interpretação de Schmitt acerca da relação entre Estado e sociedade civil em Hegel e (c) a recusa à solução especulativa que resolveria as contradições do político.

I

O argumento central que Kervégan assume é acerca da distinção entre o político e a política, referindo-se à anterioridade do político sobre o Estado. Enquanto o político se expressa como relação, ou melhor, o antagonismo ou o conflito entre forças que perpassa o corpo social; a política seria a mediação entre decisão e forma de direito, uma espécie de instituição subsidiária ao problema da representação, por exemplo, em *Römischer Katholizismus* (SCHMITT, 2008). Mesmo assim, Schmitt não leva às últimas consequências a diferença entre o político e a política, pois se o político vai além do estatal, a escolha pela ordem e pelo Estado não seria uma consequência necessária de sua teoria, mas apenas uma preferência diante do apego ao modelo westfaliano de Estado (uma espécie de "monopólio natural", ou melhor, histórico pelo Estado). Diante disso, Kervégan relaciona a reflexão sobre o político à proposta do Estado Total e, como o próprio Schmitt deixa claro no segundo pós-guerra, sustenta que era do estado já não existiria mais. Nesse contexto, Kervégan analisa as consequências da categoria do político sobre o Estado, ou seja, o Estado total como "expansão do político e decadência do Estado liberal clássico" (KERVÉGAN, 2006, p. 46), demonstrando os elementos hegelianos que Schmitt recolhe.

Ao considerar atentamente para a diferença entre o político e a política, o autor compreende bem a natureza do problema em questão: "o elemento mais notável desse texto é a ideia de que o político não é uma substância ou um domínio de objetos, mas uma relação, uma função, seria até uma tentação dizer: um modo"

(KERVÉGAN, 2006, p. 55). A análise descreve como Schmitt (2002) distingue no *Der Begriff des Politischen* entre os conceitos de Estado e de político, além de atribuir o equívoco de pensar o político como estritamente estatal, evidentemente, levando em conta o monopólio político pelo Estado na modernidade. Após esclarecer este erro persistente, o autor ressalta a leitura schmittiana de que o político se refere a qualquer atividade humana, pois um horizonte de conflito como relação originária. Por isso, Kervégan considera o "caráter relativo e relacional do político" como 2 (duas) teses distintas: (i) a indeterminação substancial da relação política e a não-impossibilidade do conflito que implica a polemicidade na natureza humana (neste ponto, o autor esboça a crítica de que o critério do político seria meramente descritivo, ou seja, seria uma antropologia política); e (ii) o conflito como o "coeficiente político da atividade humana" (KERVÉGAN, 2006, p. 57), que caracteriza o político como algo fora do Estado, determinado a partir da guerra como situação-limite, ou melhor, o horizonte de sentido a partir do qual se determina a política. Esta pressuposição sempre presente, porém, não é distinta da decisão do soberano ou da exceção, tratando sob o mesmo rótulo as diferenças entre exceção e excesso – próprias do período weimariano, sobretudo, no *Politische Theologie* (2004) – e entre o político e a política – referentes ao período tardo-weimariano, que apesar de heterogêneo pode ser delimitado pelo *Der Begriff des Politischen*. Quanto ao conceito do político, Kervégan compreende bem o que está em jogo ao afirmar que:

A dissolução, ou pelo menos a interferência dessa divisão clássica, que Schmitt pensa perceber na realidade contemporânea, torna, em contrapartida, problemática a identidade própria do político, ou a sua vocação para se manifestar de forma privilegiada mais em certos domínios de atividade do que em outros. Pareceria, portanto, que o político como grau polêmico seria, ao mesmo tempo, o que motiva a distinção entre Estado e político, preocupação maior de Schmitt, e o que ameaça a própria identidade do político, que o Estado tradicionalmente assegurava (KERVÉGAN, 2006, p. 58).

Em vez de interpretar o rompimento como favorável para uma teoria política pós-estatal, Kervégan – assim como o próprio Schmitt nesta fase da sua obra –

74 | Reflexões sobre alguns problemas filosóficos

compreende que a crise de identificação do político com o Estado provoca, ao contrário, a figura do Estado total e, por conseguinte, a guerra total, o pacifismo e a criminalização da oposição ao considerar o inimigo como exterminável, pois:

a guerra se inscreve numa lógica que não é mais aquela da grande política clássica dos poderes, mas que leva a declarar o inimigo fora da humanidade e a justificar seu extermínio (*Vernichtung*). As noções de guerra total e de inimigo total, forjadas em 1935 para caracterizar esse novo modo de ser suprapolítico da guerra, formam com aquela do Estado total os elementos de uma teoria da auto-superação em curso da guerra e do Estado e, por conseguinte, do universo clássico da política (KERVÉGAN, 2006, p. 59).

Neste contexto, Kervégan considera como o Estado Total seria a consequência da extensão e da intensificação do político, mais especificamente, como a “guinada ao Estado Total” se refere à “nova relação existente entre o Estado e os recursos da técnica moderna: para Schmitt, o Estado Total não é apenas um Estado que utiliza as técnicas para estabelecer e desenvolver sua influência sobre cada um; é também, simplesmente, o Estado da ‘era da técnica’” (KERVÉGAN, 2006, p. 68). No entanto, não apenas em relação à expressão da técnica, mas também como consequência da ruptura da separação entre sociedade e Estado, próprias das delimitações do pensamento liberal, Kervégan considera que Schmitt fora “um dos primeiros a perceber o alcance das transformações que afetam os Estados liberais-democráticos contemporâneos” (KERVÉGAN, 2006, p. 70) e faz um comentário que aproxima Schmitt de uma espécie de pré-compreensão da biopolítica:

Elas [as transformações que Schmitt analisa] não significam apenas a atribuição de novas funções ao organismo político, a par de suas obrigações tradicionais: política exterior, manutenção da ordem e justiça. A importância adquirida pela administração da vida social é traduzida por uma verdadeira mudança de natureza. A soberania do Estado (mesmo absolutista) andava junto com o reconhecimento da existência de questões não políticas. Ao contrário, o Estado social, que nisso é um Estado “total”, intervém em todas as circunstâncias possíveis e em todos os domínios da existência humana, não apenas na

economia [...], mas também nas questões culturais e sociais (KERVÉGAN, 2006, p. 70).

Ao vincular a tese da distinção entre amigo-inimigo ao Estado total autêntico, Kervégan contradiz a identidade entre Estado e política logo em seguida quando afirma que “para Schmitt, o Estado enquanto conceito de uma realidade histórica determinada está em segundo lugar em relação ao político, que designa a intensidade polêmica da relação inter-humana até seu fundamento antropológico” (KERVÉGAN, 2006, p. 73).

A reflexão sobre o Estado Total é finalizada por Kervégan com o pressuposto metafísico da teologia política. Ao lançar a interpretação da sociologia dos conceitos jurídicos como a “explicitação de tais correlações entre mentalidades e estruturas políticas”, para o autor, “a tarefa de tal sociologia é, pois, elucidar numa perspectiva histórica, o núcleo metafísico (e teológico) que as representações do Estado abrigam. A teologia política desenvolve-se assim numa antropologia cultural” (KERVÉGAN, 2006, p. 88). Kervégan interpreta como uma teoria da história as representações ou imagens que uma época elabora de si que se organizam em torno de um setor dominante (*Zentralgebiet*): objeto das representações e princípio de inteligibilidade de uma época. Em cada época da modernidade, Schmitt descreve uma imagem de mundo (*Weltbild*) que seria um campo de objetividade e representações a partir do qual há a organização da época, da subjetividade, das formas sociais. Assim, as esferas teriam se sucedido da teologia, metafísica, moral, economia até, finalmente, a técnica. Kervégan trata da historicização das formas de subjetividade, tema próximo a Dilthey e Spengler, dos quais Schmitt, afinal, se mostra devedor, mas também enfatiza a relação com Heidegger, quando este trata da “era das concepções de mundo”, que seria um prolongamento “filosófico da reflexão político-histórica de Schmitt sobre a ‘era das neutralizações e despolitizações’”; ambas analisam a concepção moderna do mundo como sistema. É significativo que tanto em Schmitt quanto em Heidegger a reflexão crítica sobre a modernidade, que insiste em sua dimensão política ou metafísica, tenha por desfecho o problema da técnica (KERVÉGAN, 2006, p. 90).

No entanto, não apenas a reflexão sobre a técnica marca a interpretação de Kervégan, mas também a peculiar filosofia da história schmittiana. O sentido da história moderna seria compreendido a partir da representação do político e das imagens do mundo, do *Zentralgebiet* como um processo de neutralização sucessivo e, por conseguinte, de despolitizações: "A História é história do deslocamento do lugar do político, ao mesmo tempo, que é das formas de ser e de agir do Estado. Quando um novo setor dominante advém, o desafio dos conflitos possíveis se desloca, e as formas mesmo do político se transformam profundamente" (KERVÉGAN, 2006, p. 92). Assim, ao trazer com clareza o tema e não ceder às tentações de associá-lo ao totalitarismo ou nazismo, o autor define o Estado total, mesmo em suas ambiguidades, como a tese schmittiana que "oferece a solução para os problemas que são apresentados pela determinação do político a partir da técnica. Somente tal Estado, 'que não conhece mais nada de absolutamente não político', é passível de investir e de dominar a esfera da técnica" (KERVÉGAN, 2006, p. 95), mesmo que nos pareça exagerado afirmar que o Estado Total "manifesta o enraizamento da *teoria jurídica positiva* e de um *discurso político* que aspira à positividade numa *metafísica da história*, entendida como teologia política. O Estado Total para Schmitt é a verdade atual do político, e ele o é porque conduz o político a se atualizar plenamente, isto é, totalmente" (KERVÉGAN, 2006, p. 95).

Kervégan ressalta a importância da tese schmittiana de contiguidade entre Estado total e democracia, ao afirmar que "fazer proceder ao 'totalitarismo' do desenvolvimento da democracia é, na época, algo original" (KERVÉGAN, 2006, p. 78-79), pista seguida à risca por Giorgio Agamben (2004). O intérprete franco-argelino percebe que nos textos pós-1945 a suspeita – e depois a convicção do jurista – surgem de uma transposição definitiva da estrutura do Estado como tal. A dissociação do político e do estatal, segundo as teses de Schmitt, adquiriria um teor de sentido historicamente concreto. Nada é mais nítido que o prefácio da reedição de *O conceito do político*, de 1963, segundo a qual "A era do Estado está em declínio. Todo comentário é doravante supérfluo" (KERVÉGAN, 2006, p. 63), poderia facilmente ter antecipado este diagnóstico já para o início da década de 1930 ou, pelo menos, nos textos que culminaram no *Der Nomos der Erde*. Assim, o intérprete elege

o tema do Estado total como aquilo que substituiria o estado liberal-constitucional parlamentar e conduz o conceito do político ao Estado total.

Se o Estado moderno se dissolve por conta do contexto geopolítico do século XX, bem como por conta da interferência entre as esferas antes segregadas do público e do privado, sociedade civil e Estado, a rigor, isso não significa o declínio do político, mas apenas sua reconfiguração em termos de Estado total.

II

O segundo aspecto da interpretação sobre Schmitt é a influência da chave de leitura hegeliana. Kervégan descreve a lógica de Hegel em três aspectos, ou melhor, discerne três momentos em tudo que possui uma realidade: (a) "o momento abstrato ou de entendimento", (b) "o momento dialético ou negatividade racional" e (c) "o momento especulativo ou positivamente racional" (KERVÉGAN, 2006, p. 352). A filosofia revelaria a partir da positividade abstrata uma negatividade que movimenta e impulsiona esta positividade, isto é, à positividade racional e especulativa do conceito. Para Hegel, segundo Kervégan, "caso seja necessário à razão olhar o negativo de frente, de persistir junto a ele [...] é sempre prejudicial tomá-lo como ponto de partida como faz o entendimento abstrato" (KERVÉGAN, 2006, p. xxi). Com efeito, ele critica a positividade que permanece fechada em si mesma. Este fechamento significa, na verdade, a recusa da oposição do imediato e da mediação, ou antes, a demonstração disso que essa oposição, ela mesma imediata é subentendida por todo o processo da mediação lógica cujo resultado seria a vontade de reconciliação. Na leitura de Kervégan, Schmitt teria compreendido esta mediação ou reconciliação em Hegel como uma concessão inadmissível ao liberalismo e à ética da discussão. Assim, Kervégan sustenta que:

Se, para Hegel, a filosofia é essencialmente filosofia do real, é evidente que ela não é uma consagração da faticidade: seus textos políticos aqueles mesmo cujo realismo Schmitt admira tanto o demonstram suficientemente. A reconciliação com a efetividade [...] não significa que essa torne toda realidade amável, mas que ensina a detectar aí os traços de um movimento cuja razão, se ela deve ser percebida e decifrada na História, não pode ser pensada somente a partir dela (KERVÉGAN, 2006, p. 356).

Além disso, a derradeira formulação da filosofia política hegeliana é rejeitada por Schmitt, pois Hegel teria inserido sua “teoria política numa doutrina do espírito objetivo, que se desenvolve como filosofia da história universal” e, por conseguinte, passaria a tratar a racionalidade histórica em ato de efetuação. No comentário preciso de Kervégan,

Hegel vê em Napoleão o instrumento da razão histórica; Carl Schmitt prefere limitar-se à materialidade do instrumento. Nada de impressionante, aliás, nessa opinião preconcebida: a filosofia da História e seu idealismo da razão são aquilo que Schmitt julga mais ambíguo em Hegel, em todo caso, aquilo que debilita, a seu ver, o vigor ‘político’ que manifestam os textos de Frankfurt ou Iena. (KERVÉGAN, 2006, p. 172).

Ao propor o Estado como uma instância da universalidade racional na ordem do espírito objetivo, Hegel afasta a compreensão do político como confronto e conflito, não obstante a força e a violência estejam presentes no seu conceito. A guerra (ou em termos schmittianos, o político como relação e antagonismo) seria para o Hegel mais maduro (aparentemente, tão distinto do Hegel da *Constituição da Alemanha* ou do artigo sobre direito natural):

a mediação, o momento negativo graças ao qual a identidade positiva da comunidade política, ela própria socialmente diferenciada e organizada, se constitui e se mantém nas coisas e nos espíritos [...] a guerra e sua eventualidade não podem ser uma determinação positiva da essência do Estado. O teor ético da guerra não é, decididamente, aquilo que permite determinar o conceito do político (KERVÉGAN, 2006, p. 175-176).

A função do Estado é a unificação, o viver em conjunto e não a guerra: “arriscando a vida e os bens de cada um [...] ela lembra factualmente a verdade filosófica segundo a qual o político não é a condição externa da vida ética, mas, antes, seu lugar próprio originário” (KERVÉGAN, 2006, p. 176). Assim, ao propor a guerra ou conflito como horizonte sempre presente da política – o elemento propriamente

político da política – Schmitt retomaria de Hegel e da ética do Estado, como Kervégan sustenta, “a continuidade entre guerra e política [que] se deve ao fato de a guerra, ou pelo menos a sua eventualidade, ser a condição da existência factual do Estado” (KERVÉGAN, 2006, p. 176). Entretanto, em Hegel há uma diferença fundamental: a guerra é uma negação da negação (das determinações finitas que se tomam por objetivos incondicionais, vida burguesa, propriedade, bens, etc.), mas aquilo que é negado pelo político (particularidade social) também é mediação daquilo que nega.

Neste contexto da leitura hegeliana, conforme Kervégan (2006, p. 352), Schmitt deveria ser compreendido, por um lado, através de uma “absolutização da positividade” – no sentido de que conserva a ordem (do Estado) acima de qualquer contradição (negatividade) interna –, por outro lado, como uma “absolutização do próprio negativo” – no sentido de que ao considerar a distinção amigo-inimigo não teria uma solução especulativa que superasse a oposição e a elevasse à positividade especulativa. Assim, para o autor, caso assumisse toda a lógica hegeliana, o jurista não teria “a força de impacto polêmica da antítese dualista’, pois a dialética oculta ou apaga [...] o ‘ou ... ou’ da decisão” (KERVÉGAN, 2006, p. 353). O autor avalia que Schmitt não assume esta consequência do hegelianismo e, por isso, “permanece num ponto de vista que se baseia no fato de deter-se no finito” (KERVÉGAN, 2006, p. 353). Neste ponto, a interpretação de Kervégan afirma que Schmitt assume a finitude, em vez do momento especulativo ou positivamente racional. É verdade que sob a influência de Hegel, como advogado da mediação, o decisionismo provocava um fechamento e, por conseguinte, a constituição da ordem implicava num conceito do político como mediação e a decisão pela representação. Entretanto, em Schmitt a negatividade não significa fechamento, pois há abertura, o outro, impossibilitando o fechamento da decisão, visto que a ordem pressupõe a relação, o conflito.

Nesse sentido, uma questão que nos parece ambígua na leitura de Kervégan é a associação entre decisionismo e emancipação da tutela filosófica. Apesar de tornar relativa tal ruptura ao reconhecer uma associação solidária entre racionalidade e entendimento, na questão da negatividade Kervégan, o autor elege a teoria decisionista como “a expressão mais consciente e radical da emancipação da tutela filosófica, mesmo que possa ser identificada como uma forma de racionalidade: de maneira paradoxal – mas este paradoxo a torna filosoficamente

interessante – ela é uma *metafísica da positividade*" (KERVÉGAN, 2006, p. xxxi). A partir desta emancipação do saber especulativo, Kervégan compreende as teses de Schmitt e o aproxima de um autêntico positivismo jurídico, pois "se o decisionismo é um pensamento 'positivo' e mesmo, a nosso ver, a sua forma acabada, é então em um sentido totalmente diferente do positivismo jurídico: ele pretende se situar não fora, mas além da interrogação que engendra a racionalidade filosófica" (KERVÉGAN, 2006, p. xxxi-xxxii). Dessa forma, Schmitt seria a expressão mais radical da emancipação do saber especulativo, nos termos do autor, de uma metafísica da positividade. Entretanto, o trecho mais evidente da interpretação que demonstra o interesse de Kervégan na ruptura provocada por Schmitt é quando afirma: "Enquanto uma filosofia como a de Hegel se esforça para acolher o entendimento positivo, para conduzi-lo para além de si mesmo e revelar a sua subordinação às exigências de ordem especulativa, o decisionismo recusa o princípio de uma racionalidade que poderia constituir o horizonte de sentido da teoria jurídica" (KERVÉGAN, 2006, p. xxxii). Ele prossegue:

A alternativa verdadeiramente positiva de uma inteligência especulativa do real [...] consiste em liberar de seus laços filosóficos o espaço da racionalidade positiva, fazendo da decisão, componente essencialmente irracional de toda ordem jurídica e política, o pré-requisito impensável do próprio direito (KERVÉGAN, 2006, p. xxxii).

A afirmação de Kervégan – segundo a qual Schmitt "exclui a perspectiva racional de reconciliação" (KERVÉGAN, 2006, p. xxxii) – sustenta que este "positivismo" não teria nada em comum com aquele em voga no início do século, pois este verdadeiro positivismo se refere à emancipação da tutela filosófica, instaurando aquilo que o autor denomina como *metafísica da positividade*. Kervégan esboça uma compreensão neste sentido e, além de elaborar uma crítica contundente à tradição política racionalista ou normativista, interpreta o conceito do político schmittiano como:

uma espécie de niilismo epistemológico refletido, radicaliza o projeto que o positivismo jurídico formulava ainda ingenuamente: o de uma razão jurídica *Wertfrei*, libertada dos valores, que não fosse mais uma razão, já que teria deixado de ser normativa. Ele consegue fazendo do 'político', ele próprio incluído, a partir da situação-limite que determina o teor, isto é, o conflito, a pressuposição irreconciliável de toda ordem jurídica concreta e de todo pensamento dessa ordem (KERVÉGAN, 2006, p. xxxii).

Para Kervégan, decisionismo seria, portanto, uma "*metafísica da decisão*, porquanto esta constitui o momento fundador, 'o político', de uma ordem qualquer" (KERVÉGAN, 2006, p. 351) – e como tal possuiria uma base metafísica expressa na teologia política e filosofia da história própria desta. Kervégan segue de perto a crítica de Löwith (1984) e de Strauss (2007): com este, afirma que Schmitt procura uma inversão sistemática do liberalismo e termina preso no horizonte daquilo que combate, configurando um liberalismo às avessas; com aquele, por não determinar o "núcleo metafísico do liberalismo" sofre com a indeterminação do próprio fundamento do pensamento que combate. Assim, para Kervégan, embora tenha reconhecido as objeções de Strauss e Löwith, Schmitt não assume, tal como gostaria, o argumento segundo o qual o político "supõe uma determinação da sua essência" (KERVÉGAN 2006, p. 119), rejeitando a determinação essencialista ou normativista. Está em jogo também uma espécie de niilismo em Schmitt: por este argumento, o autor consegue estabelecer um critério forte para opor-se as críticas de Strauss e Löwith (ausência de fundamento), pois Schmitt prescinde desta base (afinal, Löwith tem razão em seu diagnóstico, mas parece se apressar ao determinar que esta ausência é uma falha) e tanto a decisão quanto o político são irreduzíveis à racionalidade normativa.

A questão de Löwith contra o niilismo de Schmitt é a acusação de ocasionalismo, na verdade, oportunismo ou inviabilidade de um fundamento para a ação política. Da mesma forma, a crítica antinormativa ao liberalismo não fornece uma base para a ação política; ora, segundo a crítica de Strauss, ela mesma serviria como base normativa para o jurista. Assim, Schmitt ou bem descarta toda e qualquer normatividade (acusação de ocasionalismo de Löwith) ou bem assume, às avessas,

a normatividade do liberalismo (acusação de liberal e moralista de Strauss). No entanto, nem Strauss nem Löwith. Interpretamos a tentativa de Schmitt como superação da lógica do fundacionismo que escapa da “má infinidade (do fundamento) do normativismo” (KERVÉGAN, 2006, p. 120), conforme Kervégan sustenta.

Kervégan encerra a parte destinada a Schmitt com a tese de que se nem com Hobbes de Strauss, nem com o Heidegger de Löwith é possível realizar uma leitura adequada de Schmitt, seria Hegel a chave de leitura mais eficaz. Esta seria a cifra que daria acesso ao pensamento do jurista. A tese de Kervégan, portanto, afirma que Schmitt recorreu a Hegel para sua empreitada de “inversão da epistemologia liberal”, apesar de Schmitt se situar além das alternativas que ele mesmo considera inelutáveis (KERVÉGAN, 2006, p. 120). O autor acredita que Schmitt postula uma metafísica da história através da sua teologia política e o Estado total seria uma resposta aos críticos, seria “a verdade atual do político, e ele o é porque conduz o político a se atualizar plenamente, isto é, totalmente” (KERVÉGAN, 2006, p. 95), no Estado.

III

Por mais razoável que seja a interpretação de Kervégan acerca da ambiguidade de Hegel e da escolha de Schmitt pela positividade³, o problema é que a obra de Schmitt não se resume ao decisionismo. Kervégan está correto quando sustenta que Schmitt dissocia um aspecto conservador de um aspecto revolucionário em Hegel, uma vez que o jurista “subordina a problemática especulativa a uma lógica, a do ‘ou... ou’, da qual essa problemática quer ser

³ “Se Carl Schmitt endurece a oposição entre ordem política e desordem social, se insiste na semelhança estrutural entre a sociedade civil, no sentido de Hegel, e o estado de natureza hobbesiano, não é unicamente para impor a sua interpretação ‘política’ do hegelianismo em oposição às leituras ‘liberais’ dominantes. Para ele, trata-se, acima de tudo, de promover, através de Hegel, a sua própria ética decisionista do Estado e de justificar a consequência última principal que lhe parece que deve ser extraída daí: diante de uma sociedade que se tornou total, isto é, tendo imposto a sua medida no espaço do político, um Estado, ele próprio total, é a única chance de restabelecer a dimensão do universal, de preservar a condição absoluta de toda ética e de todo direito” (KERVÉGAN, 2006, p. 235). Kervégan pretende com a pesquisa mostrar o erro de Schmitt (endurece a oposição entre desordem social e ordem política) e a aporia do seu decisionismo (2006, p. 235) e, através disso, restabelecer a teoria política hegeliana. A questão estaria no ponto em que há “limites da assimilação da sociedade civil num estado de natureza puro; esta possui, ao menos, a *forma* de uma racionalidade, apesar de somente o Estado dotar essa forma de um conteúdo ético” (2006, p. 235).

justamente a rejeição [...] No caso de Schmitt, ela [a lógica da decisão] se revela particularmente operante, já que tende a fortalecer a abordagem do direito e da política que é a sua" (KERVÉGAN, 2006, p. 144). Assim, Schmitt faz da dialética o fator revolucionário do hegelianismo e considera o marxismo como herdeiro desse pensamento, mas a questão fundamental para o tema aqui tratado é que "a decisão revolucionária é a consequência rigorosa da dimensão dialética do pensamento de Hegel, ou ainda da negatividade constitutiva do racional" (KERVÉGAN, 2006, p. 144), que no pensamento sobre o político (não decisionista) de Schmitt ocorre através da dupla inscrição entre *die Politik* e *das Politischen*, em tonalidade não muito estranha à crítica marxista: o argumento do antagonismo e da concretude (finitude). Em outras palavras, a partir do contexto jurídico-político da República de Weimar, apesar de assimilar conceitos e teses hegelianas, Schmitt não admitiria a solução especulativa do sistema de Hegel, permanecendo na positividade sem a possibilidade de síntese racional. No entanto, além da recusa do campo de racionalidade definido por Hegel, Schmitt ainda apontaria uma ambiguidade no filósofo: uma tendência à afirmação do poder do Estado conviveria com a orientação liberal. Esta indecisão seria mais um motivo a partir do qual Schmitt acusaria as ilusões normativas persistentes e localiza a tensão do hegelianismo precisamente na teoria política: entre o elemento dialético e o elemento sistemático, ou seja, "o projeto íntimo de sua filosofia reconciliar – isto é, unificar na forma de uma racionalidade que faça jus ao livre movimento da diferença – a vitalidade infinita da negatividade dialética e o poder organizador da totalização sistemática" (KERVÉGAN, 2006, p. 139). Diante do que Löwith denomina de "a ambiguidade fundamental dos *Aufhebungen* dialéticos"⁴ (LÖWITH, 1969, p. 84), Schmitt tem em vista, assim como os marxistas, os conflitos que ocorrem na sociedade civil – desta análise, são extraídas as possibilidades revolucionárias do pensamento hegeliano –, bem como a doutrina do Estado e suas consequências conservadoras, até mesmo reacionárias. Conforme Kervégan (2006, p. 143), "o marxismo rompeu com aquilo que a subordinação da sociedade civil ao Estado [...] podia comportar de equívoco (do ponto de vista da lógica decisionista do 'ou... ou')".

⁴ No trecho completo, "Die Spaltung der Hegelschen Schule in Rechts- und Linkshegelianer war sachlich ermöglicht durch die grundsätzliche Zweideutigkeit von Hegels dialektischen »Aufhebungen«, die ebensogut konservativ wie revolutionär ausgelegt werden konnten" (LÖWITH, 1969, p. 84).

Neste contexto, corroborando a leitura de Kervégan, os trabalhos de Schmitt despertam, inesperadamente, um interesse em pensadores marxistas como G. Lukács, W. Benjamin e K. Korsch⁵. Com efeito, Schmitt elabora um paralelo entre estratificação social e a organização da sociedade burguesa e o Estado com a relação de Hobbes entre estado de natureza e Estado civil, mais precisamente, analisa os parágrafos 243 a 245 da *Grundlinien des Rechtsphilosophie*, onde se mostra o esboço de uma teoria da luta de classe. Nas palavras de Kervégan (2006, p. 2002), a diferenciação e diversificação própria da sociedade civil é o traço da particularidade natural que conserva em si o resquício do estado de natureza. Isso coincide com o argumento de Schmitt, bem como com a leitura de Strauss: como uma espécie de luta pelo reconhecimento, porém no interior do estado civil, assumindo ou levando a sério o paradoxo de que o direito advém da violência, ou melhor, para extinguir a violência com violência. Ora, daí as consequências do "direito dos heróis de fundar Estados" como portador de uma violência legítima. Diante disso, Kervégan faz um precisa analogia com Schmitt ao afirmar que "esse direito, que não tem nada de jurídico, ilustra, no ponto em que o estado de natureza oscila em sua negação política, uma dialética da natureza e do espírito em movimento na direção de sua objetividade ao mesmo tempo jurídica e política" (KERVÉGAN, 2006, p. 218), ou seja, como um limite entre a anomia da natureza e a ordem de direito. Kervégan prossegue, associando diretamente o tema à Schmitt:

nesse sentido, ele poderia servir para uma interpretação do tipo decisionista: da mesma forma que, em tal perspectiva, o estado de exceção é o fato extranormativo que torna efetiva a ordem normativa, assim como, no mito político das origens ao qual recorre Hegel, o ato fundador é o ponto de inflexão onde a ordem se instaura pelo próprio meio que produzia o caos (...) mas percebe-se a falha de tal interpretação: ao passo que, numa perspectiva decisionista, a exceção (a 'decisão') é a condição de validade da ordem normativa, a violência originária não é, conforme Hegel, de modo algum, princípio da ordem que ela produz factualmente. Sendo somente a negação do não-direito, o direito do herói

⁵ A breve carta de Benjamin à Schmitt consta na *Gesammelte Schriften*, Frankfurt: Suhrkamp, I/3, p. 887.

continua essencialmente do lado do não-direito e da naturalidade (KERVÉGAN, 2006, p. 218-219)

Assim como em Hegel, Schmitt aposta na força da negatividade, isto é, na oposição e cisão, mas diferentemente dele, considera esta força como originária e não apenas um começo fenomenológico dos Estados, seu início externo, não conceitual: violência e negação de direito é passagem, formam a negatividade graças a qual o Estado tem espaço. Portanto, é a negação da negação, uma reconciliação do espírito consigo mesmo. Entretanto, Hegel, diferentemente de Schmitt, não concebe o Estado a partir da situação limite do estado de natureza; pois distingue entre origem e fundamento, ser e aparecer, "começo no fenômeno" e o "princípio substancial" dos Estados. Se em Hegel, o processo da contradição e diferença (negatividade – negação da negação) leva ao fundamento, desconstruindo os dualismos (identidade e diferença, interior e exterior, sujeito e objeto) por encontrar, ao final, algo unificador; em Schmitt, não há resolução da contradição através do decisionismo, tal como Kervégan afirma: "Se Schmitt deixa de percebê-lo (a unidade racionalidade ético-política do Estado) é provavelmente menos por incompreensão ou ignorância dos textos do que porque precisaria, para reconhecê-lo, aceitar que fosse questionado o dualismo filosófico intransigente no qual se baseia a sua própria empreitada" (KERVÉGAN, 2006, p. 40). Para Kervégan, a ambivalência do pensamento de Hegel permite "o acesso ao núcleo teórico do decisionismo" (KERVÉGAN, 2006, p. 150).

Em última instância, sob este aspecto, Schmitt não seria um hegeliano nem de esquerda nem de direita, mas um anti-hegeliano, graças à Hegel, afinal, não se pode escapar dele tão facilmente: nem tanto por sua rejeição à superação da contradição, mas sim por sua compreensão da negatividade que o político implica, pois como relação e conflito originário e contínuo mesmo dentro da ordem que sempre destitui o positivo (seja o abstrato seja o especulativo/racional).

IV

A inimizade em Hegel não se refere a uma naturalização da relação, mas sim ao caráter ético e metafísico, sobretudo, quanto à soberania externa, incapaz de se reduzir à positividade abstrata. Segundo Kervégan, "o inimigo é, em sua determinação especulativa, o aspecto exteriorizado da *negatividade* constitutiva da

identidade positiva de si da vida ética" (KERVÉGAN, 2006, p. 157, grifos do autor). Inicialmente, ele trata a soberania do Estado em Hegel como uma relação de exclusão de outros Estados. Evidentemente, sem a possibilidade de um terceiro superior (a não ser a História, mas isso é obviamente problemático), os Estados teriam a forma de relação externa via guerra, ou seja, através do *jus belli*, do uso da força legítima no caso de solucionar suas pendências. Este tipo de Estado, para Schmitt, já não existe mais. Entretanto, a chave para compreender o trecho citado de Kervégan é que Schmitt assume a noção do inimigo como negativo exterior constitutivo. Na leitura de Kervégan, Schmitt assume este elemento de Hegel, mas destituída do momento especulativo e, por isso, ele teria se fixado na positividade da mera decisão entre eles ou nós, uma vez que já teria se defeito do vínculo entre político e Estado e sustentaria que "essa exterioridade mútua [...] é apenas a transposição em exterioridade da 'relação infinitamente negativa de si', que é engajada pelo ser para si positivo da totalidade política" (KERVÉGAN, 2006, p. 157). E prossegue: "É preciso, por conseguinte, pensar o negativo (o inimigo) como negatividade, isto é, não como uma realidade dada, mas como momento do processo da totalidade ético-política" (KERVÉGAN, 2006, p. 158); mais precisamente, é como se o negativo fosse um termo externo, apesar de recusar aquela "totalidade ético-política". Para Schmitt, o político e o direito comportam um caráter conflitual (antagonismo) e Kervégan reitera este aspecto quando afirma que "o político não é outra coisa senão uma relação originária da eventualidade do conflito" (KERVÉGAN, 2006, p. 350). Assim como Hegel trata do "ser para si como uma relação infinita consigo mesmo mediatizada pela relação com o outro", Schmitt também trata na relação de conflito a exposição daquilo que é constitutivo da identidade de um Estado. A releitura que elaboramos prescinde da hipótese de ordem ou do Estado como pressuposto e busca recuperar a noção de negatividade e guerra:

a equação estabelecida [por Hegel] pelo artigo sobre o direito natural entre a atividade militar e a atividade política, faces indissociáveis do modo de ser do 'estado dos homens livres', poderia ser compreendida como uma antecipação das teses de *Le concept du politique*. Numa perspectiva decisionista, a guerra, ou, mais precisamente, a hostilidade, é a situação limite que permite pensar a

possibilidade da unidade política, e o *jus belli*, isto é, 'a possibilidade real de designar o inimigo, se for preciso, por uma decisão própria', é a propriedade distintiva do Estado soberano (KERVÉGAN, 2006, p. 169)

Apesar da proximidade quanto à noção de guerra e de inimigo, neste ponto, Kervégan traz uma contribuição importante na relação entre Hegel e Schmitt: o jurista não concorda com o filósofo, pois "ela [a obra de Hegel da época de Berlim] inscreve os princípios do *jus publicum europaeum* na perspectiva puramente especulativa de um racionalidade histórica" (KERVÉGAN, 2006, p. 165) e, precisamente, seria esta capitulação diante do conflito que Schmitt parece recusar, de maneiras distintas, seja pela via do decisionismo, seja pela via do político como relação e antagonismo. Assim, nem a decisão que Kervégan aposta, nem a solução da totalidade especulativa: para Schmitt, o conceito do político, revisitado no final da década de 1920, seria o esboço de uma dialética negativa⁶. Apesar disso, Kervégan tenta associar Schmitt à Hegel quando ele "confirma a ligação íntima que une a metafísica decisionista da positividade à dialética especulativa, e isso a respeito do tema principal do ensaio sobre o conceito de política, a relação amigo-inimigo [...] 'o inimigo é a figura de nossa própria questão'" (KERVÉGAN, 2006, p. 358). Teria Schmitt assumido o inimigo como mediação? Isso nos parece bastante heterodoxo. Em todo caso, na leitura de Kervégan, Schmitt rejeita a solução da negatividade e assume a positividade racional ou especulativa ao escolher a positividade e não a dialética, além de fazer dela seu *ethos*.

Referências

AGAMBEN, G. **Homo Sacer**. O poder soberano e a vida nua. Trad. Henrique Burigo. Belo Horizonte: Editora UFMG, 2004.

BENJAMIN, W. **Gesammelte Schriften I**. Rolf Tiedmann; Hermann Schweppenhäuser (Hrsg.). Frankfurt: Surhkamp, 1991.

⁶ "No *ser-aí*, essa relação *negativa* do Estado consigo aparece como relação de um *outro* a um *outro*, e como se o negativo fosse um [termo] *externo*. A existência dessa relação negativa tem, portanto, o aspecto de um devir e de uma confusão com dados contingentes que vêm *do exterior*. Mas [essa relação] é seu momento *próprio* mais elevado, sua infinidade efetiva enquanto idealidade de tudo aquilo que está incluído nele" (*apud* KERVÉGAN, 2006, p. 158 HEGEL, RPh §323, p. 279).

KERVÉGAN, J-F. **Hegel, Carl Schmitt**: o político entre a especulação e a positividade. Tradução de Carolina Huang. Barueri: Manole, 2006.

SCHMITT, C. **Der Begriff des Politischen** (1932). Text von 1932 mit einen Vorwort und drei Corollarien. 6. Aufl. 5. Nachdruck der Ausgabe von 1963. Berlin: Duncker & Humblot, 2002.

_____. **Römischer Katholizismus und politische Form**. 5^a ed., Stuttgart: Klett-Cotta, 2008.

_____. **Politische Theologie**. Vier Kapitel zur Lehre von der Souveränität. 8^a ed., Berlin: Duncker & Humblot, 2004.

LÖWITH, K. "Der okkasionelle Dezisionismus von Carl Schmitt" (1935), in: **Sämtliche Schriften**, Band 8, Stuttgart: Metzler, 1984, p. 32-71.

_____. **Von Hegel zu Nietzsche**. Der revolutionäre Bruch im Denken des neunzehnten Jahrhunderts. Frankfurt: Fischer, 1969

STRAUSS, L. "Notes on Carl Schmitt, The Concept of the Political". In: SCHMITT, Carl. **The Concept of the Political**. Tradução de Matthias Konzen e John McCormick. Chicago: University of Chicago Press, 2007, p. 97-122.

VOEGELIN, E. "Die Verfassungslehre von Carl Schmitt". In: VOEGELIN, E. **The Collected Works of Eric Voegelin**. Vol. 13: Select Book Reviews. Jodi Cockril; Barry Cooper (Org.) Missouri: University of Missouri Press, 2001, pp. 42-66.

6. SOBRE A POSSIBILIDADE DE UMA CONSCIÊNCIA ARTIFICIALMENTE MORAL



<https://doi.org/10.36592/9786554601061-06>

Dayvide Magalhães de Oliveira¹

Resumo

O objetivo deste artigo é apresentar uma resposta aceitável para a pergunta sobre a possibilidade de atribuir consciência moral para máquinas ou sistemas de inteligência artificial. Para tanto, será necessário estabelecer um diálogo com os debates estabelecidos pela filosofia da mente sobre performance da consciência e sobre como é possível atribuir mente a terceiros. Algumas categorias conceituais nos servirão de base teórica. As principais categorias conceituais a serem articuladas aqui são: inteligência artificial (fraca e forte no sentido proposto por J. Searle), status ontológico dos agentes morais (conforme apresentado por P. Silveira), status moral maquínico (de acordo com Nick Bostrom e Eliezer Yudkowsky), tonalidade afetiva (Stimmung) e pensamento analógico (ambas conforme apresenta Byung-Chul Han), intencionalidade (consciência do agente – J. Searle mais uma vez) e background (especificamente conforme o proposto por M. Heidegger e por S. Gallagher). As mencionadas categorias (com seus respectivos anteriormente citados articuladores teóricos) nos servirão como elementos de análise e de arcabouço teórico para ajudar a oferecer uma resposta minimamente aceitável à questão diretora do presente artigo.

Palavras-chave: Inteligência artificial, mente, consciência moral

1. Introdução – Aprestando o problema

A questão diretriz (D) do presente artigo se baseia na especulação sobre se há (ou não), sistemas operacionais de linguagens artificiais tais que podemos lhes

¹ Docente da Coordenação de Filosofia da Universidade Federal do Piauí – CEAD. Coordenador da Especialização em Ciências Humanas e Sociais Aplicadas e o Mundo do Trabalho.
E-mail dayvide08@ufpi.edu.br

atribuir o status de sistemas dotados de consciências morais. Apesar das implicações éticas que o uso da inteligência artificial pode suscitar, nossa questão aqui a ser pautada não se reportará sobre os desdobramentos éticos do uso da inteligência artificial (IA). A despeito da importância de tal pauta, nosso foco parte das seguintes perguntas: Inteligências artificiais poderiam apresentar comportamentos considerados típicos de agentes morais (quando oferecidos os devidos inputs ao seu capital de linguagem de programação)? Tal comportamento, se verificável, uma vez verificado, poderia ser considerado meramente mimético (mera imitação da inteligência humana natural), ou a ele poderíamos atribuir algum nível de status de consciência moral?

A pergunta sobre a possibilidade de haver uma espécie de consciência moral em sistemas de linguagens artificiais, as IA, evidentemente demanda outras questões. Podemos dizer que requer oferecermos uma rede coerente de respostas a questões recursivamente basilares. Dito isso, é necessário oferecer uma resposta para as seguintes perguntas:

(i) O que é uma IA?

(ii) Inteligências artificiais possuem consciência? – A resposta para (ii), podemos dizer, é homônima teórica à resposta para a questão sobre se sistemas artificialmente inteligentes possuem mentes (debate canônico da filosofia da mente);

(iii) Como podemos atribuir consciência/mente a agentes biologicamente analógicos e a sistemas operacionais artificiais (IA)?

Certamente que a resposta para (iii) pode ser, ao menos superficialmente, a seguinte (R – iii): Há determinados comportamentos performáticos que são indicadores de ações típicas de agentes dotados de consciência/mente. Evidentemente, por ser uma resposta superficial, demanda-se responder o seguinte:

(iii¹) Quais são os comportamentos que podemos esperar de agentes dotados de mente/consciência?

Alguns textos e entrevistas serão considerados aqui como principais fornecedores de elementos teóricos necessários à articulação e elaboração dos argumentos que faremos uso. Para respondermos (i), serão centrais os textos "*Matter and consciousness*" (CHURCHLAND, 2013), "*The ethics of artificial*

intelligence" (BOSTRUM; YUDKOWSKY, 2011) e a entrevista sobre inteligência artificial, via YouTube (Canal Opera Mundi), dada por Miguel Nicolelis no dia 12 de junho de 2023. Para respondermos (ii), os textos basilares serão *"Computing machinery and intelligence"* (TURING, 1950), *"Minds, brains, and programs"* (SEARLE, 1980), *"The Rediscovery of the Mind"* (SEARLE, 1995), *"The mystery of consciousness"* (SEARLE; DENNETT; CHALMERS, 1997), *"Could a Machine Think? – Classical AI is unlikely to yield conscious machines; systems that mimic the brain might"* (CHURCHLAND, P.M.; CHURCHLAND, P.S., 2014) e *"Ética e Inteligência Artificial: Da possibilidade filosófica de Agentes Morais Artificiais"* (SILVEIRA, 2021). Por fim, para respondermos (iii) e (iii¹), utilizaremos *"Introdução ao pensamento de Heidegger"* (STEIN, 1966), *"The practice of the mind: theory, simulation or primary interaction"* (GALLAGHER, 2001) e *"Não-coisas: Reviravoltas do mundo da vida"* (HAN, 2022).

Algumas categorias conceituais nos servirão de base teórica. As principais categorias conceituais a serem articuladas aqui são: inteligência artificial (fraca e forte no sentido proposto por J. Searle), status ontológico dos agentes morais (conforme apresentado por P. Silveira), status moral maquínico (de acordo com Nick Bostrom e Eliezer Yudkowsky), tonalidade afetiva (Stimmung) e pensamento analógico (ambas conforme apresenta Byung-Chul Han), intencionalidade (consciência do agente – J. Searle mais uma vez) e background (especificamente conforme o proposto por M. Heidegger e por S. Gallagher). As mencionadas categorias (com seus respectivos anteriormente citados articuladores teóricos) nos servirão como elementos de análise e de arcabouço teórico para ajudar a oferecer uma resposta minimamente aceitável à questão diretora do presente artigo.

2. A tese

Dada a questão diretriz (D) deste artigo, é possível extrairmos duas teses contrárias entre si: (T1) Podemos atribuir consciência moral a ações/respostas produzidas por sistemas operacionais de inteligências artificiais; e (T2) Não podemos atribuir consciência moral a ações/respostas produzidas por sistemas operacionais de inteligências artificiais. Por exclusão, dado o critério lógico que diz

que não podemos assumir proposições contrárias, podemos apenas assumir uma das duas teses.

As premissas que darão vigor argumentativo para **(T1)** e **(T2)** serão diretamente extraídas da resposta que encontraremos para a pergunta **(ii)** e, indiretamente, inferidas a partir da resposta obtida para a pergunta **(iii¹)** – ambas já mencionadas na seção anterior deste texto. Aclarando melhor, temos: de **(ii)** podemos obter duas respostas, cada uma delas performando como premissa ou para **(T1)** ou para **(T2)**. Se obtivermos uma resposta positiva, inteligências artificiais possuem consciência, então temos uma premissa favorável a **(T1)**; se obtivermos uma resposta negativa, inteligências artificiais não possuem consciência, então teremos uma premissa para **(T2)**. De modo inferencial, por derivação, partindo da pergunta **(iii¹)**, podemos também extrair premissas para **(T1)** e para **(T2)**. Explicamos: Se considerarmos que há uma questão derivável de **(iii¹)**, que seria **(iii^{1d})** “inteligências artificiais possuem comportamentos típicos de agentes dotados de consciência?”, poderíamos extrair também duas premissas, sendo uma positiva, inteligências artificiais possuem comportamentos típicos de agentes dotados de consciência, e outra negativa, inteligências artificiais não possuem comportamentos típicos de agentes dotados de consciência. Esta última, é premissa agregável a **(T2)**, enquanto a primeira, é uma premissa agregável a **(T1)**.

Assim sendo, podemos dizer que a estrutura do argumento **(T1)** é a seguinte:

(T1a) – Se inteligências artificiais possuem consciência, então podemos, por derivação, atribuir possibilidade de consciência moral a ações/respostas produzidas por sistemas operacionais de inteligências artificiais.

(T1b) – A atribuição de consciência e a atribuição de consciência moral a inteligências artificiais, por derivação lógica, advém da possibilidade de inteligências artificiais performarem comportamentos típicos de agentes dotados de consciência.

(T1c) – Inteligências artificiais performam comportamentos típicos de agentes dotados de consciência.

(T1d) – A partir de **(T1c)**, com base no princípio de implicação empregado em **(T1a)**, inteligências artificiais podem performar comportamentos típicos de agentes dotados de consciência moral.

Logo **(T1)**, podemos atribuir consciência moral a ações/respostas produzidas por sistemas operacionais de inteligências artificiais.

O argumento a favor de **(T2)** seguirá a mesma estrutura e os mesmos princípios de implicação lógica empregados em **(T1a)** e **(T1b)**. Porém, diferente do argumento **(T1)** acima exposto, em **(T2)** teremos a negação das premissas **(T1c)** e **(T1d)** de **(T1)**, o que nos levará à conclusão de **(T2)**: não podemos atribuir consciência moral a ações/respostas produzidas por sistemas operacionais de inteligências artificiais. A estrutura do argumento **(T2)** é a seguinte:

(T2a) – Se inteligências artificiais possuem consciência, então podemos, por derivação, atribuir possibilidade de consciência moral a ações/respostas produzidas por sistemas operacionais de inteligências artificiais.

(T2b) – A atribuição de consciência e a atribuição de consciência moral a inteligências artificiais, por derivação lógica, advém da possibilidade de inteligências artificiais performarem comportamentos típicos de agentes dotados de consciência.

(T2c) – Inteligências artificiais não performam comportamentos típicos de agentes dotados de consciência.

(T2d) – A partir de **(T2c)**, com base no princípio de implicação empregado em **(T2a)**, inteligências artificiais não podem performar comportamentos típicos de agentes dotados de consciência moral.

Logo **(T2)**, não podemos atribuir consciência moral a ações/respostas produzidas por sistemas operacionais de inteligências artificiais.

A principal premissa do argumento **(T1)** é a premissa **(T1c)**. Dizemos principal premissa porque, uma vez demonstrada falsa, não há como manter **(T1d)** (dela derivada) e, por conseguinte, a conclusão do argumento de **(T1)**. Da mesma forma

que (T1c) é essencial para a manutenção da conclusão de (T1), a solidez e aceitação da conclusão de (T2) dependem especificamente da solidez de (T2c). Uma vez demonstrado que (T2c) é falsa, a conclusão de (T2) também será falsa.

Há evidentemente, um embate teórico que subjaz a aceitação de (T1c) ou a aceitação de (T2c). Em síntese, em um texto de 1950, intitulado "*Computing Machinery and Intelligence*", Alan Turing defende exatamente a ideia impressa em (T1c). Nos anos seguintes à publicação de "*Computing Machinery and Intelligence*", as ideias de Turing provocaram calorosos embates teóricos e muitos foram os argumentos contrários a Turing². Destacamos, contudo, um texto de 1980, de autoria de John Searle, cujo título é "*Minds, brains, and programs. Behavioral and Brain Sciences*". A razão de destacarmos esse texto específico de Searle é dupla: a) nele, Searle nega abertamente (de forma bastante consistente) a ideia impressa em (T1c); e b) defende abertamente a ideia impressa em (T2c)³. Nesse sentido, poderemos "terceirizar" a análise da verdade/falsidade das premissas (T1c)/(T2c) ao resultado do embate entre Turing e Searle. Nossos esforços nas próximas seções deste artigo seguirão no sentido de aderir aos argumentos de Searle.

3. Definição de (IA)

Antes de entrarmos no embate argumentativo sobre a possibilidade de atribuímos (ou não) consciência moral a sistemas de inteligências artificiais, cremos que se faz necessário estabelecer a definição de alguns termos e expressões. A finalidade é a de centrar um eixo conceitual que possa oferecer alguma segurança na compreensão do uso terminológico dos conceitos e processos que abordaremos ao longo do texto. Nesse sentido, a expressão mais basilar aqui empregada é "*inteligência artificial*". Como o título sugere, é o conceito de inteligência artificial que

² Podemos aqui destacar a Conferência ocorrida no Dartmouth College (EUA), o evento intitulado de "*Dartmouth Summer Research Project on Artificial Intelligence*", de 1956. É importante destacar que a expressão "inteligência artificial" foi empregada academicamente pela primeira vez neste evento (professor John McCarthy).

³ Trataremos mais detalhadamente desta questão em uma seção específica. Contudo, antecipando um pouco o roteiro teórico de Searle, basta dizermos agora que o ponto principal do argumento de Searle frente Turing reside na diferença entre inteligência artificial fraca e inteligência artificial forte (estabelecida por Searle).

teremos como foco principal nesta seção. Porém, acreditamos ser conveniente antes apresentarmos explicações prévias acerca de alguns elementos basilares da computação e da linguagem computacional. É o que será feito na subseção seguinte.

3.1. Hardware, software, linguagem de máquina e linguagem de programação

Paul Churchland, em um texto de 1984 ⁴, "*Matter and consciousness*", traça uma descrição dos processos e da estrutura computacional com a finalidade de aclarar como performam os sistemas de inteligência artificial. Churchland, mais especificamente no capítulo seis ⁵ do mencionado texto, consegue de forma sintética e clara descrever com precisão como processam as máquinas e como seus sistemas operam. Por ser bastante didático, o mencionado texto do Churchland nos servirá de base teórica para a presente subseção. Assim, quando nos reportarmos aos conceitos de hardware, software, linguagem de máquina ou linguagem de programação, estaremos partindo do horizonte teórico proporcionado por Churchland em "*Matter and consciousness*".

Fundamentalmente, o processamento de informações em um computador tem por base um sistema formal. Nas palavras de Churchland (2013, p.159), um sistema formal consiste em "(1) um conjunto de fórmulas e (2) um conjunto de regras de transformação para manipulá-las [as fórmulas] ⁶". As fórmulas, por sua vez, segundo Churchland (2013, p.159), "*são elaboradas pela junção, por meio de regras de formação específicas (...)*". Tais regras específicas de formação, que podemos chamar simplificada e de operações computacionais, são dadas a partir de processos controlados de transformação. As transformações ocorrem na passagem de uma fórmula para outra. O controle é feito a partir de programas ou softwares. Resumidamente, há uma parte física e uma parte gerencial não física, que são os programas e suas linguagens. O intermédio entre a parte não física e a parte física se dá a partir das chamadas portas de entrada da parte física – mais especificamente,

⁴ A primeira edição é de 1984. Além desta edição, há mais duas, sendo uma de 1988 e a mais recente, de 2013. Há uma tradução de 2004 para o português. Utilizamos como fonte a edição norte-americana de 2013. Citamos aqui a versão de 1984 apenas por critérios historiográficos.

⁵ Cujo título é "*Artificial Intelligence*".

⁶ Grifo nosso.

isso ocorre na central de processamento das máquinas ⁷. As portas de entrada recebem os comandos, operam as transformações de fórmulas, acessam a memória do computador (RAM e HD), operam nova transformação e, por fim, geram a resposta requerida – tudo com base no capital de dados e programação instalada na máquina.

É importante destacar que quando mencionamos “transformação”, estamos especificamente nos reportando à transformação de uma linguagem para outra, ou também, a transformação de uma fórmula para outra. Em termos muito gerais, com finalidade didática, podemos dizer que quando temos uma expressão numérica como, por exemplo, $x' = (5 - 2) + (3 \times 7)$ e, após efetuados os cálculos, chegamos ao resultado $x'' = 25$, podemos dizer que processamos uma transformação da expressão $x' = (5 - 2) + (3 \times 7)$ para o resultado obtido $x'' = 25$. Em termos comparativos, quando abrimos um editor de texto e digitamos uma palavra no teclado do laptop, por exemplo “**casa**”, estamos ativando a central de processamento do laptop, suas portas de entrada, a partir de uma linguagem binária, **0** e **1**. Esta linguagem é traduzida (transformada) na tela do laptop como a palavra “**casa**”. A esta linguagem binária chamamos linguagem de máquina.

A parte física, ou hardware, e os sistemas operacionais, ou software, são, portanto, mutuamente dependentes. O software determina as demandas sobre ações a serem executadas e os hardwares executam as ações demandadas. Esse processo de interação, conforme já descrito anteriormente, ocorre por meio de transformações de fórmulas – linguagem de máquina de fórmulas binárias **0**, **1**, na fórmula equivalente desejada; no caso exemplificado anteriormente, a palavra “**casa**”. Assim, por exemplo, quando instalamos um aplicativo em nossos celulares (ou computadores), estamos introduzindo neles um software que nada mais é do que um “*jogo de linguagem*” ⁸, com regras bem definidas, com a finalidade de estabelecer regras e definir a conduta operacional da máquina (computador, laptop, celular etc.).

De modo resumido, podemos dizer que as regras do “*jogo de linguagem*” computacional são definidas por padrões de ação operacional que podemos chamar

⁷ Toda vez que fora mencionada a palavra “máquina”, estamos significando, de forma genérica, computador, laptop, celular etc.

⁸ Explicar a referência

"algoritmos". Ao jogo de linguagem que performa a partir das regras de lotes de algoritmos, podemos chamar de linguagem de programação. Esta, por sua vez, uma vez instalada na máquina, performa sendo acionada a partir da geração de comandos, os *inputs*. Os inputs acessam a central de processamento da máquina, a CPU, que ativa a memória randômica (memória RAM). A memória RAM tem a finalidade de acessar aleatoriamente (vai depender do input) a memória rígida da máquina (HD) e desta extrair informações para dar a resposta (*output*) requerida pelo input gerado. Em outras palavras, quando acessamos o buscador do laptop, que desejosos de encontrar determinado arquivo, geramos um input. O input gerado acessa a CPU. A CPU então aciona a memória RAM, que vai acessar as informações guardadas no HD. A linguagem de programação vai ditar a performance, gerando a resposta desejada ao expor o arquivo rastreado – rastreamento desencadeado pelo input "**buscar arquivo**" gerado na máquina.

Mas o que falar sobre hardware, software, linguagem de máquina e linguagem de programação tem a ver com inteligência artificial? Respondendo de forma bastante direta, o processo performativo das inteligências artificiais segue basilarmente o mesmo padrão comportamental das transformações de fórmulas descritas nesta subseção: manipulação e transformação de símbolos e fórmulas através de uma linguagem de programação. No que pode dizer respeito aos hardwares, às inteligências artificiais não seria possível operar sem um hardware que as hospedasse e, uma vez requerido por elas, através de seu banco de dados (as memórias), produzissem respostas demandadas por usuários humanos.

3.2. Inteligência artificial: duas definições e uma proposta

Dentre as variadas definições que poderíamos oferecer, por questões didáticas/práticas, apresentaremos duas. Tentaremos estabelecer alguma relação comparativa entre ambas e, a partir disso, adotar uma delas – ou mesmo uma terceira definição extraída das duas que serão apresentadas nesta seção. A primeira delas foi extraída do diálogo entre nós e um tipo específico de inteligência artificial, o AIChat, desenvolvida a partir do Chat GPT e GPT-4. A segunda definição que apresentaremos foi extraída de uma entrevista sobre inteligência artificial, dada por

Miguel Nicolelis (neurocientista) ao canal de YouTube Opera Mundi (em 12 de junho de 2023).

3.2.1. O diálogo com IA

O objetivo de extrair uma definição de **IA** partir do diálogo com inteligências artificiais é duplo: tentar apresentar contextos de experiências de interação com inteligências artificiais; demonstrar que, apesar de requerermos que a inteligência artificial apresente, por "*ela mesma*"⁹, um conceito de inteligência artificial, parecendo daí extrairmos uma espécie de meta-conceito de inteligência artificial de uma **IA**, não é possível a qualquer **IA**, de per si, produzir conceitos. Tentando atingir os citados objetivos, fizemos três perguntas ao AI Chat. As perguntas, como veremos, se referem a conceito de **IA**, modus operandi de uma **IA** e sobre a diferença entre inteligência artificial e inteligência natural (humana). Passamos a transcrever o diálogo¹⁰.

(PI) – Nós: O que é inteligência artificial?

(RI) – AI Chat: *A inteligência artificial refere-se à capacidade de um sistema de computador ou máquina de simular processos de inteligência humana, como aprendizado, raciocínio, percepção e tomada de decisões. É um campo da ciência da computação que busca desenvolver algoritmos e sistemas capazes de executar tarefas complexas sem intervenção humana, além de aprender e se adaptar com base em experiências anteriores. A inteligência artificial pode ser aplicada em uma ampla gama de setores, como medicina, finanças, transporte, manufatura, entretenimento, entre outros.*

(PII) – Nós: Em que se baseia a tomada de decisões de um sistema de inteligência artificial?

⁹ Destacamos esta expressão para chamar a atenção ao fato de que sistemas de inteligências artificiais não têm (atualmente) a capacidade de produzir conhecimento/saber a partir de si. Como ficará mais claro no decorrer desta seção, o produto informacional gerado ou inferido por IA's é resultado do acesso desta aos bancos de dados, redes neurais etc.

¹⁰ AI Chat – Chat GPT e GPT4. **O que é inteligência artificial.** Chat GPT é uma inteligência artificial de linguagem natural desenvolvida pela OpenAI, que usa uma arquitetura de rede neural para gerar respostas a perguntas feitas por usuários. Acesso em 02 de julho de 2023. (<https://play.google.com/store/apps/details?id=com.ai>).

(RII) – AI Chat: A tomada de decisões de um sistema de inteligência artificial é baseada em vários fatores e técnicas, dependendo do tipo de algoritmo ou abordagem aplicada. Alguns dos principais componentes envolvidos na tomada de decisões são: 1. Dados de entrada: Os sistemas de inteligência artificial são alimentados com grandes conjuntos de dados de treinamento, que podem consistir em imagens, textos, áudios, vídeos, entre outros. Esses dados são usados para aprender padrões e criar modelos que ajudem na tomada de decisões. 2. Algoritmos de aprendizado de máquina: Uma técnica comum usada na inteligência artificial é o aprendizado de máquina. Nessa abordagem, os algoritmos são treinados em dados de treinamento para desenvolver modelos que possam fazer previsões ou classificações com base em novos dados. 3. Redes neurais artificiais: As redes neurais artificiais são um tipo de algoritmo que tenta imitar o funcionamento dos neurônios no cérebro humano. Elas são usadas para resolver problemas complexos, reconhecer padrões e tomar decisões. 4. Lógica e raciocínio: Alguns sistemas de inteligência artificial baseiam-se em regras lógicas e raciocínio simbólico para tomar decisões. Essas regras são definidas por especialistas em um determinado domínio e o sistema as aplica para chegar a conclusões ou tomar decisões. 5. Algoritmos de otimização: Em certos casos, os sistemas de inteligência artificial podem usar algoritmos de otimização para encontrar a melhor solução em um espaço de busca. Esses algoritmos tentam minimizar ou maximizar uma função objetiva, considerando restrições e preferências definidas pelo sistema. Lembrando que a tomada de decisões da inteligência artificial não é infalível e depende da qualidade dos dados de treinamento, da robustez dos algoritmos e do contexto em que o sistema está sendo aplicado.

(PIII) – Nós: Qual a diferença entre um Inteligência artificial e uma inteligência natural humana?

(RIII) – AI Chat: Existem várias diferenças entre uma inteligência artificial (IA) e uma inteligência natural humana. Aqui estão alguns pontos-chave: 1. Origem: Uma IA é um sistema desenvolvido pelo ser humano, enquanto a inteligência natural humana é inerente aos seres humanos. 2. Capacidade de Aprendizado: Uma IA é projetada para aprender com os dados fornecidos através de algoritmos específicos de aprendizado de máquina, enquanto a inteligência natural humana tem a

capacidade de aprender e assimilar informações de diversas fontes, como experiências, leitura, observação, interações sociais, etc. 3. Processamento de informações: Uma IA é altamente capaz de processar enormes quantidades de dados em velocidades muito altas, superando as capacidades humanas de processamento. Por outro lado, os humanos têm uma capacidade de processamento mais limitada, mas possuem habilidades cognitivas complexas, como raciocínio, abstração, lógica, criatividade e interpretação do contexto. 4. Consciência e Emoção: Até o momento, as IA não possuem consciência ou emoções, diferentemente das pessoas que têm consciência de si mesmas, experienciam emoções e têm uma compreensão subjetiva do mundo. 5. Flexibilidade e Adaptabilidade: As IA são altamente especializadas em tarefas específicas para as quais são treinadas, enquanto os seres humanos têm a capacidade de se adaptar a diferentes situações e aprender diferentes habilidades e conhecimentos. 6. Intuição e Criatividade: Os humanos têm a capacidade de usar a intuição em sua tomada de decisão e aplicar a criatividade para resolver problemas e gerar novas ideias. Embora IA possa simular algumas dessas habilidades, a capacidade humana nesses aspectos é altamente sofisticada e única. Em suma, enquanto uma IA pode ser programada para realizar tarefas específicas com habilidades superiores em processamento de dados, a inteligência natural humana é caracterizada por suas habilidades cognitivas complexas, consciência própria, emoções e flexibilidade adaptativa.

Das questões levantadas e respondidas pelo AI Chat elencadas acima, podemos, resumidamente, extrair o seguinte saldo informativo: a) inteligências artificiais são sistemas alimentados por algoritmos performando em máquinas com a finalidade última de mimetizar processos típicos da inteligência humana; b) apesar das IA's reproduzirem comportamento de aprendizado e de inferência similares aos da inteligência humana (natural), o fazem a partir de arcabouços de dados previamente fixados e armazenados, o que limita a articulação de informações e geração de inferências a partir de uma capital informacional já dado – há um lote de dados a partir do qual a IA processa e gera informações circunstanciadas pelo referido lote de dados; e c) apesar de poder unir, separar, ou inferir informações, tal operação não consegue projetar informações novas, completamente diferentes dos

dados de que dispõe a **IA** (bancos de dados, redes neurais etc.), o que demonstra capacidade criativa limitada ou quase nula.

Assim, uma **IA**, diferente da inteligência humana (**IH**), não é capaz de criar, de per si, novas informações. Nesse sentido, perguntar a um sistema como o AI Chat o que é inteligência artificial é gerar um input que acessa a rede neural disponível ao AI Chat, seus dados, e receber, como output da rede neural linkada ao AI Chat, uma definição de **IA**.

3.2.2. **IA e o Demônio de Laplace**

No dia 12 de junho de 2023, Miguel Nicolelis deu uma entrevista ao canal Opera Mundi. O objetivo da entrevista era debater com o neurocientista (Nicolelis) sobre inteligência artificial. Da citada entrevista, podemos extrair, também como o foi na subseção anterior a esta, um conceito de **IA**, o modus operandi de uma **IA** e a diferença entre **IA** e **IH**. Nas linhas que se seguirão nesta subseção, serão relatados os posicionamentos e conceitos sobre **IA** abordados e apresentados por Miguel Nicolelis. Apenas algumas informações prévias sobre a entrevista. A abordagem dada ao tema por Nicolelis é emoldurada pela temática política e econômica de caráter marxista. No fim das contas, a leitura conjuntural que Nicolelis faz do contexto social atual estabelecido pela inteligência artificial é de viés marxista, traçando inclusive paralelos com o advento da revolução industrial (segunda metade do século XVIII à primeira metade do século). Não trataremos aqui sobre esta abordagem política e econômica feita por Nicolelis. Considerando o escopo deste artigo, ressaltaremos apenas o caráter conceitual e descritivo sobre **IA** proposto pelo neurocientista.

Nas linhas que se seguirão nesta subseção serão feitos o relato da entrevista do Nicolelis ao canal Opera Mundi e, paralelo a isso, nossos comentários acerca do conteúdo da citada entrevista.

Quando perguntado pelo entrevistador (Breno Altman) sobre qual seria o conceito de inteligência artificial, Nicolelis afirma que Noam Chomsky partilha da sua mesma perspectiva, que é a de que não há de fato inteligência artificial. Pelo menos não como o queriam fazer crer um grupo norte-americano de matemáticos e

cientistas da computação da década de cinquenta. Estes, segundo Nicoletis, defendiam a tese de que é possível à inteligência artificialmente criada a capacidade de pensar inteligentemente como humanos e, de forma meio escatológica, a capacidade de suplantar e dominar a humanidade. Nas palavras de Nicoletis, essa perspectiva de IA teria sido apenas uma jogada de marketing para atrair investimentos do governo estadunidense. Ainda segundo Nicoletis, inteligência é um atributo apenas de "*organismos*" e que depende de um meio "*analógico*"¹¹. A inteligência, nesse sentido, não ocorreria a partir de meios digitais, uma vez que, segundo Nicoletis, entre o **0** e **1** da linguagem de computador há um vasto universo "*operativo*"¹² de coisas que os cérebros humanos e animal são capazes de realizar e que a IA não o é.

Contudo, há um algo a que corriqueiramente é atribuído o rótulo de inteligência artificial e que Nicoletis, apesar de sua perspectiva sobre não podermos atribuir inteligência a dispositivos e sistemas digitais, não pode ignorar. Nesse sentido, Miguel Nicoletis apresenta a seguinte definição para IA: "*a inteligência artificial seria a aplicação de métodos estatísticos para, basicamente, fazer mineração de grandes bancos de dados e extrair correlações, projeções e estimativas de futuro – como se o futuro fosse reproduzir o passado*". Ainda nas palavras de Nicoletis, a IA trabalha com o mesmo princípio da probabilidade de Laplace, bastante famoso do século XIX, que parte do pressuposto de que, em termos probabilísticos, o futuro pode ser previsto a partir dos dados obtidos no passado. Seria o "*futuro sem futuro*", como afirma Nicoletis. Alegorizando a definição aqui por ele apresentada de IA, Nicoletis diz que: "*A inteligência artificial deveria ser chamada de o demônio de Laplace*"¹³.

¹¹ No livro "*Não-coisas, reviravoltas do mundo da vida*", de 2022, Byung-Chul Han utiliza o mesmo termo para opor inteligência não artificial a inteligência artificial.

¹² A palavra "*operativo*" é empregada aqui por nós. Não é uma palavra que o Nicoletis tenha usado na entrevista aqui em questão. E empregamos "*operativo*" para diferenciar dos processamentos de máquina que a eles é atribuída a palavra "*operacional*". De forma muito resumida, enquanto "*operativo*" indica ação deliberada a partir da intencionalidade de um sujeito, "*operacional*" remete a um software disponível para trabalhar em termos de input-output.

¹³ No texto "*Théorie analytique des probabilités*" (1812, p. 08), Laplace, ainda na introdução, diz: "Devemos, portanto, considerar o estado atual do universo como o efeito de seu estado anterior e como a causa do que se seguirá" (no original "*Nous devons donc envisager l'état présent de l'univers comme l'effet de son état antérieur et comme la cause de celui qui va suivre.*"). Este trecho citado trata exatamente do mesmo princípio probabilístico por trás do processamento operacional das IA's de que fala Nicoletis. De modo mais radical ainda, eis aí o princípio da razão suficiente. O próprio Laplace reconhece que sua teoria da probabilidade é inspirada na defesa de Leibniz do princípio da razão

A segunda pergunta levantada por Breno Altman foi referente às possíveis limitações, em termos comparativos, da inteligência artificial em relação à inteligência humana. A pergunta, mais precisamente, foi sobre o que a inteligência humana é capaz de fazer e a inteligência artificial, não. De forma bastante direta, Nicolelis responde: *"Criar! Nenhum sistema digital cria; inova; nenhum sistema digital poetiza; nenhum sistema digital cria do nada algo completamente revolucionário; ele simplesmente imita, ou compila tudo aquilo que foi produzido pela mente humana e gera um relatório estatístico"*. Nesse trecho da entrevista, há um momento bastante relevante sob o ponto de vista teórico da filosofia da mente. Nicolelis cita Alan Turing ¹⁴ e diz que aparentemente a maioria das pessoas não lê a tese de Turing até o final. Ele, Nicolelis, afirma que Turing já alertava na segunda parte da sua tese sobre certas coisas serem impossíveis de serem feitas nos (pelos) sistemas digitais. Fazendo uma citação livre de Turing, para certas coisas, diz Nicolelis, seria necessário um oráculo. E, como conclui o próprio Nicolelis, tal oráculo seria exatamente o humano.

Ainda na tentativa de estabelecer pontes comparativas entre **IA** e **IH**, o entrevistador Breno Altman levanta um questionamento que parece tentar contradizer o que Nicolelis defendeu sobre sistemas digitais não serem capazes de criar informações e a partir do nada. O Altman apela para o fato de que o conhecimento humano seria produzido, inclusive novos conhecimentos, a partir da síntese de conhecimentos do passado da humanidade. Nesse pressuposto, pergunta Altman: O texto criado pelo Chat GPT, quando alguém lhe faz uma pergunta, e quando o Chat digita a resposta no formato de texto, esse tal texto não seria uma criação – embora seja ele baseado numa compilação de dados que alimentam o sistema? O Nicolelis diz que não, que tal tarefa do Chat GPT não passa de um exercício de plágio a partir do compilado de informações já existentes – como agravante de que você acessa uma única resposta, que é a que o chat quer oferecer.

suficiente. Em termos alegóricos, e para tratar dos princípios da probabilidade, Laplace menciona um certo *"Esprit"* ou Demônio (como muitos traduziram, e como o próprio Nicolelis denomina), com capacidade de cálculo maior que os humanos, que poderia, sem erros, prever todo o futuro com base nos dados do passado.

¹⁴ Certamente Nicolelis esteja se referindo ao texto de 1950, *"Computing Machinery and Intelligence"* – embora não o diga abertamente na entrevista.

Uma experiência rápida de diálogo com o AI Chat pode facilmente comprovar a assertiva de Nicolelis de que as respostas oferecidas por sistemas do tipo Chat GPT são plágios. No dia 12 de julho de 2023 abrimos um breve diálogo com o AI Chat ¹⁵, cuja temática se baseava nos conceitos de fenômeno em Kant e em Heidegger. Fizemos três perguntas. As duas primeiras foram: 1) O que é fenômeno para Kant e o que é fenômeno para Heidegger? 2) Sendo que apresentam definições diferentes para fenômeno, qual dos dois está certo, Kant ou Heidegger? Para a primeira pergunta, o AI Chat apresentou a definição de fenômeno em ambos os autores e concluiu apontando uma breve distinção entre os conceitos dados por Kant e Heidegger sobre fenômeno. Para a segunda pergunta, o AI Chat deu como resposta novamente a definição de fenômeno para ambos os filósofos. Porém, destacou que a proposta de Heidegger é uma crítica à definição kantiana. Conclui, no entanto, de modo genérico, ao dizer que a escolha de qual definição é correta vai depender do quadro filosófico adotado e dos objetos de reflexão.

Observemos que as duas respostas oferecidas pelo AI Chat podem ser elaboradas sem requerer maior esforço reflexivo ou crítico. Um bom arcabouço de dados, como uma rede neural, é capaz de oferecer os dados informacionais necessários para a elaboração (por compilação de informações) das repostas mencionadas no parágrafo anterior. Na terceira pergunta, resolvemos testar a capacidade de oferecer uma resposta crítica/reflexiva. A pergunta foi: Mas para o este sistema de inteligência artificial, qual dos dois está certo sobre o conceito de fenômeno – uma vez que uma proposta é crítica da outra? A resposta dada pelo AI Chat foi:

Como uma inteligência artificial, não tenho uma opinião ou perspectiva própria. O que posso fazer é fornecer informações e explicar as diferentes abordagens filosóficas de Kant e Heidegger em relação ao conceito de fenômeno. Caberá a você, como usuário, analisar e avaliar qual abordagem faz mais sentido ou se alinha com suas próprias concepções filosóficas.

¹⁵ AI Chat – Chat GPT e GPT4. **O que é fenômeno para Kant e o que é fenômeno para Heidegger?** Chat GPT é uma inteligência artificial de linguagem natural desenvolvida pela OpenAI, que usa uma arquitetura de rede neural para gerar respostas a perguntas feitas por usuários. Acesso em 12 de julho de 2023. (<https://play.google.com/store/apps/details?id=com.ai>).

Em confirmação ao já preconizado por Nicolelis, o AI Chat, como qualquer outro sistema de inteligência artificial, é incapaz de gerar uma resposta nova, que seja baseada na criatividade. Na entrevista ao Opera Mundi (foco desta subseção), o entrevistador diz ao Nicolelis que poderia ter sido um ótimo aluno de química com o Chat GPT (quando de sua época de escola). O Breno Altman não deixa de ter razão quando o contexto demanda questões objetivamente extraídas de dados preexistentes. O Nicolelis, contudo, contradita a Altman ao dizer que: *"Por outro lado, você seria um péssimo aluno de filosofia; isso porque o Chat GPT, quando confrontado com perguntas existenciais, falha miseravelmente porque ele nunca existiu, nunca experimentou o que é ser um ser humano..."*. Com nossa experiência aqui relatada sobre solicitar do AI Chat um posicionamento crítico acerca do conceito filosófico de fenômeno, percebemos que até mesmo em questões filosóficas que não envolvem dilemas existências, até mesmo nesse quesito os sistemas de inteligências artificiais falham miseravelmente com as questões filosóficas.

3.2.3. A síntese dos conceitos

Apresentadas as duas definições de inteligência artificial, uma a partir do diálogo com o AI Chat, a outra, extraída de uma entrevista dada pelo neurocientista Miguel Nicolelis, percebemos que ambas apresentam pontos bastante similares e, apesar das semelhanças, ambas também apresentam elementos de distanciamento. O AI Chat define **IA** como um sistema capaz de imitar a inteligência humana a partir do acesso a um robusto provedor de dados. O Miguel Nicolelis define a inteligência artificial como sistema de aplicação de métodos estatísticos com a finalidade de estabelecer correlações, projeções e estimativas de futuro a partir da mineração de informações acessadas em grandes bancos de dados. Como elemento de similaridade, ambas as respostas convergem para o fato de que os sistemas de inteligência artificial necessitam de um vasto banco de dados para cumprir sua função operacional de oferecer respostas e produtos com base na compilação e inferência de dados previamente existentes. Se distanciam, contudo, quando o AI Chat assume, tacitamente, que os sistemas de **IA** são sistemas inteligentes e o Miguel Nicolelis, contrariamente a isso, declara que inteligência não é um atributo de

sistemas digitais, mas de sistemas orgânicos – a rigor, não se poderia falar em sistemas digitais artificialmente inteligentes.

Como síntese das duas propostas conceituais, assumiremos a seguinte definição:

(D^{IA}) – Inteligência artificial é, por definição, um sistema de processamento de dados com a finalidade de oferecer respostas similares às oferecidas pela inteligência animal nos níveis em que não é exigido capacidade caritativa/reflexiva própria de inteligências analógicas não artificiais.

No decorrer deste artigo, resguardadas as situações que demandarem definições do horizonte teórico dos debates em filosofia da mente, quando mencionarmos sistemas de inteligências artificiais, estaremos trabalhando com o significando de **IA** expresso na definição **(D^{IA})**.

4. Retorno aos argumentos sobre a possibilidade de atribuição de consciência moral à IA

Relembrando o conteúdo apresentado na segunda seção, foram expostos dois argumentos naquele momento. Um favorável à tese de que podemos atribuir consciência moral aos sistemas de inteligência artificial, que identificamos simbolicamente por **T1**, e outro contrário à tese de que podemos atribuir consciência moral aos sistemas de inteligência artificial, que podemos identificar simbolicamente por **T2**. As duas primeiras premissas de **T1** e **T2** coincidem. São duas premissas que indicam, respectivamente, dois princípios de implicação lógico. As premissas **(T1a)** e **(T2a)** estabelecem que se há consciência em sistemas artificiais de inteligência, então, por derivação, pode ser atribuído consciência moral a tais sistemas. As premissas **(T1b)** e **(T2b)** constituem uma implicação lógica de tal modo que estabelecem um critério para atribuição de consciência e de consciência moral para sistemas de inteligências artificial. O critério é dado a partir da verificação de comportamentos que podem ser identificados como comportamentos típicos de agentes dotados de consciência. Em outras palavras, se inteligências artificiais

performam como agentes dotados de consciência, então a elas podemos atribuir consciência.

Partindo da definição de **IA** aqui apresentada pelo AI Chat, e aceitando a teoria do Alan Turing ¹⁶ sobre mentes e computadores, a resposta é sim, inteligências artificiais podem performar como agentes dotados de consciência e, portanto, a elas podemos atribuir consciência e, por derivação, consciência moral – o que corroboraria a conclusão de **T1**. Chamaremos esse posicionamento de tese do tipo Turing, ou simplesmente de **TT**. Por outro lado, partindo da definição de **IA** apresentada por Miguel Nicolelis, e da definição de inteligência artificial forte proposta por John Searle ¹⁷, a resposta para **(T1b)/ (T2b)** é não; não podemos atribuir comportamento típico de agentes dotados de consciência a inteligências artificiais, e, portanto, a estas, por implicação, não podemos atribuir consciência ou consciência moral – o que confirmaria a conclusão de **T2**. Chamaremos a esta tese de tese do tipo Searle, ou simplesmente de **TS**.

Enquanto **TS** confirma a conclusão do argumento **T2**, negando o antecedente da implicação contida em **(T1a)/(T2a)**, dando assim suporte argumentativo para as premissas **(T2c) (T2d)** de **T2**, **TT** confirma o consequente da implicação contida em **(T1b)/ (T2b)**, dando assim suporte à conclusão de **T1** via confirmação, por derivação, das premissas **(T1c) (T1d)** de **T1**. Nas seções que seguirão, o foco será, respectivamente: verificar se podemos confirmar o consequente **(T1b)/ (T2b)**, o que, em conjunto com **(T1a)/ (T2a)**, confirmaria a conclusão de **T1**; e verificar se podemos confirmar o antecedente de **(T1a)/ (T2a)**, o que, em conjunto com **(T1b)/ (T2b)**, confirmaria a conclusão de **T1**.

5. Como podemos saber de outras mentes

A verificação sobre a veracidade do consequente de **(T1b)/(T2b)** está intimamente conectada com um dos problemas centrais da filosofia da mente, que é responder à pergunta sobre como podemos saber de outras mentes. Ao considerarmos uma resposta para a referida questão, temos que levar em

¹⁶ Veremos sobre a tese de Turing mais adiante.

¹⁷ Veremos sobre a tese de Turing mais adiante.

consideração o debate travado entre três propostas teóricas do cenário da filosofia da mente:

- Teoria da simulação – é possível saber de outras mentes a partir das experiências mentais, em primeira pessoa, do próprio agente que atribui estados mentais a um outro;
- Teoria da teoria – é possível atribuir estados mentais a outras pessoas a partir de teoria, elaboradas em terceira pessoa, sobre comportamentos de outros indivíduos.
- Teoria práticas incorporadas – é possível atribuir estados mentais a outros a partir de práticas intersubjetivas incorporadas, em nível de segunda pessoa, dentro do background histórico-social.

Segundo Shuan Gallagher (2001, p. 84), no texto "*The practice of the mind: theory, simulation or primary interaction*", a:

... teoria da simulação, argumenta que ninguém teoriza sobre outras pessoas, mas usa suas próprias experiências mentais como um modelo interno para outras mentes [...]. Para entender as outras pessoas, eu simulo os pensamentos e os sentimentos que eu poderia experimentar se estivesse no lugar do outro [...] ou eu construo em minha própria mente crenças imaginadas, desejos ou estratégias que eu uso para compreender o comportamento dos outros. Minha fonte para tais simulações não é uma teoria que tenho. Antes, eu tenho um modelo real de mente à minha disposição, isto é, tenho minha própria mente e posso usá-la para gerenciar e criar simulações.¹⁸

Levando em consideração tanto a tese **TT** quanto a tese **TS**, uma proposta teórica sobre a atribuição de outras mentes, como a descrita acima por Gallagher, em primeira pessoa, nos parece inócua quando o que está em questão é ou negarmos

¹⁸ ... simulation theory, argues that one does not theorize about the other person but uses one's own mental experience as an internal model for the other mind [...] To understand other person, I simulate the thoughts or feelings that I would experience if I were in the situation of the other [...] or I create in my own mind pretend beliefs, desires or strategies that I use to understand the other's behavior. My source for these simulations is not a theory that I have. Rather, I have a real model of the mind at my immediate disposal, that is, I have my own mind, and I use it to generate and run simulations. (Tradução nossa).

ou afirmarmos os consequentes nas implicações das premissas (T1b)/(T2b) de T1/T2. Isso porque há obviamente uma variante que impede o reconhecimento de estados mentais em sistemas de inteligência artificial a partir de uma teoria da simulação: não há como estabelecer similitude entre a estrutura das máquinas (e os sistemas de IA de que dispõem) com a estrutura cerebral e a dinâmica randômica dos pensamentos humanos. Um sujeito **S** que pretendesse atribuir mente/consciência aos sistemas de IA, que para isso partisse de seus próprios estados mentais (em primeira pessoa), não disporia de material comparativo suficiente ou para afirmar ou para negar que é possível atribuir mente/consciência a sistemas de IA – sob a pena de incorrer em arbitrariedade.

No que diz respeito à teoria da teoria, Gallagher (2001, p. 83), no mesmo texto citado anteriormente, nos diz:

O maior dogma da teoria da teoria está baseado sobre experimentos científicos bem-designados que mostram que crianças desenvolvem uma compreensão de outras mentes por volta dos quatro anos de idade. Uma versão da teoria da teoria mostra que esta compreensão é baseada sobre mecanismo inatamente especificado [neurônios espelhos, p.e.]¹⁹, um mecanismo de domínio específico designado para 'leitura' de outras mentes [...] Uma versão alternativa alega que crianças atingem esta habilidade através de curso de desenvolvimento em que testam e aprendem a partir do ambiente social [...] Comum a ambas as versões de teoria da teoria é a ideia de que crianças atingem sua compreensão de outras mentes por implicitamente empregar uma instância teórica. Esta instância envolve postular a existência de estados mentais nos outros e usar tais postulações para explicar e prever o comportamento de outras pessoas.²⁰

¹⁹ Grifo nosso.

²⁰ The major tenets of theory theory are based on well-designed scientific experiments that show children develop an understanding of other minds around the age of four. One version of theory theory claims that this understanding is based on an innately specified, domain specific mechanism designed for 'reading' other minds [...] An alternative version claims that the child attains this ability through a course of development in which the child tests and learns from the social environment [...] Common to both versions theory theory is the idea that children attain their understanding of other minds by implicitly employing a theoretical instance. This instance involves postulating the existence of mental states in others and using such postulations to explain and predict another person's behavior. (Tradução nossa)

De forma bastante resumida, por inferência do conteúdo constante na citação acima, podemos atribuir estados mentais a outras pessoas e, por similaridade, a sistemas de **IA**, a partir das experiências que temos com outras pessoas sabidamente dotadas de mente/consciência. Se pessoas ou sistemas de **IA** apresentam comportamento, ainda que simulados, iguais aos comportamentos de pessoas sabidamente dotadas de mente/consciência, então podemos atribuir mente/consciência a tais pessoas e a tais sistemas de **IA**. Nesse caso, se a teoria da toeira for confirmada como verdadeira, então a tese **TT** também o será, uma vez que, como diz a definição de **IA** apresentada pelo AI Chat, uma das funções dos sistemas de **IA** é imitar mentes humanas, e uma vez que a tese central de Turing é a de que sistemas computacionais podem imitar mentes humanas ²¹. Mas a questão que se impõe agora é: A mera imitação de comportamentos de pessoas sabidamente dotadas de mente/consciência serve como prova suficientemente forte de que sistemas de **IA** são dotados de mente/consciência? Acreditamos que o salto inferencial aí projeto parece sugerir aí o estabelecimento de indução fraca, via analogia.

O Gallagher não está disposto a assumir nem a tese da teoria da simulação, nem a tese da teoria da teoria. Ao contrário dessas propostas, sua oposição se baseada na ideia de um sujeito contextualizado, imerso em conjunto com os outros através de práticas intersubjetivas. Em adição a tal postura, ainda pode ser dito da abordagem de Gallagher que o entendimento sobre outras mentes e sobre o comportamento de outras pessoas se dá a partir de um nível pré-teórico, portanto não racional. Como ilustração da tese de Gallagher, podemos apresentar o seguinte exemplo: há determinadas expressões faciais que, sem que antes delas sejam emitidas palavras, são capazes de nos dizer da intenção de quem as faz; há, por suposição, casos em que basta um gesto facial para entendermos que o outro está triste ou preocupado, contrariado, feliz, zangado etc. E o referido entendimento se faz de modo pré-teórico porque é constituído sem que seja necessário construir uma teoria sobre a situação, ou sem que seja preciso a ocorrência de uma rotina de simulação em primeira pessoa. Nesse sentido, a compreensão é pré-teórica porque

²¹ Poderíamos mesmo, por comparação, encaixar o Turing como um teórico da teoria.

se dá através de um background (horizonte sócio-histórico) que promove práticas contextualizadas. A atribuição de crenças a outros a partir de gestos corporais só é possível porque há um contexto que lhes confere sentido e significado; o contexto da ação gestual confere seu significado e torna possível a um agente traduzir gestos em crenças e desejos. Nas palavras de Gallagher (2001, p. 86), o que ocorre é que:

... antes que eu possa desenvolver uma teoria da mente, eu devo já ter uma compreensão do outro e de sua experiência – incluindo o outro como o sujeito de uma ação intencional. Anterior à possibilidade de conhecimento da mente de outros, ou ao modo teórico ou ao modo de simulação, já é antes requerida (a) uma compreensão do que significa ser um sujeito da experiência; (b) uma compreensão do que significa que certos tipos de entidades (mas não outros) no ambiente são de fato tais coisas; e (c) uma compreensão que em alguns casos, tais entidades são similares e em outros casos deferentes de si mesmo. Ademais, para formar uma teoria sobre ou uma simulação do que outras pessoas acreditam ou desejam, nós já devemos ter conhecimento específico pré-teórico sobre como pessoas agem em contextos particulares.²²

Nesse jogo de contexto de ações intersubjetivamente incorporadas a partir de um background, subjaz um elemento novo no debate que Gallagher menciona citação acima: a consciência de um eu a partir de ações que denotam intencionalidade. Um gesto no rosto é capaz de denotar mentes demonstrando sua aprovação ou desaprovação mediante determinadas circunstâncias. Uma mera simulação de comportamento se diferencia disso porque falta-lhe a dimensão da intencionalidade. Ao passo que demonstra intencionalidade a partir de um simples gesto no rosto, um franzir de testa por exemplo, a interpretação de tal ação dotada de intencionalidade só possível porque o sujeito que interpreta ou outro está, junto com o outro, imerso

²² ... before I can develop a theory of mind I must already have an understanding of other and their experience – including the other as the subject of intentional action. Prior the possibility of knowing the other's mind in either a theoretical or simulation mode, one already requires (a) an understanding of what it means to be an experiencing subject; (b) an understanding of what it means that certain kinds of entities (but not others) in the environment are indeed such subjects; and (c) an understanding that in some way these entities are similar to and in other ways different from oneself. Furthermore, to form a theory about or to simulate what another person believes or desires, we already need to have specific pre-theoretical knowledge about how people behave in particular contexts. Tradução nossa.

em um background, ambos incorporados nesse caldo sócio-histórico de tal modo que o captar a intencionalidade de outro não se faz por intermédio de cálculos probabilísticos (ao modo das IA's), mas de maneira imediata. Uma tese assim, se demonstrada verdadeira, corroboraria a tese TS.

6. Nossa posição

Conforme dito na seção anterior, uma teoria como a teoria da simulação se mostra inócua, ou mesmo constitui uma base arbitrária quando o que está em questão é oferecer resposta minimamente plausível para pergunta sobre como podemos saber de outras mentes. Resta-nos saber qual teoria se mostra mais eficaz, ou mais satisfatória: a teoria da teoria ou a teoria das práticas intersubjetivamente incorporadas de Gallagher. Para toarmos o posicionamento devido, lembremos do embate Turing/Searle.

Paulo Silveira (2021, p.38-39), no texto *"Ética e Inteligência Artificial: Da possibilidade filosófica de Agentes Morais Artificiais"*, consegue sintetizar de modo bastante claro a tese de Turing:

Turing sugere um artifício denominado 'jogo da imitação', no qual um entrevistador precisa adivinhar quem seria o humano em uma entrevista às cegas. A pergunta do autor é o que aconteceria no caso de o entrevistador tomar por erro a máquina como se fosse o humano. Para ele esse questionamento substituiria a tradicional pergunta 'podem as máquinas pensar?'. Na impossibilidade de se definir com clareza o conceito de 'inteligência' ou 'pensar', o autor [Turing] ²³ propõe verificar se uma máquina poderia imitar com sucesso o comportamento humano a ponto de confundir um observador imparcial.

Observemos o quão similar a proposta de Turing é à tese da teoria da teoria. Destaquemos, um trecho da citação. O Silveira diz abertamente que *"Na impossibilidade de se definir com clareza o conceito de 'inteligência' ou 'pensar', o autor propõe verificar se uma máquina poderia imitar com sucesso o comportamento*

²³ Grifo nosso

humano a ponto de confundir um observador imparcial". Claramente há aqui o desafio da questão. O problema deixa de ser se podemos atribuir mente/consciência a máquinas, e passa a ser se máquinas podem ou não imitar comportamentos humanos a ponto de confundir nosso julgamento sobre a diferença entre humanos e máquinas. Parece aqui que Turing sai pela tangente. Nesse sentido, do modo como está configurado o problema proposto pro Turing, não possível atribuir consciência a sistemas de **IA**.

Mas suponhamos que imitação/simulação de mente seja um elemento importante a ser considerado sobre a possibilidade de atribuição de mente a sistemas de IA. Lançamos outra questão: A imitação da mente humana, mesmo sendo um elemento importante para atribuição de consciência a sistemas de **IA**, seria um elemento necessário? A proposta teórica de John Searle parece apontar que não. Em "*Minds, brains, and programs. Behavioral and Brain Sciences*", de 1980, Searle, logo no resumo, propõe o seguinte argumento:

Este artigo pode ser visto como uma tentativa de explorar as consequências de duas proposições. (1) A intencionalidade em seres humanos (e animais) é um produto de características causais do cérebro. Eu presumo que seja um fato empírico sobre as relações causais reais entre processos mentais e cérebros. Diz simplesmente que certos processos cerebrais são suficientes para a intencionalidade. (2) A instanciação de um programa de computador nunca é por si só uma condição suficiente de intencionalidade. Essas duas proposições têm as seguintes consequências: (3) A explicação de como o cérebro produz intencionalidade não pode ser que ele o faça instanciando um programa de computador. Esta é uma consequência lógica estrita de 1 e 2. (4) Qualquer mecanismo capaz de produzir intencionalidade deve ter poderes causais iguais aos do cérebro. Isso deve ser uma consequência trivial de 1. (5) Qualquer tentativa literalmente de criar intencionalidade artificialmente (IA forte) não poderia ser bem-sucedida apenas projetando programas, mas teria que duplicar os poderes causais do cérebro humano. Isso decorre de 2 e 4.²⁴

²⁴ This article can be viewed as an attempt to explore the consequences of two propositions. (1) Intentionality in human beings (and animals) is a product of causal features of the brain I assume this is an empirical fact about the actual causal relations between mental processes and brains It says simply that certain brain processes are sufficient for intentionality. (2) Instantiating a computer

Observemos mais uma vez uma similaridade aí expressa. Dessa vez, com relação à teoria das práticas intersubjetivamente incorporadas de Gallagher. O elemento crucial tanto em Gallagher quanto em Searle é a intencionalidade. Ambos partem do pressuposto que sem intencionalidade não há como atribuir consciência a um ente ou sujeito. Em sua premissa (2), Searle é taxativo: "*A instanciação de um programa de computador nunca é por si só uma condição suficiente de intencionalidade*". Isso porque, inferindo das passagens aqui citadas de Gallagher, a intencionalidade não é uma prática calculada simulável por máquinas. Nesse sentido, a tese de Turing, além de sair a tangente apresenta elementos insuficientes para assegurar que máquinas (ou sistemas de inteligências artificiais) possuem consciência. Com isso, acreditamos ter negado a conclusão do argumento **T1** apresentado na segunda seção deste artigo e ter demonstrado que, se não verdadeira, pelo menos bastante aceitável a conclusão de **T2**, de não podemos atribuir consciência moral a ações/respostas produzidas por sistemas operacionais de inteligências artificiais.

Referências

AI Chat – Chat GPT e GPT4. **O que é inteligência artificial.** Chat GPT é uma inteligência artificial de linguagem natural desenvolvida pela OpenAI, que usa uma arquitetura de rede neural para gerar respostas a perguntas feitas por usuários. Acesso em 02 de julho de 2023. (<https://play.google.com/store/apps/details?id=com.ai>).

_____. **O que é fenômeno para Kant e o que é fenômeno para Heidegger?** Chat GPT é uma inteligência artificial de linguagem natural desenvolvida pela OpenAI, que usa uma arquitetura de rede neural para gerar respostas a perguntas feitas por usuários. Acesso em 12 de julho de 2023. (<https://play.google.com/store/apps/details?id=com.ai>)

program is never by itself a sufficient condition of intentionality The main argument of this paper is directed at establishing this claim The form of the argument is to show how a human agent could instantiate the program and still not have the relevant intentionality. These two propositions have the following consequences (3) The explanation of how the brain produces intentionality cannot be that it does it by instantiating a computer program. This is a strict logical consequence of 1 and 2. (4) Any mechanism capable of producing intentionality must have causal powers equal to those of the brain. This is meant to be a trivial consequence of 1. (5) Any attempt literally to create intentionality artificially (strong AI) could not succeed just by designing programs but would have to duplicate the causal powers of the human brain. This follows from 2 and 4. (Tradução nossa).

BOSTROM, N; YUDKOWSKY, E. **The ethics of artificial intelligence**. In: The Cambridge Handbook of Artificial Intelligence, K. Frankish e W. M. E. Ramsey, Eds. Cambridge University Press, 2014, pp. 316–334.

CHURCHLAND, P. M. **Matter and consciousness**. Cambridge, Massachusetts: MIT Press, 2013.

_____. **Could a Machine Think? – Classical AI is unlikely to yield conscious machines; systems that mimic the brain might**. California. EUA, 2014.

FRICKER, M. **Epistemic Injustice: Power and the Ethics of Knowing**. Nova York, EUA: Oxford University Press, 2007.

GALLAGHER, S. **The practice of the mind: theory, simulation or primary interaction**. [S.l.: s.sn.], 2001. p. 83-108.

HAN, B. **Não-coisas: Reviravoltas do mundo da vida**. Trad. Rafael Rodrigues Garcia. Rio de Janeiro: Vozes, 2022.

MARTIN, H. **Ser e tempo – Parte I**. 12 ed. Petrópolis, RJ: Vozes, 2002.

OPERA MUNDI. **Inteligência artificial: tudo o que você precisa saber – Miguel Nicolelis**. YouTube, 12 de junho de 2023. Disponível em: <https://www.youtube.com/watch?v=pb4b4_MINwo . Acesso em: 14 de junho de 2023.

SEARLE, J. **Minds, brains, and programs**. *Behavioral and Brain Sciences*, v.3, n. 3, p. 417- 457, 1980.

_____. **The rediscovery of the mind**. Cambridge: MIT, 1995.

_____. **The mystery of consciousness: John R. Searle and exchanges with Daniel C. Dennett and David Chalmers**. New York: The New York Review of Books, 1997.

STEIN, E. **Introdução ao pensamento de Heidegger**. Rio Grande do Sul: ITHACA, 1966.

SILVEIRA, P. **Ética e Inteligência Artificial: Da possibilidade filosófica de Agentes Morais Artificiais**. Porto Alegre, RS: Editora FI, 2021.

TURING, A. M. **Computing Machinery and Intelligence**. *Mind*, v. 49, p. 433-460, 1950.

