

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS HUMANAS, LETRAS E ARTES
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA
DOUTORADO EM FILOSOFIA

***INTELIGÊNCIA ARTIFICIAL:
Um perguntar pelo homem?***

JOSÉ ATÍLIO PIRES DA SILVEIRA

JOÃO PESSOA - PB
2017

JOSÉ ATÍLIO PIRES DA SILVEIRA

***INTELIGÊNCIA ARTIFICIAL:
Um perguntar pelo homem?***

Trabalho apresentado ao Programa de Pós-Graduação
em Filosofia da Universidade Federal da Paraíba para
a obtenção do título de doutor em filosofia.

Orientador: Prof. Dr. EDMÍLSON ALVES DE
AZEVÊDO

JOÃO PESSOA - PB
2017

Catalogação na publicação
Setor de Catalogação e Classificação

S587i Silveira, José Atilio Pires da.

Inteligência artificial: um perguntar pelo homem? / José Atilio Pires da Silveira.
– João Pessoa, 2017.
219 f. : il.

Orientador: Edmílson Alves de Azevêdo.
Tese (Doutorado) – UFPB/CCHLA

1. Filosofia. 2. Inteligência Artificial - projeto IA. 3. Homem - inteligência e
artifício. 4. Tecnologia - homem. 5. Volta estranha - noção. I. Título.

UFPB/BC

CDU -

JOSÉ ATÍLIO PIRES DA SILVEIRA

INTELIGÊNCIA ARTIFICIAL:
Um perguntar pelo homem?

BANCA EXAMINADORA

Prof. Dr. Edmílson Alves de Azevêdo (UFPB)

Prof. Dr. André Leclerc (UnB)

Prof. Dr. Garibaldi Sarmento Monteiro (UFPB)

Prof. Dr. Gilfranco Lucena dos Santos (UFPB)

Prof. Dr. Iraquitan de Oliveira Caminha (UFPB)

JOÃO PESSOA -PB
2017

Ao meu irmão Celso Romário, o primeiro filósofo que conheci e *in memoriam* de nossos pais, avós, tias e tios.

AGRADECIMENTOS

À minha companheira amada Jane que sempre me apoia de forma crítica, pela leitura, correção ortográfica e gramatical do texto e à nossa querida Leona, filha do Pingo, cujo aconchego aquece nossos melhores sentimentos e emoções.

Ao meu irmão Antônio Laffayete, Áurea, Carol, Ciça, Pedro, Cali e Carmela que mesmo na distância me alimentam e fortalecem com o seu amor.

À vó Celestina, às tias Nice, Elcira, Sebastiana, Thereza e Maria Loyde que sempre me incentivaram e auxiliaram nesse caminho de conhecimento.

À minha querida “família trapo”, Henrique, Myrna, Yuri, Moana e Cosmo que tornam a Paraíba o lugar dos meus sonhos com a sua presença amável.

A Marília, Sérgio e Tereza que a vida nos aproximou pelo amor e carinho por Yuri.

Aos amigos e colegas Mateus, Fernando, Luciano e Antônio Fábio cujas conversas nas reuniões da *Revista Problemata* trouxeram valiosas contribuições para este trabalho.

Ao meu cunhado Ibanez, que nesses anos de doutorado manteve com zelo, carinho e dedicação o nosso cantinho em Toledo.

Ao meu orientador Prof. Edmílson, que com grande paciência e dedicação soube me incentivar na realização desse trabalho.

A Beth, Leon, Júlia e Preto cujo carinho me acompanham desde Porto Alegre.

A Luciano, Sânia e Daniel pela amizade e especial consideração.

A Márcio e Dina pelo compartilhamento de sua profunda sensibilidade pelo outro.

A d. Mariaugusta, minha amiga dos almoços e conversas na padaria sobre o momento pelo qual passa o nosso país, sua vida tão rica e vigorosa, uma lembrança que guardarei para sempre com carinho.

Ao meu amigo Julio, cuja inteligência e esforço sempre me inspiraram.

A s. Nelson que com suas divertidas histórias da vida em condomínio me ligam mais a este lugar.

A d. Penha que nos recebeu sempre de braços abertos com a simpatia e leveza que só o povo do Nordeste possui.

A d. Ilka cujo carinho e acolhimento nos farão sentir saudades dos bons anos vividos na Paraíba.

O pensamento difuso deve ser considerado não deplorável se ele torna possível a solução de problemas que são complexos demais para uma análise precisa.

Lotfali Askar-Zadeh

INTELIGÊNCIA ARTIFICIAL: Um perguntar pelo Homem?

SUMÁRIO

Introdução.....	1
Cap. 1 Da busca pela regularidade ao funcionamento do cérebro.....	8
1.1 A longa gestação do projeto de IA. Da Grécia até a primeira metade do século XX.....	8
1.2 O nascimento do projeto de IA. Alan Turing e o esforço de guerra.....	23
1.3 Os pioneiros da IA.....	26
1.4 Algumas das principais vertentes e questões da IA..	31
1.5 Os novos rumos do projeto de IA.....	43
Cap. 2 Inteligência Artificial: ilusão ou possibilidade?.....	47
2.1 O caráter divergente das pesquisas em IA.....	48
2.2 A influência da dicotomia e sua origem cartesiana..	50
2.3 A proposta turingeana de superação da dicotomia...	53
2.4 A crítica de John Searle ao pragmatismo de Turing.	59
2.5 A resposta de Daniel Dennett a John Searle.....	64
2.6 Bennett e Hacker: Uma proposta de trabalho convergente.....	71
2.7 Colocação do problema: o projeto de IA é admissível?.....	75
Cap. 3 Como tratar o problema do projeto de IA.....	83
3.1 A dicotomia como entrave do debate filosófico.....	85
3.2 A estratificação como modelo de compreensão.....	99
3.3 A noção de <i>volta estranha</i> como superação da dicotomia.....	112
Cap. 4 Um triângulo de impossibilidade: <i>homem, inteligência e artifício</i>	127
4.1 O Conceito de homem.....	131
4.2 Algumas definições de homem.....	138
4.3 O animal simbólico.....	139

4.4 Um ser que produz tecnologia.....	143
4.5 Um <i>ser propenso à autodestruição</i> e o medo da tecnologia.....	162
4.6 O conceito de <i>inteligência</i>	166
4.7 A relação entre o cérebro humano e a inteligência...	169
4.8 Algumas definições de <i>inteligência</i>	176
4.9 O <i>saber calcular</i>	176
4.10 O saber elaborar uma obra artística.....	183
4.11 O conceito de <i>artifício</i>	185
4.12 Uma <i>volta estranha</i>	192
Considerações Finais.....	204
Bibliografia.....	207

LISTA DE FIGURAS

Figura 1	<i>Kurzweil, 2007</i>	p. 77
Figura 2	<i>Drawing Hands</i> de M.C. Escher (1948)	p. 112
Figura 3	Penrose (1998)	p. 114
Figura 4	<i>Ascending and Descending</i> de M.C. Escher (1960)	p. 119
Figura 5	<i>Systematic Study</i> de M.C. Escher (1936)	p. 194
Figura 6	Ilustração extraída da internet	p. 195
Figura 7	<i>Relativity</i> de M.C. Escher	p. 196
Figura 8	Triângulo de Impossibilidade	p. 197
Figura 9	Penrose (1999)	p. 198
Figura 10	Triângulo <i>Homem, Inteligência e Artifício</i>	p. 201

RESUMO

Esta tese tem por objetivo debater a viabilidade do projeto de IA na sua versão forte. Nossa hipótese de trabalho pergunta se o projeto de IA não corre o mesmo risco que a Alquimia, qual seja, de almejar algo considerado impossível. Para tratar dessa questão tomamos como embasamento filosófico a crítica de John Langshaw Austin à dicotomia, a *concepção estratificada do tempo* de Reinhart Koselleck e a noção de *volta estranha* apresentada por Douglas Hofstadter. Pressupomos que o debate em torno dos objetivos do projeto de IA se encontra fortemente influenciado pela forma dicotômica de arranjo de ideias. Tal modelação produz um impasse no debate, pois os interlocutores que dele participam não cogitam a possibilidade de que o seu parceiro de diálogo possa ter um mínimo de razão. Nos propomos, aqui, a apresentar uma proposta que consiga superar os limites impostos pela forma dicotômica e sustentar que o projeto de IA é, ao menos, uma proposta admissível.

Palavras-chave: Inteligência Artificial, Homem, Artifício, Tecnologia, Volta Estranha

ABSTRACT

This thesis aims to discuss the feasibility of the AI - Artificial Intelligence - in its strong version. Our working hypothesis asks if the AI project does not run the same risk as Alchemy, that is, to aim for something considered impossible. To deal with this question we take as philosophical basis John Langshaw Austin's critique of the dichotomy, the stratified conception of Reinhart Koselleck's time, and Douglas Hofstadter's notion of a strange turn. We assume that the debate about the objectives of the IA project is strongly influenced by the dichotomous form of arrangement of ideas. Such modeling produces a deadlock in the debate, since the participants who participate do not consider the possibility that their dialogue partner can have a minimum of reason. We propose here to present a proposal that can overcome the limits imposed by the dichotomous form and maintain that the IA project is at least an admissible proposal.

Keywords: Artificial Intelligence, Man, Artifice, Technology, Strange Tour

INTRODUÇÃO

A presente tese possui como principal temática o problema da efetividade do projeto de inteligência artificial (IA). Nela perguntamos se o propósito de alcançar inteligência semelhante à humana de modo artificial é, ao menos, admissível.

Um dos objetivos centrais deste trabalho é destacar o papel que a reflexão filosófica ocupa nesse cenário tão marcadamente científico e tecnológico. Se almejamos superar o impasse no qual se encontra o debate em torno do projeto de IA, julgamos ser a transposição de um obstáculo de natureza metafísica, como a dicotomia, uma das tarefas indispensáveis ao esforço de atingir-se tal superação. É inegável a relevância e as influências que a ciência e a tecnologia exercem nesse contexto. Contudo, acreditamos que o horizonte desse debate também seja constituído por elementos e concepções que, embora não sejam de natureza científica, nem tecnológica, possuem o mesmo grau de importância que esses no esforço de produzir-se uma resposta satisfatória ao problema. A problematização do projeto de IA reveste-se de importância filosófica porque, também, nos induz a refletir sobre o que seja o próprio homem e a especular sobre os limites e capacidades do nosso conhecimento.

A princípio, talvez devêssemos delimitar essa reflexão nos atendo tão somente às questões que dizem respeito às origens e significados dos conceitos de *inteligência, artifício, pensamento, consciência*, dentre outros. Contudo, pensamos que essa abordagem do problema restringiria muito o escopo desta reflexão, visto que as questões suscitadas pela intenção de se produzir inteligência artificialmente não dizem respeito exclusivamente ao âmbito de uma disciplina, de uma ciência ou de uma área do conhecimento. Tratando-se, sobretudo, de um esforço conjunto de todas elas.

Não obstante, isto não significa dizer que iremos abrir mão da tarefa de perguntar pela origem dos significados dos termos empregados na elaboração das teorias que buscam explicar o fenômeno do pensamento e da inteligência, nem deixaremos de lhe dar relevância nesse trabalho.

A publicação do artigo de Alan Turing, no ano de 1950, intitulada *Computação e Inteligência* (1996) no qual descreve a arquitetura do que ele chamou de *máquina-de-estado-discreto*, máquina com a capacidade de realizar atividades consideradas, até então, exclusivas do pensar humano, promoveu o surgimento de uma nova área de interesse do

conhecimento denominada “inteligência artificial”. No entanto, esse termo só foi empregado pela primeira vez seis anos após a publicação de Turing quando Marvin Minsky, Allen Newell, Herbert Simon, John McCarthy e outros pesquisadores reuniram-se, no ano de 1956, em New Hampshire/Estados Unidos, para debater sobre a possibilidade de construção artificial de máquinas inteligentes.

Para aferir a semelhança de capacidade de pensar da máquina-de-estado-discreto com a humana, Turing propôs como teste o *jogo da imitação*, em que um dos jogadores participantes seria uma dessas máquinas. Dois objetivos foram atribuídos a essa nova área de investigação, sendo um bem mais audacioso do que o outro. O objetivo mais arrojado seria o de construção artificial de um *novo homem*, dado que almejaria reproduzir artificialmente uma característica natural que distingue o homem de todas as demais coisas; o mais modesto visaria somente à construção de um dispositivo capaz de substituir o homem no desenvolvimento de atividades executadas exclusivamente por este, antes do advento da nova tecnologia da máquina-de-estado-discreto. Esses dois objetivos distinguem o projeto de IA forte do projeto de IA fraca. Quando nos referimos, aqui, ao projeto de IA, nos reportamos à sua concepção forte, pois entendemos que a concepção fraca de IA não enfrenta as mesmas críticas que a IA forte, visto que aquela concebe a máquina-de-estado-discreto como uma mera bancada de testes no esforço de compreensão sobre como se dá a atividade de pensamento humana.

A declaração da IA forte de pretender alcançar um objetivo tão ambicioso quanto o da construção artificial de algo que compartilharia com o homem a capacidade que distingue este de todas as demais coisas é provocativa e perturbadora. Esses efeitos resultam do fato da ideia de construir-se um *novo homem* ir de encontro a um conjunto de crenças e convicções de naturezas filosófica, religiosa e científica idealizadas ao longo da história do pensamento e por provocarem indagações advindas de diferentes áreas de interesse da racionalidade humana, as quais perguntam sobre os limites do conhecimento, a possibilidade de nos definirmos satisfatoriamente, a dependência da tecnologia no processo de evolução humana e a negação da ideia de que a inteligência humana seja o resultado de um ato de criação divina.

Embora não envolva dificuldade compreender-se em que consista o propósito último do projeto de IA, seu entendimento, contudo, não é fácil de ser assimilado. Isso se deve a uma série de razões que vão desde as imprecisões na definição dos termos “inteligência”, “artifício” e “homem”, por não possuírem um sentido unívoco; até o próprio modo de organização das ideias e conceitos envolvidos na tarefa de esclarecer questões relacionadas ao objetivo de construção do “novo homem”. Essas condições propiciaram o surgimento de diferentes abordagens e propósitos da reflexão em torno do objetivo do projeto de IA atribuindo-lhe, com isso, a característica de um conjunto de esforços muitas vezes divergentes.

Mesmo que diferentes áreas do conhecimento concorram no esforço de desenvolvimento da tecnologia necessária para a obtenção de IA, isso não possibilitou, ainda, uma demarcação epistêmica clara e, consequentemente, capaz de garantir uma convergência de esforços. A falta de clareza do objeto a ser compreendido para, a seguir, ser reproduzido artificialmente é uma das dificuldades que nos indica o estágio inicial de desenvolvimento no qual se encontra essa área da atividade humana, pois, a nosso ver, ainda permanece sem solução a questão de sabermos sobre a quem devemos atribuir a responsabilidade do agir inteligente humano: à mente, a uma faculdade pensante, a um órgão pensante ou a uma estrutura consciente? Como podemos constatar no debate em torno da pergunta pela natureza das operações de pensamento e das atividades do intelecto, chamadas também de *estados mentais*. Seriam elas consequências das interações entre os componentes físicos que constituem os nossos corpos ou, são fenômenos de outra natureza, não sendo físicos, nem materiais?

Por ir contra várias concepções consideradas ao menos aceitáveis em nossa época, o objetivo último do projeto de IA é recebido com bastante desconfiança e ressalvas. Não é a primeira vez, na história do pensamento humano, que uma ideia é apresentada num contexto que lhe é adverso. Contribui para tanto o aparente exagero de propósitos, que se destaca como a característica comum às ideias assim recepcionadas, pois os resultados indicados por elas não são considerados realizáveis. Mas, mesmo que não julguemos tais resultados impossíveis de serem alcançados, as suas realizações são de tamanha complexidade que nos induz a considerá-los praticamente como tais.

O propósito de construção artificial de algo que presumimos ser atributo exclusivo da natureza não é uma ideia original do projeto de IA, a Alquimia também se propunha a algo semelhante. Essa prática antiga não possuía somente como objetivos a transmutação de metais e a obtenção do elixir da longa vida. Ela pretendia, também, construir vida humana de forma artificial, o *Homunculi*, ideia essa diretamente relacionada à concepção judaica de *Golem*. Podemos verificar que há uma semelhança de características, tanto de objetivos gerais quanto de propósitos específicos, entre o projeto de IA e a Alquimia, se levarmos em consideração essa construção de raciocínio. Assim sendo, cabe-nos colocar uma questão, a saber: o projeto de IA não corre o risco de chegar à mesma situação na qual se encontra a Alquimia? O projeto de IA não promete mais do que seríamos capazes de realizar? Não estaríamos diante de uma pretensão de caráter meramente ilusório?

Certamente este não é um problema que demande uma forma simples de tratamento. Ele envolve questões que são provenientes de diferentes áreas do conhecimento e de distintas necessidades de fundamentação, além dos fatos como os apontados acima.

Acreditamos que muitas das dificuldades apontadas pelos críticos do projeto de IA devem-se a um conjunto de situações e conceitos mal definidos e incompreendidos nos seus usos que merecem ser examinados, somente assim evitaremos os enganos provocados por falsos problemas, como nos advertiu John Langshaw Austin em *Sentido e Percepção* (1993).

Todavia, não seguiremos as advertências de Austin como a um programa de pesquisas que seria empregado ao longo de todo este trabalho, mesmo que em boa parte ele seja utilizado como um dos modelos de abordagem dos problemas aqui tratados. Concordamos com sua afirmação de que muitos fatos, situações, palavras e seus usos merecem ser melhor esclarecidos quando está em discussão o projeto de IA e seu objetivo principal. Talvez o problema maior não esteja na parte prática do projeto, mas no conceitual disponível para a elaboração das explicações dos avanços tecnológicos alcançados nesta área. Por outro lado, a falta de clareza conceitual induz àqueles que desenvolvem tecnologia na área de IA, muitas vezes, a acreditar que uma determinada ação desempenhada por um computador digital se trata de um agir genuinamente inteligente, quando na verdade é apenas uma imitação deste tipo de ação.

A produção e reconhecimento de inteligência artificial não dependem somente de avanços científico/tecnológicos, pois acreditamos que não podemos prescindir de avanços no campo da linguagem e seus usos como condição de possibilidade para a superação das dificuldades que nos impedem de alcançar tal tipo de reconhecimento.

Pressupomos que a formulação de uma resposta satisfatória à questão suscitada pelos objetivos da IA é difícil e complexa, e depende, dentre tantas outras coisas, da compreensão do que seja o *homem*, a *inteligência*, a *consciência*, a *vontade* e de como funciona a linguagem comum. Não arriscaríamos assumir o ponto de vista de que seja possível chegar-se, um dia, a uma definição de *homem* que esgote todas as possibilidades de significação do mesmo. No entanto, não deixamos de mencionar como uma das razões que sustenta a suspeição de que não é possível construir-se um objeto dotado de inteligência como a humana, a falta de clareza do que seja o próprio *homem*.

Alguns pesquisadores acreditam que uma vez removida a barreira da falta de informação sobre o funcionamento do cérebro estará aberto o caminho para o avanço das pesquisas que buscam desvendar aquilo que até agora era considerado algo inexpugnável, a intimidade dos nossos pensamentos. Parece ser uma questão de tempo o desvelamento do diálogo da alma consigo própria, mesmo sem o seu consentimento e testemunho.

Entretanto, há um conjunto de questionamentos que precisam ser tratados a fim de avaliarmos com maior rigor e profundidade suas possíveis consequências. Pode ser que estejamos no limiar de uma nova virada de pensamento na história do conhecimento e, consequentemente, do homem. Pode ser que a busca por inteligência artificial nos permita mudar a maneira como pensamos nós mesmos. Pode ser que esta busca dissipe muitos problemas e entendimentos no campo das ciências empíricas e experimentais.

A primeira dificuldade a ser superada na tentativa de elaborarmos uma resposta às hipóteses de trabalho acima colocada é de natureza metafísica. Como veremos, o debate acerca do projeto de IA é bastante influenciado por uma maneira específica de organização de ideias, o que John Langshaw Austin em *Sentido e Percepção* (1993) chamou de *Gleichschaltung*, a dicotomia.

A necessidade de reconhecermos a forte influência que o modo de pensar dicotômico exerce sobre o debate em torno dos objetivos do projeto de IA é condição *sine*

qua non para nos afastarmos desse vórtice metafísico, uma situação da qual parece não haver saída. A crítica austiniana ao modo de pensar dicotômico é o ponto do qual partimos na tentativa de elaboração de uma resposta ao problema do projeto de IA. A força de coesão que une os conceitos que formam um par dicotômico deve ser anulada, pois os fenômenos mostram-se muito mais complexos do que esse modo de arranjo de ideias é capaz de representar.

Apesar de Austin ter criticado de maneira vigorosa a dicotomia como forma de organização de ideias implicadas na tarefa de elaborar uma explicação sobre o que é o mundo e aquilo que nele se dá, ele não chegou à apresentação de um modelo alternativo de organização das mesmas.

Na obra de Reinhart Koselleck *Estratos do Tempo* (2014), encontramos um exemplo de organização de ideias que rompe com o modelo dicotômico por permitir a relação entre mais do que dois conceitos numa dada representação de mundo. O sistema estratificado de Koselleck não sofre as limitações impostas a um modelo que se fundamenta num antagonismo dicotômico entre ideias, o que o torna apto a produzir representações mais complexas do mundo, ao poder incorporar um número maior de diferentes ideias nessas representações.

No entanto, embora o sistema de Koselleck possa ser considerado como modelo alternativo ao criticado por Austin, ele ainda não nos permite elaborar uma resposta satisfatória ao problema do projeto de IA, pois ainda é tributário de um arranjo de natureza dicotômica, por pressupor que um determinado conjunto de ideias deva ser arranjado de acordo com os critérios de anterioridade e posterioridade. Necessitamos de um sistema que não seja tão marcadamente espacial.

Encontramos nas obras de Roger Penrose, *O grande, o pequeno e a mente humana* (1998) e na de Douglas Hofstadter, *Gödel, Escher e Bach. Um entrelaçamento de gênios brilhantes* (2001), exemplos de um modelo com as características que possibilitam a elaboração de uma resposta ao problema do projeto de IA. Esse modelo de representação do mundo nos permite relacionar livremente conceitos, sem que seja necessário levarmos em consideração critérios dualistas como os de anterioridade e posterioridade, por exemplo.

A noção de “volta estranha”, Hofstadter (2001), presente no pensamento matemático de Gödel e na concepção artística de Escher é empregada na análise que fazemos da relação entre os diferentes significados dos conceitos de “homem”, “artifício” e “inteligência”, a fim de sustentar a tese de que o objetivo forte do projeto de IA é admissível, não podendo ser considerado uma promessa de conhecimento semelhante a que foi a Alquimia no passado.

CAPÍTULO 1

DA BUSCA PELA *REGULARIDADE* AO FUNCIONAMENTO DO CÉREBRO

Esta tese pretende apresentar uma resposta à questão acerca da admissibilidade do projeto de IA. Para tanto, primeiramente faremos uma breve exposição da história deste projeto tendo em vista colocar em evidência suas diferentes origens, problemas e caminhos de investigação. Com isso, procuramos ressaltar a importância de que a pergunta sobre a possibilidade de obtenção de inteligência artificial só poderá ser respondida adequadamente se levarmos em consideração seu caráter multi-disciplinar.

Os gregos determinaram a configuração do problema da IA ao conceberem o homem como um ser constituído por um corpo material e uma porção imaterial.

A longa gestação do projeto de IA.: dos gregos até a primeira metade do séc. XX

Conforme Hubert Lederer Dreyfus (1929-2017), em *O que os computadores não podem fazer* (1975), afirma que a discussão sobre os objetivos do projeto de IA pode ter iniciado com os gregos ao inventarem a lógica e a geometria. Esse debate está latente na cultura humana desde o momento em que o homem alimentou o sonho de ter seu trabalho substituído e executado por uma máquina.

É pouco provável que os gregos tivessem pensado na possibilidade de construir-se uma máquina capaz de realizar as operações de pensamento que um ser humano realiza, mas pensaram na hipótese de os seres humanos serem estruturas biológicas capazes de realizar operações de pensamento padronizadas, um dos pressupostos básicos presente desde os primórdios do projeto de IA.

[...] a história da inteligência artificial bem poderia ter início em torno de 450 a.C. quando (segundo Platão) Sócrates exige de *Euthyphro*, um colega ateniense que, em nome da piedade, está para entregar o próprio pai, acusado de homicídio: - Quero saber que característica da piedade é essa que torna todas as ações pias... para que eu possa voltar-me para ela e usá-la como um padrão pelo qual julgue suas ações e a de outros homens. (DREYFUS, 1975, p. 17)

No período antigo, a literatura grega nos apresenta muitos exemplos de autômatos e mecanismos complexos que vão desde a coruja de Minerva até as estátuas de *Hefesto*, que, segundo a mitologia grega, fez um gigante para proteger a ilha de Creta. A capacidade grega de não só imaginar, mas, também, de construir autômatos e máquinas complexas, pode-se ver no que restou de uma antiga calculadora mecânica, possivelmente empregada para realizar cálculos astronômicos, chamada *Antikythera*.

A ideia de se construir um homem mecânico, capaz de realizar as tarefas humanas, era bastante difundida na Europa do séc. XVIII. O fascínio sobre as plateias que a ideia de um mecanismo que funcionaria sem a interferência do ser humano provoca tornou notório o barão Wolfgang von Kempelen, que enganava as cortes europeias com a *máquina de jogar xadrez*. Na verdade, não se tratava de um verdadeiro autômato, pois ele era operado por um ser humano que ficava escondido dentro do dispositivo. Porém, muitos relojoeiros desse período inventaram autômatos e mecanismos complexos empregados na celebração de cultos religiosos e como brinquedos para a diversão de crianças e reis.

A suposição, levantada pelo personagem de Platão, de que o comportamento humano é padronizável pode ser considerada como o momento inaugural do pensamento que nos leva hoje a especular sobre os objetivos do projeto de IA. Tal suposição serve de fundamento para muitos pesquisadores envolvidos no desenvolvimento do projeto de IA ao conceberem o pensamento como um fenômeno que exibe um modo de desenvolvimento que pressupõe uma organização, que pode ser conhecida e reproduzida a partir do conhecimento que dela poderemos ter. De acordo com essa concepção, a hipótese de que o objetivo forte do projeto de IA não é um exagero de expectativas é correta.

A preocupação platônica em julgar corretamente uma ação visava evitar que alguém fosse condenado ou condecorado por algo que não praticou e, por consequência, não mereceria. A orientação para o desenvolvimento desse julgamento consiste no reconhecimento de que uma determinada ação seria realizada por quem quer que fosse o seu agente particular. A correspondência ao padrão determinaria o reconhecimento justo da ação praticada. O procedimento de universalização apresenta-se como modelo de bem

julgar. Uma ação qualquer que vier a ser realizada não pode ser considerada condenável ou louvável se a mesma, dadas as circunstâncias, for executada sempre e da mesma forma por qualquer outro agente. Parece haver, de acordo com a preocupação socrática, uma tríade de valores que podem ser atribuídos a uma determinada ação praticada. Ela pode ser julgada condenável, louvável ou trivial.

Entretanto, não são somente questões acerca de como devemos julgar que estão presentes na indagação socrática, também está presente

o que os teóricos modernos do computador chamariam de procedimento útil (*effective procedure*) – um conjunto de regras que nos diz exatamente, de momento a momento, como nos comportarmos. (DREYFUS, 1975, p. 17)

Conforme Dreyfus (1975), embora Platão não fosse um cibernetico, porque “... procurava mais critérios semânticos do que sintáticos”, ele não deixou de buscar compreender o comportamento humano através de um conjunto de regras que o determinem.

Consoante Platão, todo conhecimento deve ser expresso em definições explícitas que qualquer um possa aplicar. Se alguém não pudesse expressar seu “know how” em termo de tais instruções explícitas – se seu saber *como* (*knowing how*) não se pudesse converter em saber *que* – isso não seria conhecimento, mas simples crença. (DREYFUS, 1975, p. 17)

A ideia platônica de que somente demonstramos conhecimento quando somos capazes de apontar as razões de determinado acontecer, por meio da enumeração das regras que o determinam, pode ser considerada como um dos gametas que tornou possível o surgimento do projeto de IA. Uma vez que ela pressupõe como modo adequado e rigoroso o procedimento de formalização, isto é, a busca pela identificação dos componentes e suas relações num determinado arranjo. Platão via nessa forma de proceder a garantia de universalização do conhecimento, deixando de lado aspectos que dificultassem essa condição. De acordo com Dreyfus (1975)

Para o projeto platônico atingir plenamente uma nova etapa, é necessário que todo apelo à intuição e ao discernimento seja eliminado. Posto que Galileu descobriu que se poderia encontrar um puro formalismo para se descrever o movimento físico ignorando-se as qualidades secundárias e considerações teleológicas, poder-se-ia supor que um Galileu do comportamento humano teria êxito na redução de todas as considerações semânticas (apelo aos significados) e às técnicas de manipulação sintática (formal). (DREYFUS, 1975, p. 19)

Essa perspectiva de produção de conhecimento vê o procedimento de formalização dos fenômenos como condição necessária a ser efetivada, a fim de que possamos compreender a realidade como ela é. Descobrir quais sejam as regras que configuram os diferentes acontecimentos é uma tarefa indispensável para que seja levado a bom termo o propósito de compreensão do mundo. Tal perspectiva de tratamento da questão orienta as investigações de muitos grupos de pesquisadores envolvidos na criação de inteligência artificial.

A crença de que tal formalização do conhecimento deve ser possível logo passou a dominar o pensamento ocidental. Ela já expressava uma exigência básica, moral e intelectual e o êxito da ciência da física pareceu sensibilizar os filósofos do século dezesseis, como ainda parece sugerir a pensadores como Minsky que a exigência poderia ser satisfeita. (DREYFUS, 1975, p. 19)

Contudo, acreditamos que a descoberta de leis que governam os fenômenos não seja suficiente para reconhecer-se a criação de inteligência artificial, pois julgamos ser necessário, também, um desenvolvimento concomitante da linguagem que a torne capaz de dar significado à novas situações e usos de palavras.

Embora seja mais perceptível a característica de natureza científico-tecnológica do projeto de IA, seu surgimento, na verdade, dá-se em duas etapas. A primeira etapa é filosófica; a segunda, científico-tecnológica. Tais etapas possuem abordagens e contextos históricos distintos, o que não significa dizer que não estejam profundamente relacionadas uma com a outra.

Concordamos com Hubert Dreyfus (1975) que o pensamento sobre a possibilidade de IA já esteja latente desde as primeiras reflexões filosóficas do período clássico, porém, é no pensamento moderno de Descartes que esta discussão adquire os contornos e os

limites que apresenta na atualidade. Na modernidade, a ideia de que o agir humano pudesse ser algoritmizável era considerada altamente relevante para a maioria dos filósofos desse período. A busca pelas condições de elaboração de uma ética universal é um propósito que caracteriza a reflexão de muitos filósofos modernos e, para tanto, é indispensável uma descrição da ação humana com o maior detalhamento possível, condição indispensável para chegar-se à dedução dos princípios, leis e regras que determinam seu agir.

O cenário do debate envolvendo questões profundamente relacionadas aos objetivos do projeto de IA é configurado e influenciado, ainda hoje, pela tese cartesiana do *dualismo de substâncias*. De acordo com essa tese, nossos pensamentos se desenvolvem num contexto semelhante ao de uma peça teatral da qual somos, além de espectadores, o próprio palco, cenário e atores que a encenam. No entanto, julgamos como mais pertinente e apropriado entendermos tal tese como um dispositivo de contenção, uma camisa de força metafísica imposta aos atores em cena, pois ao exercerem os seus papéis eles ficam impedidos de realizar uma grande gama de ações e performances, já que os mesmos se encontram bem definidos e determinados.

No *dualismo de substâncias* o homem é concebido como um ser que possui dois tipos de substâncias que não se comunicam e nem interferem uma na outra. O que diferencia o homem das demais coisas é o fato dele ser portador de uma substância que nenhuma outra criatura possui, a que lhe dá a capacidade de pensar, a *res cogitans*.

Embora Descartes não tenha desprezado a substância material, a *res extensa*, como afirma Paul Churchland (*The Ontological Problem*, 2013, posição 61, *kindle*), ele acredita, contudo, que o princípio de certeza se encontra na *res cogitans*. Seguindo, assim, o pensamento clássico de que devemos buscar a explicação do erro em nós mesmos.

Descartes baseou seu sistema filosófico no leito de pedras da realidade de seus próprios pensamentos. Para nossos propósitos, o aspecto crucial do ponto de vista cartesiano sobre o mundo será a noção de que há uma clara distinção entre o corpo físico (inclusive o cérebro) de um lado e a mente não-física de outro. Este suposto dualismo mente/corpo tem desempenhado um papel importante no pensamento sobre a atividade intelectual desde Descartes. Os filósofos, de fato, têm escrito críticas longas e detalhadas sobre a maneira cartesiana de encarar o mundo. Certamente o tipo de disjunção entre mente e corpo implícita na

estrutura cartesiana não se ajusta muito bem ao que sabemos agora sobre o cérebro. Não obstante, há a impressão de que algo que é parecido com o procedimento de Descartes permanece válido para a questão da consciência humana. (TREFIL, 1999, p. 159)

Concordamos com essa última impressão de Trefil, consideramos que a influência cartesiana sobre a reflexão em torno do projeto de IA é bastante forte, mesmo nos dias de hoje. A concepção cartesiana é tributária de uma forma de lidar com ideias que está presente no cenário da reflexão filosófica desde o seu início, ela é uma instanciação da dicotomia. Entendemos que a dicotomia é um problema que prejudica o avanço e a compreensão em torno da possibilidade e reconhecimento de inteligência artificial. Procuraremos mostrar, no capítulo 5, que é possível uma superação do modelo dicotômico de tratamento das questões envolvendo os objetivos do projeto de IA.

Immanuel Kant (1724-1804) na *Lógica* (1992) também coloca no horizonte de sua reflexão a possibilidade de compreender-se o agir humano como consequência de um ordenamento *a priori* das nossas ações. Ele afirma que todo acontecer na natureza nada mais é do que um seguimento de regras. Tão convicto está dessa certeza que afirma que se algo se apresentar a nós como caótico isto significa dizer que desconhecemos as regras que determinam este acontecer, não que ele seja produzido pela ausência de regras. Para Kant, até mesmo o pensamento humano está sujeito a regras que determinam o seu uso, embora não seja fácil descobri-las.

O exercício de nossas faculdades se verifica conforme leis, as quais, desde logo nos ajustamos sem ter consciência disso, até o ponto de que chegamos inconscientemente ao conhecimento das mesmas pelos fatos da experiência e pelo uso contínuo das próprias faculdades. (KANT, 1992, p. 29).

A maneira como concebemos o conhecimento científico foi fortemente influenciada pela compreensão de Kant. Ele difundiu a ideia de que o conhecimento humano deve ser objetivo, universal e determinista. Essas três características dariam à racionalidade científica as condições de realizar algo tão almejado pelo homem, a capacidade de fazer antecipações e previsões acerca dos fenômenos.

Entretanto, com o surgimento da Teoria da Relatividade de Einstein e da Mecânica Quântica, a forma de compreender a atividade científica foi profundamente remodelada. Quando Werner Heisenberg (1901-1976) apresentou o *princípio da incerteza*, cujo enunciado afirma que não é possível saber-se a posição e a velocidade exatas de uma partícula atômica, mas somente indicá-las de maneira probabilística, o modelo clássico das ciências sofreu grande impacto, pois o paradigma da determinação e da certeza foi substituído pelo da probabilidade.

A incerteza, a qual o princípio de Heisenberg se refere, deve-se ao fato de que ao observarmos um objeto, nós alteramos as suas condições. Ao contrário do que ocorre com o desenvolvimento da argumentação que segue o modelo kantiano, em que o passo seguinte na cadeia inferencial está claramente demarcado na nova abordagem fundamentada no *princípio da incerteza*. Não há um e somente um caminho a seguir na cadeia inferencial, pelo contrário, abre-se um leque de possibilidades dedutivas relevantes ao conjunto de premissas dadas.

Em *Physics and beyond* (1971), Heisenberg narra um debate ocorrido com a filósofa e matemática Grete Hermann (1901-1984) em que podemos ver o contraste entre os dois modos de conceber a atividade científica. Hermann era aluna de Leonard Nelson (1882-1927), professor de filosofia que ensinava em Göttingen e partilhava a mesma posição neo-kantiana com o filósofo e naturalista Jakob Friederich Fries (1773-1843). Nesse debate Hermann defende a tese kantiana de que as leis causais são inabaláveis, pois elas são o fundamento último das explicações científicas.

Um dos pressupostos da escola de Fries e também do círculo de Nelson, era que toda questão filosófica deveria ser tratada com o rigor normalmente reservado à matemática moderna. Foi seguindo essa rigorosa abordagem que Grete Hermann acreditou poder provar que a lei causal – como na forma dada por Kant – era inabalável. (HEISENBERG, 1971, p. 117, tradução nossa)

De acordo com a argumentação de Hermann, sem a lei causal nossas impressões sensíveis seriam nada mais do que sensações subjetivas. Sem nenhum correlato objetivo, devendo essa regra (causa e efeito) ser pressuposta se pretendemos fazer ciência, pois é ela quem torna objetivas nossas observações, possibilitando, assim, a verificação das

experiências. Após essas considerações iniciais ela formula uma questão dirigida aos cientistas que participavam do debate.

A lei causal é uma ferramenta mental com a qual tentamos incorporar a matéria prima das nossas impressões sensíveis na nossa experiência, somente assim poderemos entender os objetos da ciência natural. Sendo esse o caso, como pode a mecânica quântica pensar ser possível abrir mão da lei causal e, ainda assim, permanecer sendo um ramo da ciência? (HEISENBERG, 1971, p. 117, tradução nossa)

Dentro da tradição kantiana, Hermann sustenta que a possibilidade de se alcançar rigor na atividade científica é encadeando um fato ao outro segundo numa relação de causa e efeito. A resposta dada por Heisenberg demonstra a profunda diferença entre o modelo kantiano de fazer ciência e o da mecânica quântica.

Sempre que tentamos deduzir leis de nosso estudo do fenômeno atômico, descobrimos que não correlacionamos mais processos objetivos no espaço e no tempo, mas somente situações observacionais. Somente assim podemos derivar leis empíricas. Os símbolos matemáticos com os quais nós descrevemos tais situações observacionais representam possibilidades, aos invés de fatos. (HEISENBERG, 1971, p. 122, tradução nossa)

O modelo de Kant segue na esteira do personagem platônico que vê na regularidade dos acontecimentos a única possibilidade e meio adequados à sua compreensão. No entanto, a atividade consciente parece ser algo especial, já que ela é desempenhada de maneira distinta quando comparada a outras atividades realizadas por nós. De acordo com John Searle (1932), essa diferença já nos é apontada por Kant na *Crítica da Razão Pura* (2001) quando ele nos mostra como a consciência se relaciona com as condições de possibilidade do conhecimento. Embora sejamos capazes de descrever com relativa facilidade os objetos que nos cercam, é diametralmente oposto descrever-se a experiência consciente dos mesmos, o que parece ser um paradoxo, pois os fenômenos da consciência são o que de mais íntimo pode haver para nós.

Desde Kant, estamos cientes de uma assimetria no modo como a consciência se relaciona com o espaço e com o tempo. Embora

experimentemos objetos e eventos tanto espacialmente extensivos como de duração temporal, nossa consciência em si não é experimentada como espacial, embora seja experimentada como temporalmente extensiva. Na verdade, as metáforas espaciais para a descrição do tempo parecem, da mesma forma, praticamente inevitáveis para a consciência, como quando falamos, por exemplo, do “fluxo da consciência”. (SEARLE, 1997, p. 183)

A atenção kantiana sobre a consciência evidencia uma característica dos estudos da mente que já está presente em Descartes. De antemão a mente se mostra fugaz diante das nossas tentativas de conhecê-la. Seu funcionamento parece fugir àquilo que se espera de qualquer fenômeno, qual seja, de que funcione seguindo um ordenamento *a priori*. Como é peculiar à maioria dos pensadores do período moderno, Kant destaca que a dificuldade de se alcançar o conhecimento tem muito mais a ver com nossas limitações do que por algum obstáculo imposto pelas próprias coisas.

A consciência, vista dessa perspectiva kantiana, possui características semelhantes às das *situações observacionais* indicadas pelos teóricos da mecânica quântica, ao mostrar-se difícil de ser enquadrada num conjunto de leis e princípios de organização e funcionamento.

Blaise Pascal (1623-1662) também foi um pensador que deu importante contribuição para o surgimento do projeto de IA ao aproximar a filosofia da tecnologia com a construção de uma máquina de calcular mecânica chamada *pascalina* que realizava as operações de soma e subtração. As realizações das operações aritméticas de sua máquina o levaram a compará-las com a atividade humana de pensar, questões que somente encontrariam condições de serem novamente colocadas com o advento do computador digital. Atualmente uma das questões de maior relevância na discussão em torno do projeto de IA é o papel da vontade, da capacidade de escolher e da liberdade no desempenho das nossas ações.

A máquina aritmética produz efeitos que se assemelham ao pensamento, mas do que qualquer coisa que os animais possam fazer, no entanto, nenhum dos seus resultados nos permite declarar que ela esteja dotada de vontade como os animais. (PASCAL, fragmento 741, Pensées, 2006. www.ub.uni.freiburg.de/fileadmin/ub/referate/04/pascal/pensses.pdf - tradução nossa)

A observação de Pascal nos aponta um caminho possível para o encontro de um critério de reconhecimento de inteligência como a humana. Para ele a ação inteligente seria uma conjunção destes dois atributos, algo que se assemelha à nossa atividade de pensar acompanhado de vontade e querer.

Embora a construção da máquina de calcular de Pascal tivesse sido uma realização notável, ela não calculava com grande eficiência, nem era a mais desenvolvida em seu tempo. Quando Pascal apresentou sua primeira versão da máquina de calcular em 1642, já existia um dispositivo construído por Wilhelm Schickard (1592-1635), em 1624, chamado de *relógio de calcular*, capaz de realizar operações com melhor performance do que a *pascalina*. Como podemos verificar (ADAMSON, 1995), alguns a consideram como algo nada além de um dispositivo interessante e curioso.

Nove anos após a morte de Pascal, Leibniz inventou uma máquina mais versátil graças à uma roda escalonada, que lhe permitia multiplicar ou dividir por meio de repetidas adições ou subtrações. Contudo, nem mesmo a máquina de Leibniz podia ser considerada confiável. Numa exposição realizada em 1992 pelo Museu de Ciências de Londres, as máquinas de Schickard, Pascal e Leibniz foram corretamente chamadas de “curiosidades requintadas”: “por toda sua engenhosidade”, numa legenda era dito que, “esses dispositivos eram... brinquedos seriamente concebidos, que nem sempre eram confiáveis e que possuíam diferentes procedimentos de uso”. (ADAMSON, 1995, p. 24, tradução nossa)

A contribuição de Pascal, no entanto, não se restringiu à confecção de uma máquina capaz de realizar atividades semelhantes às do intelecto humano. Em suas investigações no campo da física, Pascal procurou responder à questões relacionadas à mecânica dos fluidos, ramo antigo da física que estuda o comportamento dos fluidos. Ele desenvolveu experimentos para saber o que produzia a formação de um vazio na extremidade superior de um tubo de vidro contendo líquidos como água, vinho e mercúrio.

Todavia, foi no campo da matemática que ele deu sua contribuição mais relevante para o surgimento do projeto de IA. Em colaboração com Pierre de Fermat (1601-1665) ajudou a desenvolver as primeiras investigações no que hoje se chama de *cálculo de probabilidade*.

Em cinco cartas sobre o cálculo de probabilidade, Pascal e Fermat fundam um ramo da matemática que era praticamente novo. Iniciado nessas cartas e aprofundado no trabalho sobre binômios, presente no *Tratado Sobre o Triângulo Aritmético*, essa investigação do cálculo de probabilidade decorreu do fascínio de Pascal pelos jogos de azar, o que aprendera a gostar com os seus novos amigos. O “Chevalier de Mere” era um hábil jogador. Na companhia desse e do duque de Roannez, Pascal deve ter jogado ou visto jogarem piquet, gamão e vários outros jogos de dados. Duas das questões discutidas com Fermat em sua correspondência dizem respeito a jogos de azar e lhe foram apresentadas por Mere, como a que pergunta sobre quantos lances de dois dados deverão ser feitos para que possamos esperar, pelo menos uma vez, uma dupla de seis? (ADAMSON, 1995, p. 35, tradução nossa)

Essa invenção conjunta de Fermat e Pascal constitui-se como um procedimento importante tanto no desenvolvimento de uma estratégia de abordagem do problema da incerteza, como veremos a seguir, quanto na elaboração de redes de dados.

A invenção do *cálculo infinitesimal* também representa um avanço de conhecimento relevante para o surgimento do projeto de IA. O cálculo infinitesimal goza da mesma importância e consideração que um dia foram dispensadas à lógica silogística, considerada pronta e acabada por muitos séculos desde a publicação da obra de Aristóteles.

Um dos instrumentos conceituais e analíticos mais básicos e importantes inventados pelos homens é o cálculo infinitesimal. Sem ele a história da ciência – o que em boa medida também quer dizer a história da humanidade – haveria sido diferente, muito diferente. É, ademais, uma ferramenta que goza de um *status* que outras, incluindo algumas das maiores sínteses teóricas produzidas na ciência, seguramente jamais poderão alcançar. Com efeito, parece ser o destino inevitável das teorias científicas serem substituídas por outras mais precisas, mais gerais. Assim sucedeu-se, por exemplo, com a mecânica e a teoria da gravitação newtonianas, ou com o eletromagnetismo maxwelliano. Com o cálculo infinitesimal não ocorrerá o mesmo. Poderão melhorar – isso já acontece e continuará a acontecer no futuro – seus pilares estruturais (por exemplo, a noção de infinitesimal ou limite), porém sua ideia, propósitos e procedimentos continuarão sendo os mesmos no futuro, e o que chamamos de “cálculo infinitesimal” seguirá sendo, ainda que sua apresentação seja diferente (o que provavelmente não ocorrerá), basicamente igual ao que seus inventores fizeram. (NEWTON & LEIBNIZ, 2006, p. 7)

O advento dessa ferramenta matemática também está inserido na vertente de pensamento que busca na regularidade a possibilidade de explicação dos fenômenos.

Dois personagens contemporâneos um do outro e de grande envergadura, tanto filosófica quanto científica, contribuíram de maneira definitiva para o advento do cálculo infinitesimal, Isaac Newton (1643-1727) e Gottfried Wilhelm Leibniz (1646-1716). Ambos pensadores chegaram quase que simultaneamente aos mesmos princípios do cálculo por vias distintas. Isso fez com que houvesse, por parte de Newton, a suspeita de que tivesse sido copiado por Leibniz, o que veio a ser provado um equívoco.

O cálculo infinitesimal é o instrumento de pensamento que tornou possível uma grande ampliação da nossa capacidade de compreensão e explicação de fenômenos que não podiam ser estudados pela física de Arquimedes, que não possuía a capacidade de fazer análises sobre as mudanças de estado dos fenômenos. Com o cálculo infinitesimal foi possível ir além da contagem, medição e descrição das formas das coisas. Ele possibilita o estudo do funcionamento de máquinas, do comportamento dos planetas, da fluidez dos líquidos, da eletricidade e outros fenômenos. Com essa ferramenta a matemática, tornou-se capaz de apreender as mudanças e variações de estado dos fenômenos.

Hoje o contexto de justificação científica já incorporou outros princípios e formas de compreender o mundo que o mostram diferente da representação matemática newtoniana.

Enquanto Newton havia descrito um mundo de perfeição mecânica como um relógio no qual partículas e objetos de todos os tamanhos seguiam padrões altamente disciplinados e previsíveis, as leis da termodinâmica descreveram um mundo de caos. De fato, é isso o que o calor é. O calor é o movimento caótico – imprevisível – das partículas que compõem o mundo. (KURZWEIL, 2007, p. 31)

Leibniz julgou ser a algoritmização do pensar humano a solução para uma série de dificuldades enfrentadas por nós, principalmente no que diz respeito às tomadas de decisões que envolvem opiniões contraditórias, o que passou a ser chamado de *o sonho de Leibniz*. Quando, em 1666, escreveu *Dissertacion Acerca del Arte Combinatorio* (1992), ele tinha por propósito produzir uma demonstração da tese da existência de Deus

com o mesmo rigor e grau de certeza que as demonstrações matemáticas. Seu sonho cibernetico consistia na definição de novos termos que possibilissem a elaboração de uma nova linguagem matemática capaz de realizar cálculos que provassem qual, dentre duas afirmações contraditórias, deveria ser levada em consideração numa determinada discussão. Sua tentativa de elaborar uma mecânica do pensamento o levou a construir uma máquina de calcular mecânica, capaz de realizar as quatro operações matemáticas. Para tanto, foi necessário desenvolver um sistema simbólico capaz de descrever os diferentes estados e mudanças ocorridas com os componentes constitutivos da máquina na realização das operações de cálculo.

Tão importante quanto a invenção do cálculo infinitesimal para o surgimento do projeto de IA foi o desenvolvimento da lógica do cálculo de predicados feito por Gottlob Frege (1848–1932). A lógica do cálculo de predicados permite a avaliação de construções argumentativas formadas por proposições que possuem uma estrutura mais complexa do que as da lógica silogística.

Frege escreve o artigo 'A Lógica Calculatória de Boole e a Conceitografia', só postumamente publicado, em que expõe e compara sua conceitografia com o que ele denomina de 'lógica calculatória de Boole'. Nele, Frege mostra que não poderia, para os fins a que se propunha, empregar a lógica de Boole, pois necessitava de um sistema com maiores recursos, e assim defende a superioridade de sua conceitografia em relação à lógica booleana. (Frege, 2009, p. 18)

Frege identifica no uso da linguagem ordinária a origem das dificuldades que não permitem a clareza de sentido e rigor no emprego dos elementos implicados numa argumentação. Em *Sobre a justificação científica de uma conceitografia*, publicado em 1882, ele afirma que:

Nas partes mais abstratas da ciência torna-se cada vez mais inequívoca a falta de um meio que permita, ao mesmo tempo, evitar incompreensões quanto ao pensamento (*Denken*) de outrem, e também equívocos sobre o nosso próprio pensamento. Tanto um como o outro tem sua causa na imperfeição da linguagem, já que temos de usar sinais sensíveis para pensar. (FREGE, 2009, p. 59)

Para que isso pudesse ser realizado era necessária a produção de uma demonstração, única forma aceitável para provar se uma determinada proposição sobre as propriedades dos números naturais era válida ou não. Frege desenvolveu um sistema lógico mais poderoso do que o de Aristóteles, pois ele possibilita a realização de análises de argumentos constituídos por tipos de proposições que não estavam no escopo da lógica aristotélica, ampliando consideravelmente a capacidade de representação simbólica e de avaliação lógica.

O método de prova (*Beweisführung*) mais seguro consiste, obviamente, em seguir estritamente a lógica, que, abstraindo as características particulares das coisas, apoia-se exclusivamente nas leis sobre as quais se baseia todo o conhecimento. Por esta razão, dividimos todas as verdades que requerem prova em duas espécies: aquelas cuja prova pode ser conduzida por meios puramente lógicos e aquelas cuja prova se apoia em fatos empíricos. (FREGE, 2009, p. 11)

O desenvolvimento desse sistema na primeira metade do séc. XIX, que tinha por propósito servir de fundamentação para a aritmética, proporcionou a realização de análises lógicas que consistem na execução de um procedimento que atribui valores numéricos e simbólicos, ao invés de palavras, às características de um objeto pertencente a um determinado conjunto. Um dos objetivos da investigação de Frege foi a construção de uma linguagem que não sofresse dos problemas de imprecisão, ambiguidade e polissemia presentes no uso da linguagem ordinária. Ele separa a parte lógica de funcionamento da linguagem de todas as demais. Acreditava que assim alcançaria o maior rigor possível na construção e análise de argumentos, elemento básico de toda e qualquer explicação científica.

[...] o desejo de clarificar as noções fundamentais da matemática e de tornar exata a concatenação dessas noções - como evidencia a resenha de um insignificante manual de aritmética publicada ainda em 1874- talvez tenham sido os fatores que desencadearam e motivaram a maioria de seus trabalhos, a saber, a procura de uma fundamentação para a aritmética. (FREGE, 2009, p. 11)

Essa delimitação do uso lógico da linguagem levou-o a fazer considerações que influenciam ainda hoje o debate em torno do projeto de IA. Uma delas é de que os elementos psicológicos da linguagem não possuem relevância na compreensão do funcionamento lógico da mesma, o chamado *anti-psicologismo* de Frege, que consiste na ideia de que a Lógica é distinta da Psicologia, porque enquanto aquela trata com fenômenos que são de natureza objetiva e universal; essa trata com fenômenos subjetivos e particulares.

Um homem pode imaginar mais e outro pode imaginar menos. Pois aqui não nos encontramos no domínio da psicologia, da representação, do que é subjetivo, mas no âmbito do objetivo e do verdadeiro. (FREGE, 2009, p. 59)

[...]

A linguagem não é regida por leis lógicas, de maneira que a mera observância da gramática seja suficiente para garantir a correção formal do curso do pensamento. (FREGE, 2009, p. 61)

Susan Haack (2002, p. 310) afirma que com relação aos processos mentais e a lógica existem três posições distintas: a primeira afirma que a lógica descreve os processos mentais, a segunda que ela diz como deveríamos pensar e a terceira que ela nada tem a ver com os processos mentais. A autora considera que a posição fregeana encontra-se entre as do terceiro tipo.

Sempre fortemente anti-psicologista, Frege enfatiza que o sentido, ou significado, de uma expressão deve ser distinguido da ideia que pode acontecer estar associada a essa expressão. Assim, quando ele diz que o sentido de uma sentença é o pensamento (*Gedanke*) que ela expressa, ele quer dizer ‘proposição’, em vez de ‘ideia’. (HAACK, 2002, p. 97)

Ao fazer esse corte epistêmico, Frege inseriu uma distinção que exerce forte influência no debate em torno dos objetivos do projeto de IA, a separação entre *logicistas* e *psicologistas*. Para ele, a descoberta dos mecanismos lógicos de funcionamento da linguagem não pode ser confundida com a descoberta das regras que regem o funcionamento de todo pensamento. Essa concepção da tarefa da lógica serve de

fundamento para a ideia de que o pensamento pode ser realizado por uma estrutura distinta da biológica devido ao seu caráter objetivo e universal.

O nascimento do projeto de I.A.: Alan Turing e o esforço de guerra

Russel & Norvig (2004) afirmam que Warren McCulloch e Walter Pitts são reconhecidamente os precursores das publicações que tratam de questões relacionadas ao projeto de inteligência artificial ao publicarem *A logical calculus of the ideas immanent in nervous activity*, em 1943, no *Bulletin of Mathematical Biophysics*, quase uma década antes do ensaio de Turing que trata sobre o reconhecimento de inteligência artificial.

Partiram de três fontes: conhecimento sobre a fisiologia básica e funcionamento dos neurônios no cérebro, a análise formal da lógica proposicional de Russell e Whitehead e a teoria da computação de Turing. Propuseram um modelo constituído por neurônios artificiais, no qual cada um deles se caracteriza por estar “ativado” ou “desativado”; a ativação se dava em resposta à estimulação produzida por uma quantidade suficiente de neurônios vizinhos. O estado de um neurônio via-se como equivalente, de fato, a uma proposição com alguns estímulos adequados. (RUSSEL & NORVIG, 2004, p. 19, tradução nossa)

Podemos perceber que o contexto de surgimento do projeto de IA é nutrido pelo cruzamento de conhecimentos advindos de áreas distintas. Ao compararem o funcionamento de uma célula viva (neurônio) com o de um dispositivo eletrônico (válvula, circuito, relé, transistor, chip) e identificarem uma característica comum a ambos, a de possuírem dois estados de funcionamento, McCulloch e Pitts estabeleceram uma conexão entre o mundo da biologia e o da eletro-eletrônica, entre o mundo natural e o do homem, sendo esse o primeiro passo dado na direção dos objetivos da IA. Tal conexão estabelece uma forte força de coesão entre os elementos básicos que compõem o cenário de apresentação e desenvolvimento do projeto de IA. É essa conexão, que une o mundo natural ao do conhecimento humano, que serve de modelo para a organização e estruturação das ideias implicadas neste contexto.

Entretanto, foi o conflito da Segunda Guerra Mundial o grande catalisador que promoveu o surgimento do artefato que daria o impulso definitivo para a elaboração do projeto de IA, o computador. Essa máquina não foi construída originariamente com o propósito de obter-se inteligência artificial, ela foi construída para ser empregada na realização de cálculos para decifrar códigos de guerra, na construção dos primeiros mísseis de cruzeiro, na construção da bomba atômica, no aumento da eficiência da artilharia. Ela foi criada visando ao cumprimento de objetivos de guerra.

A equipe de Alan Turing construiu, em 1940, o primeiro computador operacional com funcionamento eletromecânico, chamado “Heath Robinson”, com um único propósito: decifrar mensagens alemãs. Em 1943 o mesmo grupo desenvolveu o “Colossus”, uma máquina potente de propósito geral construída com válvulas à vácuo. O primeiro computador operacional programável foi o “Z3”, inventado por Konrad Zuse na Alemanha em 1941. Zuse também inventou os números de ponto flutuante (forma de representação dos números reais no computador) e a linguagem de programação de alto nível, Plankalkül. O primeiro computador eletrônico, o “ABC”, foi criado por John Atanasoff junto com o seu discípulo Clifford Berry, entre 1940 e 1942 na Universidade Estadual de Iowa. (RUSSEL & NORVIG, 2004, p. 16, tradução nossa)

As principais potências ocidentais* travaram uma disputa silenciosa para alcançarem o domínio de conhecimento necessário à construção do computador digital, que se mostrou uma arma de guerra tão poderosa quanto o mais devastador artefato bélico.

Somente após o término da Segunda Guerra Mundial que o projeto de IA foi concebido nos moldes como o conhecemos, isto é, o uso do computador digital como uma plataforma capaz de suportar e desenvolver atividades que exigem o emprego de inteligência na sua realização. Por essa razão, consideramos o ensaio *Computação e Inteligência* de Alan Turing (1912-1954), publicado em 1950, o marco inicial do projeto

* Um fato a salientar é não terem levado muito a sério a iniciativa privada de construção dos primeiros computadores, tanto na Alemanha quanto nos Estados Unidos (RUSSEL & NORVIG, 2004). Konrad Zuse e John Atanasoff tiveram muitas dificuldades para conseguir financiamento para construir e desenvolver suas máquinas. Os ingleses, contudo, agiram de forma diferente. Talvez por seu caráter pragmático, eles vislumbraram um uso imediato dos computadores, a substituição de seres humanos na realização dos cálculos para desvendar códigos de guerra alemães. No caso de Atanasoff, nos Estados Unidos, sua dificuldade foi aumentada porque o projeto de construção da bomba atômica desenvolvia seu próprio calculador eletrônico.

de IA, porque nele é expressada, pela primeira vez, a preocupação com o problema do reconhecimento de inteligência, por meio da aplicação de um teste de competência no uso da linguagem ordinária.

Uma das aspirações dessa área de interesse do conhecimento é a elaboração de uma teoria do comportamento inteligente, porém, uma teoria que não tivesse exclusivamente como ponto de partida as ações humanas e nem restringisse seu poder explicativo a elas. Essas questões ganharam relevância com o advento do computador digital, máquina capaz de desempenhar atividades consideradas exclusivas do homem, como jogar xadrez, organizar o fluxo de veículos no trânsito de uma cidade, vigiar passageiros num aeroporto, escolher um alvo dentre muitos, regular a temperatura do cobertor pelo diferencial de temperatura entre o ambiente e o corpo humano, etc.

Desde que Turing lançou o desafio do *jogo da imitação*, dois ramos da inteligência artificial se distinguiram, um denominado fraco e o outro forte. O primeiro é considerado fraco porque seu propósito visa meramente a empregar o computador digital como bancada de testes para a implementação de programas que teriam por objetivo servir de modelo para a descrição da arquitetura e funcionamento do intelecto humano. O segundo ramo é considerado forte, porque seu nível de afirmação eleva ao grau máximo os objetivos do projeto de IA, ou seja, esse ramo tem por propósito construir uma máquina capaz de alcançar o mesmo *status* ontológico que o de um ser humano ao dotá-la das capacidades de pensar e decidir.

Quando nos referimos aqui ao conceito de inteligência artificial, o fazemos levando em consideração o ramo forte do projeto de IA. Está na radicalidade da sua pretensão a razão do maior grau de exigência das explicações em torno da questão da admissibilidade desse projeto, pois sua versão forte objetiva algo bem mais ambicioso do que uma simulação, objetiva uma reprodução. O caráter ambicioso da versão forte do projeto de IA dá relevância aos problemas de natureza filosófica implicados nesse projeto. Tais problemas necessitam ser abordados se pretendemos dizer algo a respeito da admissibilidade a respeito de um projeto com tamanha envergadura.

Os pioneiros da I.A. e seus problemas

Uma das primeiras realizações em inteligência artificial foi a máquina de John von Neumann (1903-1957), esse dispositivo permitiu que os programas fossem gravados numa mesma unidade de memória juntamente com os dados que seriam analisados. A forma como as operações são desenvolvidas nesse tipo de máquina é linear, ou seja, há um fluxo de diretivas realizadas umas após as outras que seguem um rígido ordenamento. Isso objetivava evitar que erros de funcionamento passassem despercebidos e determinassem negativamente o resultado final de todo o processamento. Esse modo de funcionamento não serve como representação metafórica adequada para a mente devido à sua incapacidade de realizar várias operações diferentes simultaneamente. O grau de eficiência (capacidade de realizar determinado número de operações por segundo) de uma máquina com uma arquitetura desse tipo é bastante reduzido, seu funcionamento é lento, pois cada etapa do processamento tem de seguir uma ordem de sequência de tal modo que mesmo uma etapa mais simples não tem como ser realizada antes que as mais complexas se desenvolvam completamente de acordo com o ordenamento. Deste ponto de vista, o agir *inteligente* consistiria no desenvolvimento de ações, que concorrem para um mesmo propósito, perfeitamente encadeadas e previamente ordenadas.

De certo modo, o projeto de IA resgata a teoria mecanicista na medida em que concebe o homem como máquina, como um conjunto de diferentes partes que existem umas em razão das outras, constituindo uma totalidade. Esse resgate, na verdade, coloca o mecanicismo num outro patamar, já que as máquinas projetadas pela IA não são concebidas para serem meros autômatos que seguem as leis da mecânica sem realizar nada além daquilo que é esperado que façam.

Por sua vez, o *conexionismo* não vê nos tipos de mecanicismo, como são os modelos clássico de Von Neumann e o funcionalista de Hilary Putnam (1926-2016), aqueles capazes de representar adequadamente o modo de funcionamento da mente, pois nossos pensamentos não se restringem a somente um fluxo contínuo de ideias. A noção de inteligência é sutilmente alterada devido à forma como o conteúdo mental é organizado. Ao invés de um modelo sequencial que segue um fluxo, no conexionismo a forma de

organização é em rede, núcleos autônomos de processamento que se relacionam independentes uns dos outros. Esse tipo de arranjo permite a realização de operações em paralelo, o que possibilita a execução de várias tarefas simultaneamente, algo que se aproxima mais do funcionamento da mente humana.

Há o pressuposto de que a maioria dos pensadores e cientistas envolvidos no projeto de IA concorda com a ideia de que é no cérebro que ocorrem os principais fenômenos envolvidos numa ação inteligente, ação essa a qual podemos aferir um propósito, uma intenção, uma vontade, não meramente um movimento mecânico onde tais características estão ausentes.

Hoje vive-se uma *corrida ao cérebro*, pois ele é o *dispositivo* mais elaborado que poderíamos copiar para reproduzir artificialmente aquilo que nos torna humanos. O estudo do cérebro humano não é um campo de investigação exclusivo de uma área de conhecimento. Dependendo da área e do tipo de abordagem, o cérebro pode ser visto tanto como modelo, quanto de exemplo a ser superado. As diversas disciplinas que se debruçam sobre o cérebro desenvolvem seus próprios métodos de abordagem, empregando diferentes estratégias para fazer com que ele nos forneça as informações que necessitamos saber para, assim, compreendermos adequadamente seu funcionamento. Por exemplo, a versão forte do projeto de inteligência artificial cogita a possibilidade de construir-se cérebros artificiais mais desenvolvidos que os nossos atuais, com capacidades de elaborar pensamentos que não temos nem como imaginar. Esses novos cérebros exerceriam, no futuro, um papel importante no processo evolutivo do nosso próprio cérebro.

O projeto de IA, como de resto qualquer outro produto da criação humana, assume características que dizem respeito à própria época em que foi concebido. Umas das características mais marcantes do séc. XX foram os fenômenos de massa impulsionados pelos *media*, em que a propaganda e a publicidade se destacaram como os seus mais eficientes aplicativos. O exagero, o oxímoro, o acento demais, o chamar a atenção e o interesse do maior número de pessoas possível, tiveram bastante influência no contexto de surgimento do projeto de IA com previsões e expectativas exageradas de desenvolvimento como podemos ver nas obras de Hans Moravec (*Mind Children: the future of robots and human intelligence*, 1988) e Ray Kurzweil (*A era das máquinas*

espirituais, 2007). A versão forte da IA sempre teve maior arrojo em propor candidatos digitais ao título de inteligência genuína.

Um dos exemplos de exagero cometido por alguns pioneiros da IA, foi a afirmação de Herbert Simon (1916-2001), intelectual americano que trabalhou em várias áreas do conhecimento, de que as máquinas já em 1957 podiam pensar, pois ele, juntamente com seu colega Allen Newell (1927-1992), desenvolveram um programa, o *Logic Theorist*, capaz de solucionar problemas matemáticos, tendo resolvido vários dos teoremas apresentados por Alfred North Whitehead (1861-1947) e Bertrand Russel (1872-1970) nos *Principia Mathematica* (1997).

A execução desse programa era considerada uma demonstração do uso de inteligência obtida artificialmente, já que na resolução de teoremas a capacidade criativa é uma condição indispensável a ser cumprida para que seja aberto um caminho, no uso dos axiomas do sistema, que proporcione uma solução para os problemas que se apresentam. É compreensível que a apresentação do trabalho de Newell e Simon tenha provocado bastante debate na ocasião do Simpósio de New Hampshire, pois o programa *tomava decisões* à medida que elaborava as estratégias de prova dos teoremas.

A afirmação de Newell e Simon além de ser de difícil constatação, também é de difícil aceitação, porque ela vai de encontro a um axioma da antropologia clássica ao considerar que o homem não seria o único ente capaz de ter pensamentos complexos. Talvez a razão da não aceitação da afirmação de Newell e Simon seja que os conceitos que funcionam bem na formulação da compreensão de mundo que desenvolvemos até aqui, não apresentam o mesmo grau de eficiência quando o ambiente é alterado com a introdução de um ente que age de forma inteligente devido ao fato dos seres humanos, com o seu conhecimento, terem-no dotado dessa capacidade. As ferramentas conceituais necessárias à elaboração de uma compreensão de mundo ou podem estar sendo mal-empregadas ou são deficientes quando se trata da elaboração de uma explicação mais robusta do que seja o pensar e a ação inteligente.

O conceito de inteligência apresenta-se como um problema por aparentar ser algo que possui diferentes graus, nos permitindo afirmar que algo é mais, ou menos, inteligente do que outro. Uma das maneiras de definir a inteligência seria a aplicação do método

behaviorista e observar o que é possível de ser observado no agir inteligente, isto é, o comportamento exibido pelo agente em consideração. Contudo, essa estratégia de estudo para definir a inteligência não consegue preencher todas as lacunas explicativas, pois o comportamento sofre influência do ambiente e isto pode levar o observador a equívoco.

A noção de comportamento imprevisível, ou alteração de comportamento devido a mudanças no ambiente, foi empregada por William Grey Walter (1910-1977) para explicar o comportamento de pequenos robôs construídos por ele. Seus pequenos robôs possuíam um sensor de luz e motores elétricos que dirigiam as rodas em direção a uma fonte de luz. Essas pequenas máquinas iam em direção a uma fonte de luz, porém, faziam outra coisa, evitavam fontes de luz com brilho intenso.

Um dos experimentos a que foram submetidos esses robôs consistia em colocar no ambiente deles uma fonte de luz com brilho intenso servindo de ponto de reabastecimento das cargas das suas baterias. O comportamento esperado dos robôs é que eles continuassem buscando uma fonte de luz e evitassem a fonte de luz com brilho intenso. Entretanto, quando a carga das baterias diminuía a sensibilidade ao brilho extremo era afetada. Ao precisarem recarregar as baterias, eles deixavam de evitar a fonte de luz com brilho intenso. Isso foi interpretado por Grey Walter como um comportamento semelhante ao do ser humano que, embora não sendo pré-programado para enfrentar uma determinada situação específica, ajusta seu comportamento à ela.

Sua explicação é de que foram as alterações no ambiente (fonte de luz com brilho intenso como ponto de recarga da bateria do robô) e, nos próprios robôs (diminuição da carga da bateria com consequente diminuição da percepção de fonte de luz com brilho intenso), que ocasionaram a mudança de comportamento dos mesmos. Com isso, ele acreditava ter demonstrado que as mudanças de comportamento dos robôs eram uma resposta inteligente às alterações ocorridas no seu ambiente.

Uma das críticas dirigidas à explicação de Grey Walter é de que ele já esperava, quando montou o experimento, que a diminuição da carga da bateria poderia afetar o sensor de luminosidade dos robôs e por isso montou um ponto com luz intensa para servir de referência, já que os pontos com luz sem brilho intenso não teriam como atrair o robô,

cujo sensor de luz teve seu funcionamento afetado. Por essa razão posicionou o carregador exatamente no ponto em que estava a fonte de luz com brilho acima do normal.

Pouco antes do início da II Guerra Mundial, foi construído nos Estados Unidos o chamado *analisador diferencial de Bush*, um computador analógico capaz de resolver equações diferenciais. O objetivo de construção dessa máquina era a realização de cálculos que tornassem a artilharia mais precisa. Ao estudar o princípio de funcionamento dessa máquina, Claude Elwood Shannon (1916-2001) descobriu que os componentes que formavam os circuitos, na época eram empregadas relés, possuíam dois estados, fechado e aberto, que se ajustavam de maneira precisa às leis descritas pela álgebra de George Boole (1815-1864).

Em 1854, Boole publicou *An investigation of the laws of thought*, obra em que apresenta seu procedimento de formalização no tratamento dos sistemas lógicos. Ele matematizou a lógica, isto é, desenvolveu um sistema matemático para ser empregado na representação e resolução de problemas lógicos. Tal procedimento se apoia em princípios que regem todo pensamento que pretenda alcançar rigor científico.

Toda ciência consiste de verdades gerais, mas, dessas verdades, somente algumas são primeiras e fundamentais, as outras são secundárias e derivadas. As leis do movimento elíptico de Kepler são verdades gerais da astronomia, mas elas não são verdades fundamentais. Isso também ocorre nas ciências puramente matemáticas.

[...]

Vamos definir como fundamentais as leis e princípios nos quais todas as demais verdades gerais das ciências podem ser deduzidas, e nos quais todas podem ser resolvidas. Erramos ao considerar isso como a verdadeira ciência da lógica que, ao estabelecer as leis elementares confirmadas pelo testemunho verdadeiro da mente nos permite, então, deduzir por processos uniformes a cadeia inteira de suas consequências secundárias, fornecendo-nos, para aplicação prática, métodos de generalizações perfeitas? (BOOLE, 1854, p. 5, tradução nossa)

De acordo com o pensamento de Boole, para que alcancemos rigor no uso de argumentos, é necessário que sejamos capazes de analisar as relações lógicas estabelecidas entre os elementos que constituem o dado argumento. Para evidenciar o caráter lógico dessas relações, ele abstraiu o conteúdo semântico das proposições por

meio da sua substituição por símbolos. Ele desenvolveu um sistema algébrico cujos símbolos denotam proposições e os operadores lógicos que as relacionam, sendo que as mesmas possuem somente dois valores (verdadeiro ou falso).

Com uma linguagem assim constituída é possível observar e avaliar logicamente as diferentes construções de pensamento que empregam tais elementos. Assim, torna-se possível demonstrar se uma determinada ideia pode ser, ou não, deduzida de outras.

Que a linguagem é um instrumento da razão humana e não meramente um meio de expressão do pensamento é uma verdade geralmente admitida... Em várias etapas dessa investigação seremos levados a considerar a constituição da linguagem, compreendida como um sistema adaptado a um fim ou propósito; investigar seus elementos; buscar determinar sua mútua relação e dependência; e perguntar de que maneira eles contribuem para a realização do fim ao qual, como parte de um sistema de coordenadas, eles dizem respeito. (BOOLE, 1854, p. 5, tradução nossa)

Ao aplicar a lógica booleana no estudo e elaboração de circuitos lógicos, além de descobrir a linguagem que permite a compreensão, construção e análise de diferentes circuitos lógicos, empregados no desenvolvimento das máquinas digitais, ele desenvolveu uma teoria que estima a quantidade de informação propagada por um determinado meio, a *Teoria da Informação*. Tal teoria é importante no desenvolvimento de IA, pois a velocidade da informação propagada num determinado meio é um dado de suma relevância na construção das máquinas digitais. Se o objetivo do projeto de IA é construir um dispositivo tão capaz quanto a capacidade humana de pensar, então é necessário saber-se qual a medida dessa realização, fator esse criado por Shannon com a introdução do bit (*binary digit*) como unidade básica de informação.

Algumas das principais vertentes e questões da I.A.

Uma das primeiras e principais linhas de pesquisa e investigação no campo da inteligência artificial foi o *funcionalismo*. Ele foi uma reação ao modelo alemão de investigação psicológica do século XIX encabeçado por Wilhelm Maximilian Wundt

(1832-1920) que tinha por objetivo investigar a consciência a partir dos elementos comuns envolvidos em cada experiência de consciência. Na obra *Principles of physiological psychology* (2002), Wundt expõe sua tentativa de aproximar a psicologia da fisiologia e a justifica por se tratarem de ciências que se ocupam em explicar fenômenos vitais.

A Fisiologia e a Psicologia abrangem o campo dos fenômenos da vida. Elas tratam com fatos da vida em geral e, em particular, com os fatos da vida humana. A Fisiologia se interessa por todo fenômeno vital que se apresenta a nós como percepção sensível ou processos corporais, e que constituem parte do que chamamos mundo externo. Por outro lado, a Psicologia busca dar conta da interconexão dos processos que são evidenciados por nossa própria consciência ou que inferimos das manifestações da vida corporal de outras criaturas que nos indicam a presença de uma consciência similar à nossa. (WUNDT, 2002, p. 1, tradução nossa)

Essa proposta de investigação visa obter um acesso às experiências de consciência por meio do emprego de um método introspectivo de descrição dos seus elementos mais básicos, as sensações. As observações seriam coletadas por um psicólogo da descrição da experiência de consciência feita por outro psicólogo. Desse modo, pretendiam alcançar maior rigor na obtenção e análise dos elementos que deveriam ser levados em consideração na compreensão da experiência de consciência. Através da descrição rigorosa das sensações presentes numa dada experiência de consciência, seria possível identificar a estrutura do mental.

Distinto do modelo alemão do séc. XIX, o *funcionalismo* busca identificar a estrutura psicológica humana a partir das funções e processos realizados pela mente, não nas sensações e sentimentos. Para essa corrente de estudos da mente, os verdadeiros objetos de estudo da psicologia são as capacidades de sentir, pensar e decidir. A denominação *funcionalismo* é em razão de que, para essa concepção, nossas atenções, ao procurarmos entender a ação inteligente, devem se voltar para as funções desempenhadas pelos componentes da estrutura cognitiva que torna possível tal ação, não para os elementos que a constituem.

O que merece, de fato, ser levado em consideração são os resultados que essas funções efetivam ao serem executadas, constituindo-se nos materiais básicos de toda ação inteligente. Neste caso, não é relevante saber do que é constituída uma determinada estrutura cognitiva e inteligente, mas sim saber, se ela é capaz de desempenhar as funções inerentes à produção de uma ação considerada inteligente. Para essa forma de abordagem, é importante não deixarmos de dar a devida relevância aos processos envolvidos nas ações inteligentes. Os componentes materiais que constituem aquilo no qual são desenvolvidas essas ações não possuem o mesmo grau de relevância que elas, já que uma mesma ação pode ser desenvolvida por diferentes dispositivos constituídos de distintos componentes.

Se, por um lado, as pesquisas envolvidas no projeto de IA revigoraram, por exemplo, o *mecanicismo* como modelo de compreensão do homem, por outro, elas provocaram o abandono de métodos que visam investigar o problema de saber como se dá uma ação inteligente.

Uma das formas de estudo do comportamento animal é o *behaviorismo*, cujo método consiste na adoção de um procedimento rigoroso de observação, com o intuito de chegar-se aos elementos mais básicos de um comportamento inteligente, o que nos colocaria em condições de elaborar uma explicação de como e porque uma determinada ação inteligente ocorre. Para o *behaviorismo* o fundamental, para entender-se o comportamento animal, está naquilo que ele possui de observável. Não sendo necessário pressupor a existência de entidades mentais ou de qualquer outro tipo de evidência para que possamos elaborar uma explicação suficiente do comportamento.

O *behaviorismo* surge a partir de uma crítica ao *funcionalismo*, pois, de acordo com essa concepção, não é confiável a postulação de eventos internos e não observáveis para alcançar-se a compreensão da ação inteligente. Se, por um lado, o *funcionalismo* pressupõe uma atividade interna como condição de compreensão da ação inteligente, de outro, o *behaviorismo* procura restringir-se ao que é dado e observável, sem apelar à existência de processos internos inerentes a uma estrutura mental.

O advento da máquina de estado discreto de Turing tornou possível a proposição de um outro tipo de método empregado na elaboração de explicações do comportamento inteligente. Ao contrário do behaviorista, naquela forma de investigação são pressupostos

processos mentais como a memória, a capacidade de aprender e usar uma linguagem, saber tratar com ideias, etc. Esse novo método, chamado *cognitivismo*, pretende aprofundar a observação através de um procedimento de teste: organiza na máquina a sequência de ações, observa como esta funciona e, por último, verifica se o mesmo ocorre com os seres humanos.

Tal forma de investigação tem como pressuposto a ideia de que as operações que o pensamento humano realiza são da mesma natureza que as realizadas pela máquina de Turing, os chamados *cálculos computacionais* que, ao serem realizados, produzem uma complexidade que daria suporte às qualidades emergentes do sistema, como a consciência, por exemplo, fenômenos que não podem ser apreendidos pela simples observação do comportamento.

O cognitivismo entende que a compreensão da ação inteligente deve-se dar a partir do estudo de processos internos dos agentes inteligentes. Para tanto, são considerados mais relevantes os casos em que usamos a inteligência para alcançar conhecimento, uso esse que envolve sermos capazes de aprender novas coisas, de lidar com ideias, de usar uma linguagem. De acordo com essa corrente, esses processos devem ser compreendidos de maneira mais aprofundada se pretendemos produzir inteligência artificial, sendo que os mesmos devem ser, também, considerados processos computacionais, isto é, processos que podem ser decompostos em todas as suas etapas de realização.

O objetivo da abordagem cognitivista é determinar quais os tipos de operações de pensamento devem ser realizadas para que o surgimento da inteligência ocorra. Uma vez descobertas quais sejam essas operações a próxima tarefa consistiria na elaboração de um programa que teria, em tese, as mesmas propriedades emergentes da mente humana. Nessa perspectiva, não é importante saber de que tipos de materiais são formados os componentes que executam as operações de pensamento. Para essa corrente, as condições que devem ser cumpridas, para que tenhamos condições de saber o que seja o agir inteligente são: sermos capazes de identificar com precisão quais sejam as operações que produzem a emergência da inteligência e qual a sequência exata em que elas são empregadas.

A ideia de que essas operações são de natureza computável é criticada por John Searle que não acredita ser o pensamento a resultante da capacidade de combinar símbolos, como ele demonstra no *argumento do quarto chinês*. Na formulação de seu experimento mental, Searle demonstra que mesmo que sejamos capazes de combinar corretamente os vocábulos de uma linguagem desconhecida, isso não nos permite deduzir que tenhamos compreensão dos significados dos símbolos que manipulamos de forma eficiente.

Responder à questão se máquinas, como os computadores digitais, podem pensar ou não, envolve uma série de dificuldades. A principal delas, o próprio significado de *pensar*. Talvez por essa razão, Turing, ao propor uma resposta ao problema, não buscou esclarecer o significado de conceitos como *pensar, inteligência, consciência, vontade*. Ele sugeriu uma saída de natureza prática quando apresentou o *jogo da imitação*.

Mas, como o próprio nome do jogo evidencia, o alcance máximo dessa estratégia é a realização de uma imitação, não de um desempenho genuinamente inteligente. Isto não significa dizer que uma máquina deveria ser considerada, no máximo, uma imitadora competente de uma capacidade exclusiva do homem. A saída *turingeana* nos indica que as dificuldades em se atribuir a capacidade de pensar às máquinas devem-se mais a distinções e a definições conceituais do que propriamente a dificuldades de natureza tecnológica e científica. Definir em que consiste o pensamento é uma tarefa semelhante à definição do que seja o tempo, pois embora sejamos capazes de pensá-lo, isso não torna mais fácil a tarefa de definir em que ele consiste.

De acordo com John Searle, mesmo que aceitássemos a hipótese de que uma determinada máquina fosse agente de ações inteligentes, após ser submetida ao teste, a aplicação deste não nos permite compreender nada a respeito do funcionamento dessa máquina, não saberíamos como ocorre, nem a partir do que se dão as supostas ações inteligentes executadas por tal dispositivo. Searle não acredita que uma combinação de elementos sem capacidade de entendimento seja capaz de gerar entendimento. A inteligência é algo que tem de ser pressuposto, ela não pode ser compreendida como uma espécie de acidente, a mera resultante das interações entre os constituintes do ente capaz de pensar.

A crítica de Searle ao projeto de IA é dirigida aos programas elaborados com o fim de participar do teste de Turing e, mais especificamente, ao programa de Roger Shank (1946) e Robert Paul Abelson (1928-2005) de 1977. Roger Schank é um pesquisador da área de inteligência artificial que desenvolveu os primeiros programas aplicativos educacionais. Para ele, a ciência cognitiva representa a salvação dos sistemas educacionais ao desenvolver modelos que descrevem como a mente humana trabalha. Searle em *Mentes, Cérebros e Programas* afirma que o objetivo do programa de Shank e Abelson, ao participarem do teste de Turing, foi "... simular a habilidade humana para compreender histórias". (TEIXEIRA, 1996, p. 64).

Contudo, de acordo com Searle, ser capaz de realizar operações que pressupõem o uso de símbolos a partir de regras sintáticas, que determinam como os símbolos devem ser relacionados, não explica como ocorre a compreensão. Ela é uma atividade que envolve a capacidade de preencher-se as lacunas de uma determinada narrativa, o que pressupõe a elaboração de uma interpretação.

É característico na habilidade dos seres humanos para compreender histórias que estes possam responder questões sobre elas, mesmo se a informação não estiver explicitamente dada no texto. (TEIXEIRA, 1996, p. 64)

Dotar um dispositivo com regras de manipulação simbólica não é suficiente para que o resultado da manipulação dos símbolos, mesmo que eficiente, seja considerado uma ação inteligente acompanhada de compreensão e entendimento. A compreensão não é algo que possa fazer parte de um programa de computador, ela não é obtida porque seguimos um programa que nos permite realizar operações computacionais.

[...] se o programa de Schank não constituir uma explicação completa da compreensão, talvez constitua uma parte de tal explicação. Ou seja, assumimos como possibilidade empírica, embora sem razões para supor que ela seja verdadeira (uma vez que ela é apenas sugerida e não demonstrada) que o programa de computador é irrelevante para a minha compreensão da história. ...O que queremos dizer é que por mais que se coloque no computador princípios formais isto não será suficiente para a compreensão, uma vez que um ser humano será capaz de seguir tais princípios formais sem compreender nada. (TEIXEIRA, 1996, p. 69)

Na opinião de Searle o que nos induz a pensar que as operações computacionais realizadas pelo computador digital são da mesma natureza que as operações que nos propiciam alcançar compreensão, deve-se ao fato de que “em muitas discussões encontramos muita confusão sobre a palavra ‘compreensão’”. (TEIXEIRA, 1996, p. 70) Esse estado de confusão é gerado pela riqueza de significados que o termo “compreensão” possui, fazendo com que seus diferentes sentidos sejam empregados em situações em que eles não são pertinentes. Outra razão para esse estado de confusão se efetivar seria o hábito que temos de atribuir a objetos, de maneira metafórica, as qualidades que são atributos nossos. Não há problema algum no emprego de metáforas para se compreender o que se passa com os objetos, desde que não nos esqueçamos de que se trata de um recurso de linguagem, não representando uma definição exata do objeto em questão. A estratégia de emprego de metáforas como via auxiliar na compreensão de algo só faz sentido se entendermos que o paralelo estabelecido funciona como uma abertura de sentido que nos permitirá um entendimento mais aprofundado daquilo que pretendemos conhecer.

A razão pela qual fazemos essas atribuições é interessante e tem a ver com o fato de que estendemos nossa própria intencionalidade para os artefatos. Nossos instrumentos são extensões de nossos propósitos e assim achamos natural fazer atribuições metafóricas de intencionalidade a eles; mas estes exemplos não resolvem nosso problema filosófico. O sentido no qual uma porta automática “compreende instruções” através de sua célula fotoelétrica não é de jeito nenhum o sentido no qual eu comproendo inglês. (TEIXEIRA, 1996, p. 71)

A razão principal da não aceitação, por parte de Searle, de que o computador, no qual tivesse instalado o programa de Shank, não alcança compreensão alguma das histórias que lhe são contadas é porque este não possui *intencionalidade*, isto é, não é capaz de fazer referência a objetos distintos dele mesmo.

Não é porque eu sou a instanciação de um programa de computador que eu sou capaz de entender inglês e ter outras formas de intencionalidade (eu sou, suponho, a instanciação de qualquer programa de computador) mas pelo que sabemos é porque eu sou um certo tipo de organismo com uma certa estrutura biológica (física e química) e, esta estrutura, em

termo causais, é capaz, sob certas condições, de produzir a percepção, a ação, a compreensão, o aprendizado e outros fenômenos intencionais. (TEIXEIRA, 1996, p. 71)

Portanto, não está no estabelecimento de regras formais que constituem um determinado programa a condição que deve ser preenchida para que sejam produzidos estados intencionais. Estes são resultados das determinações de natureza biológica que configuram os seres humanos. Searle não afirma que é impossível dotar-se máquinas de inteligência, mas sim que ainda não sabemos exatamente como se dá o surgimento desta.

Intencionalidade é a propriedade de muitos estados e eventos mentais pelos quais eles são direcionados para, ou sobre, ou de objetos e estados de coisas no mundo. Se, por exemplo, eu tenho uma crença, deverá ser a crença de que tal é o caso; se tenho medo, deverá ser medo de alguma coisa ou de algo que irá ocorrer; se tenho um desejo, deverá ser desejo de fazer algo ou de que alguma coisa aconteça ou seja o caso; se tenho uma intenção, ela deverá ser a intenção de fazer alguma coisa.

Tais estados são acompanhados de crenças e desejos, entretanto, estados não direcionados não são idênticos a crenças e desejos. Em minha abordagem, se um estado *S* é intencional, ele deverá responder a questões como: *S* é sobre o quê? De quê é *S*? O que é esse *S*? (SEARLE, 1983, p. 1)

A noção de *intencionalidade* para Searle se caracteriza como um movimento direcionado a elementos extrínsecos ao sujeito que a possui. É essa capacidade que permite aos seres humanos interferir no mundo e dar sentido ao que ocorre à sua volta. Por essa razão os estados mentais portadores de intencionalidade não são reflexivos, no sentido de se dirigirem a quem os possui. Isso decorre da estrutura biológica dos seres humanos, não se devendo a algo de natureza transcendental que estivesse num âmbito distinto dos acontecimentos naturais. Como ele próprio afirma, alguns estados mentais não são portadores de intencionalidade, como os estados de euforia, depressão e ansiedade, porque eles não possuem essa característica de *direcionalidade*.

Algumas realizações da inteligência humana não parecem ser o resultado do emprego exclusivo de operações computáveis. Um dos exemplos disso é o fato de nós podermos fazer estimativas acerca de algo que, num certo sentido, estaria fora do nosso alcance. Para nos convencermos da verdade de algumas proposições, não recorremos a

um processo de demonstração em que todas as suas etapas são ordenadas adequadamente, de tal modo, que ao final não teríamos outra coisa a fazer a não ser concordar com o resultado apurado.

O teorema de Kurt Gödel (1906-1978), o qual afirma que em qualquer sistema formal contendo como subsistema a aritmética de Peano-Dedekind em I ordem (AP), existem sentenças indecidíveis admitindo-se a consistência de AP, necessitando, portanto, de uma linguagem superior para realizar tal tarefa, é um caso em que, mesmo não tendo como demonstrar-se a verdade de uma proposição por meio de um procedimento dedutivo, isso não nos impede de a aceitarmos enquanto tal. O uso de nossa inteligência parece envolver fenômenos mais complexos do que somente a realização de cálculos computáveis e produção de demonstrações.

De acordo com Roger Penrose (1931), em *O grande, o pequeno e a mente humana* (1998), a inteligência humana somente será compreendida se avançarmos no conhecimento da complexidade de funcionamento dos componentes constitutivos dos neurônios.

Um componente primordial do cérebro é o seu sistema de neurônios. Uma parte importante de cada neurônio é uma fibra muito longa, conhecida como o seu axônio. Os axônios bifurcam-se em fios separados em vários lugares e cada um destes termina finalmente em algo chamado sinapse. Essas sinapses são as junções em que são transferidos sinais de cada neurônio para (sobretudo) outros neurônios, através de substâncias químicas chamadas neurotransmissores. Algumas sinapses são de natureza excitatória, com neurotransmissores que tendem a intensificar o disparo do próximo neurônio, e outras são inibitórias, tendendo a suprimir o disparo do próximo neurônio. Podemos referir-nos à confiabilidade de uma sinapse na transmissão da mensagem de um neurônio para o outro como a intensidade da sinapse. Se todas as sinapses tivessem intensidades fixas, o cérebro seria muito parecido com um computador. No entanto, é certamente verdade que essas intensidades sinápticas podem mudar e existem várias teorias acerca da maneira como mudam. (PENROSE, 1998, p. 136)

Uma investigação mais detalhada e minuciosa dos neurônios indica que existem muitas áreas escuras que precisam ser melhor esclarecidas se pretendemos compreender como ocorrem a produção da inteligência e da consciência. Assim como já havia feito em *Shadows of the mind* (1994), Penrose afirma que,

são os microtúbulos que as transportam (substâncias neurotransmissoras) pelos axônios e dendritos e, portanto, sua atividade influenciaria a concentração dessas substâncias químicas na extremidade do axônio e nos dendritos. Isso, por sua vez, poderia influenciar a intensidade da sinapse. Outra influência do microtúbulo estaria no crescimento e na degeneração do neurônio, alterando a própria rede de conexões neurônicas. (PENROSE, 1998, p. 141)

Embora tenhamos a possibilidade de compreender a consciência e o agir inteligente pelo que se descontina como um continente a ser explorado, que é o funcionamento do cérebro, tal esforço pode ser em vão se não nos for possível fazer uma abordagem que não seja meramente computacional do funcionamento desse órgão. Assim, um computador digital não teria condições de reproduzir os fenômenos que determinam o surgimento da inteligência.

O tipo de abordagem proposta por Penrose vai de encontro a proposta de Allen Newell e Herbert Simon. Esses acreditavam ser possível elaborar-se modelos de funcionamento das diferentes funções desempenhadas pelo intelecto na produção de conhecimento, modelos que seriam apresentados na forma de programas (*software*), posteriormente implementados e testados num computador.

A construção de uma máquina inteligente dependeria de uma pré-construção que teria por objetivo reproduzir, passo-a-passo, todos os processos envolvidos num comportamento inteligente. Esse tipo de abordagem pressupõe que seja concebida, antes, uma plataforma de testes, um modelo básico, que contenha os elementos envolvidos no desempenho das funções inerentes a uma ação inteligente. Contudo, mesmo a elaboração de um esquema detalhado do processo de produção da inteligência não é suficiente para que compreendamos como o mesmo ocorre. A tarefa de produzir inteligência artificialmente não é simples, pois não basta que saibamos identificar todos os elementos e suas relações envolvidos na sua produção.

Para contornar os problemas advindos da sua organização altamente complexa foi proposta a estrutura modular do cérebro. A ideia de *modularidade* é empregada para enfatizar que o cérebro possui diferentes partes, que executam diferentes funções, que, por sua vez, produzem diferentes resultados. Essa concepção foi proposta e apresentada por Jerry A. Fodor (1935) em *The Modularity of Mind* (1983), ela consiste na ideia de

que o pensamento resulta da execução de diferentes processos produzidos em diferentes partes da mente quando tratariam com conteúdos específicos da percepção.

A “psicologia das faculdades” está começando a ser respeitável novamente depois de séculos em companhia de frenologistas e outros tipos duvidosos. Por “psicologia das faculdades” quero significar, grosso modo, a visão de que muitos tipos fundamentalmente diferentes de mecanismos psicológicos devem ser postulados para se explicar os fatos da vida mental. (FODOR, 1983, p. 1, tradução nossa)

Podemos nos perguntar se Fodor propõe um modelo de estrutura da mente ou, um modelo de estudo dessa. Para ele, ambas estão intimamente relacionadas, ele é influenciado pelo pensamento de Descartes que apresenta um modelo de funcionamento da mente determinado pela operação de diferentes capacidades.

Como praticamente todos sabem, a doutrina das idéias inatas de Descartes se faz presente novamente e é explicitamente interpretada como uma teoria sobre como a mente está estruturada em faculdades psicológicas ou “órgãos”. Estou inclinado a vislumbrar este despertar cartesiano como um grande benefício. (FODOR, 1983, p. 3, tradução nossa)

De acordo com essa teoria, alguns problemas ligados à percepção são evidências de que diferentes módulos mentais funcionam de maneira independente uns dos outros. Uma ilusão seria a comprovação do funcionamento em módulos da mente, já que ela pode ser compreendida como o produto da desconsideração mútua dos diferentes resultados alcançados pelo funcionamento independente de cada módulo.

É fundamental na formulação desse tipo de abordagem o conceito de *faculdade mental*, pois a ideia de que a mente executa diferentes funções em diferentes módulos, exige que sejam discriminados os diferentes processos executados pela mente e seus respectivos setores de produção. Esta concepção do mental não comprehende a mente como uma coisa só, em que todas as suas partes executam de maneira indistinta qualquer uma das suas diferentes funções. O pensamento resultaria de uma elaborada rede de processos executados de forma independente e organizada.

A psicologia das faculdades leva a sério a aparente heterogeneidade do mental e está impressionada com tais diferenças *prima facie* como entre, digamos, sensação e percepção, volição e cognição, aprendizagem e memória, ou linguagem e pensamento. De acordo com a psicologia das faculdades, a causalidade mental do comportamento envolve, characteristicamente, a atividade simultânea de uma variedade de mecanismos psicológicos distintos, a melhor estratégia de pesquisa parece ser a de dividir e conquistar: primeiro estudar as características intrínsecas de cada uma das faculdades presumidas, depois, estudar as formas como elas interagem. (FODOR, 1983, p. 1, tradução nossa)

A concepção em módulos salienta o arranjo compartmentalizado de funcionamento da mente. Fodor destaca que o tema acerca da estrutura da mente possui diferentes abordagens que concordam em vários aspectos e divergem em outros tantos. O neocartesianismo, que é uma dessas abordagens, afirma que as faculdades mentais se distinguem pelos seus conteúdos proposicionais. Por sua vez, para os formuladores da estrutura horizontal da mente, as faculdades mentais não possuem conteúdos específicos e são capazes de lidar com conteúdos distintos. Para os propositores da estrutura vertical da mente, as faculdades mentais são capacidades comuns às pessoas e são invariáveis, possuem domínio específico e são independentes em seu funcionamento; enquanto para os associacionistas não existem faculdades mentais, o máximo que poderíamos falar seria na capacidade que a mente possui de estabelecer associações entre os seus diferentes conteúdos.

Fodor distingue três diferentes sistemas psicológicos ou mecanismos que atuam no funcionamento da mente: os transdutores, os sistemas de entrada e os processadores centrais. A descrição do funcionamento de cada um desses sistemas é apresentada como a demonstração do funcionamento compartmentalizado da mente, onde cada sistema opera com regras e procedimentos próprios. Por exemplo, é apontada como uma das características próprias dos sistemas de entrada a necessidade de realização dessa função uma vez em contato com o seu conteúdo.

O fato dos sistemas de entrada serem praticamente obrigados a serem aplicados sempre que forem possíveis é, quando pensamos nisso, bastante notável. Existem fortes razões para acreditarmos que, em geral, as relações computacionais que os sistemas de entrada mediam – grosso

modo, as relações entre transdutores de saída e percepções – são bastante remotas. (FODOR, 1983, p. 53, tradução nossa)

Assim, um módulo é definido como um sistema que possui conteúdo próprio e está articulado a outros sistemas sem serem dependentes uns dos outros. A perspectiva de trabalho da concepção modular é a de simplificar o funcionamento da mente por meio da sua decomposição em partes, resolvendo, assim, o problema em torno da sua complexidade. Dessa maneira, torna-se mais fácil a compreensão dos diferentes mecanismos e suas funções presentes na produção de pensamentos. Não se trata de um tipo de abordagem que abarca o mundo na sua totalidade, pelo contrário, ela o segmenta para, logo a seguir, recompô-lo num todo articulado.

Os novos rumos do projeto de IA.

Fodor critica a chamada concepção clássica do projeto de IA por entender que essa concepção fundamenta-se na ideia de que a inteligência humana poderia ser reproduzida artificialmente por meio de um sistema simbólico capaz de realizar as mesmas operações de pensamento que um ser humano e, com isso, alcançar a mesma capacidade de compreensão que esse. Segundo a concepção clássica do projeto de IA, seria suficiente que uma máquina demonstrasse ter a capacidade de fazer uso da linguagem como um ser humano para ser considerada cognitivamente semelhante a esse.

Entretanto, os pesquisadores em IA não conseguiram a realização de tal feito, porque o desafio de construir IA envolve questões mais complexas do que a construção de um dispositivo com capacidade de operar simbolicamente no mesmo nível que os seres humanos. Tal situação levou ao surgimento de novas formas de abordar os objetivos do projeto de IA que, ao invés de tratar os processos mentais como programas de computador, estabelece um paralelo entre os processos mentais e a atividade cerebral.

Antes de qualquer outra coisa, a capacidade de inteligência dos seres humanos está profundamente relacionada com as respostas de adaptação dadas às circunstâncias determinadas pelo ambiente. De acordo com a concepção clássica da IA, seria necessária

a elaboração de um programa que fosse robusto o suficiente para dar conta das situações inusitadas apresentadas pelo ambiente no qual os seres inteligentes se desenvolvem. Isso é o que os biólogos chamam de *plasticidade fenotípica*, a capacidade de lidar com situações que não são esperadas.

O modelo de abordagem proposto pela concepção clássica da IA entende que a resposta produzida pela *plasticidade fenotípica* é um processo que envolve a percepção do que ocorre. A compreensão dessa ocorrência, o delineamento de uma estratégia de resposta e a ação a ser executada. Esse modelo possui uma linearidade das atividades supostamente envolvidas na elaboração de uma resposta eficiente. Tal linearidade produz como consequência uma resposta lenta e ineficaz de acordo com determinadas circunstâncias apresentadas pelo meio. Os seres vivos não interrompem sua ação quando colocados em situações em que o que acontece representa algo inusitado e inesperado, não fazem uma parada para realizar o processamento da informação conforme concebido pela concepção clássica da IA. Isso não ocorre com os seres inteligentes que dão respostas rápidas aos eventos excepcionais, levando à conjectura de que talvez o modo de funcionamento da inteligência não se dê conforme a descrição do modelo clássico de IA.

Cada vez mais os pesquisadores de IA terão de investigar de maneira mais aprofundada as interações que se dão entre os organismos inteligentes e o ambiente no qual eles existem. A inteligência não é mais considerada como sendo meramente a capacidade de tratar formalmente com conteúdos informacionais. A inteligência faz parte de uma totalidade complexa que exige de nós um esforço de compreensão que seja capaz de descrever as diferentes funções e relações que se dão entre os elementos que constituem um ser inteligente e os elementos que integram o ambiente no qual ele vive.

Talvez o maior desafio aos pesquisadores do projeto de IA num futuro breve seja demover a opinião pública de que este projeto não representa um risco aos seres humanos. O número de abril de 2017 da *MIT Technology Review* traz um artigo bastante inquietante em que é relatado o modo de funcionamento de uma nova tecnologia empregada na substituição de motoristas humanos. O título do artigo é chamativo, *O lado escuro do coração da IA* (tradução nossa), mas não deixa de ser exato com o conteúdo da matéria.

No último ano, um carro com uma auto-condução estranha, trafegou pelas tranquilas estradas de Monmouth County, Nova Jérsei. O veículo experimental, desenvolvido pelos pesquisadores da fabricante de chips Nvidia, não parecia diferente de outros carros autônomos, mas era diferente do que qualquer coisa demonstrada pelo Google, Tesla ou General Motors, mostrando um aumento da capacidade da inteligência artificial. O carro não seguiu nenhuma instrução fornecida por um engenheiro ou programador. Ao invés disso, ele contou inteiramente com um algoritmo que lhe ensinara a dirigir ao observador um ser humano fazê-lo. (<https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ia> Tradução nossa)

As possíveis consequências éticas e jurídicas deverão ocupar maior destaque no tocante às questões relativas aos objetivos do projeto de IA. Mais preocupante do que as possíveis consequências sobre a condição humana que essa nova tecnologia poderá produzir é a falta de conhecimento necessário à explicação de como se dá o funcionamento do carro inteligente da *Nvidia*. Os pesquisadores que o desenvolveram não sabem explicar como o carro aprende, o que eles fazem é relacionar diferentes tecnologias e dispositivos cibernéticos esperando pelos resultados advindos das interações entre os componentes do dispositivo (carro) e o conjunto de informações colocado à sua disposição.

Isso provoca questões interessantes. Conforme a tecnologia avança, poderemos ultrapassar um certo limite, além do qual o uso de IA requererá um salto de fé. (<https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ia>, tradução nossa)

Turing já havia cogitado a possibilidade de que os engenheiros e demais pesquisadores envolvidos no projeto de IA não necessitariam ter de explicar como se daria o funcionamento do computador que, por ventura, viesse a passar no jogo da imitação. Não podemos esquecer que o projeto de IA está inserido, também, num ambiente fortemente experimental.

Vimos, nesta breve exposição da história do projeto de IA, que os problemas e questões relacionados aos objetivos desse projeto são oriundos de diferentes áreas de conhecimento. De algum modo, todas essas questões convergem para o ser humano, o

que faz com que seja necessário se dar atenção às questões inerentes a esse se pretendemos reconhecer aquilo que lhe é próprio. Sem a compreensão do que consiste a atividade inteligente, de como e por que se dá uma decisão, do que ocorre quando empregamos símbolos, do que seja o próprio homem. Sem compreender essas questões de natureza filosófica, ficamos sem as condições de reconhecer a inteligência que possa, por ventura, vir a ser produzida artificialmente.

A Filosofia tem, portanto, um papel relevante na formulação de respostas à questão acerca da admissibilidade do projeto de IA. Esse projeto não somente deu uma nova direção a várias áreas de pesquisa e desenvolvimento científico-tecnológico, ele se apresenta como um desafio à própria Filosofia ao exigir-lhe a apresentação de um cenário conceitual e teórico que envolva as questões e problemas inerentes aos seus objetivos, produzindo, com isso, uma aproximação entre a reflexão filosófica e as demais áreas de conhecimento.

CAPÍTULO 2

INTELIGÊNCIA ARTIFICIAL: ILUSÃO OU POSSIBILIDADE?

No primeiro capítulo fizemos uma breve exposição de alguns dos momentos mais importantes que constituem o *background* do projeto de IA. Procuramos mostrar com isso que, desde os seus primórdios, o projeto de IA relaciona diferentes problemas e áreas de conhecimento. Essas condições favoreceram o surgimento de propostas de investigação e pesquisa com características próprias de abordagens e propósitos, como podemos ver na distinção entre IA fraca e IA forte. Acreditamos que a visão de que as diferentes linhas de pesquisa e investigação relacionadas ao objetivo do projeto de IA devem competir entre si prejudica o esforço de conhecimento engendrado pelo mesmo. Construir um ambiente em que as noções de complementaridade e colaboração concorram para o desenvolvimento do conhecimento e modo de solução de problemas talvez represente um dos maiores desafios a ser superado na realização dos objetivos do projeto de IA. Tal projeto, tem na sua face científico-tecnológica o seu aspecto mais notório, induzindo-nos a pensar que o mesmo é estritamente dessa natureza. Contudo, a outra face, que é de natureza filosófico-especulativa, embora seja menos notável, é tão importante quanto a primeira. Como veremos, os papéis dessas diferentes áreas de pesquisa no tocante ao reconhecimento de inteligência artificial são bem distintos, pois o contexto de justificação não é o mesmo que o da pesquisa estrita. O aporte conceitual das disciplinas filosófico-especulativas é necessário como fundamento que justifica as diferentes abordagens com fins a atingir os objetivos desse projeto.

No presente capítulo procuramos discorrer sobre algumas das diferentes abordagens de pesquisa relacionadas ao projeto de IA; a influência do modo de pensar dicotômico presente nas suas diferentes concepções; a proposta de Turing para vencer a barreira dicotômica e uma crítica ao dualismo dicotômico que apresenta uma proposta de trabalho convergente, envolvendo diferentes áreas de conhecimento. Por fim, estabelecemos uma comparação entre a Alquimia e o projeto de IA a fim de mostrar nossa hipótese de trabalho que indaga sobre a admissibilidade do projeto de IA.

O caráter divergente das pesquisas em IA

O evento que precipitou o surgimento da *máquina de estado discreto* de Turing foi a necessidade de substituição da *mente-de-obra-humana* empregada na realização de cálculos que tornassem mais precisa a artilharia, desvendassem as linguagens em código criptografado e possibilitassem o desenvolvimento da bomba atômica durante a Segunda Guerra Mundial. Essas atividades demandaram o emprego de muitos matemáticos, necessários à realização dos cálculos. Embora muitas regras, algoritmos, equações e fórmulas tenham sido descobertas por eles na realização dessas operações, ainda assim, a demanda por mais seres humanos capazes de realizar cálculos complexos era um fator de limitação quanto ao alcance dos resultados. Diferente do engodo que foi a *máquina de jogar xadrez* do barão Von Kempelen, a *máquina de estado discreto* de Turing é uma máquina capaz de realizar, de fato, cálculos e propor soluções para problemas.

O objetivo último do projeto de IA é a construção de um *novo homem*, um ser engendrado pelo conhecimento e domínio de tecnologias. Esse desafio, uma vez lançado, foi aceito por vários e distintos desafiantes. Como é um continente a ser descoberto, as propostas de descobrimento são de naturezas e abordagens específicas e particulares. As etapas de realização desse empreendimento parecem indicar que, num primeiro momento, a construção do *novo homem* depende do próprio homem, que é quem busca conhecimento e desenvolve tecnologias com vistas a alcançar tal propósito.

Porém, já existem realizações feitas pelos computadores que não teriam como ser executadas por um ser humano. Isto é, parece haver um ponto na linha de desenvolvimento dessa construção especial, em que o elemento humano perderá preponderância com os resultados práticos alcançados, sendo, por si mesmos, suficientes para garantir o avanço em direção ao propósito último em vista. Um exemplo disso foi o sucesso dos primeiros programas de cálculo que resolveram vários paradoxos matemáticos traçando uma estratégia de resolução própria, ou seja, agiram como se fossem matemáticos humanos.

Por ser uma área de conhecimento necessariamente multidisciplinar, não existe um acordo sobre qual metodologia de pesquisas e estudos é a mais adequada, tendo em vista

o objetivo último do projeto de IA. Enquanto alguns grupos de pesquisadores se dedicam a desenvolver uma linguagem de programação que possua o mesmo grau de complexidade e interação que as linguagens naturais; outros preferem se dedicar ao estudo do cérebro para desenvolver uma máquina que seja capaz de realizar as mesmas atividades que este. Bem como, há os que veem no desenvolvimento da eletrônica nossa única chance de conseguir algum resultado satisfatório, sem falarmos nos que preferem desenvolver métodos e teorias que visem a compreensão da psicologia inerente a uma estrutura pensante.

O projeto de IA abriga abordagens que são excludentes à primeira vista, por tratar-se, também, de uma área recente de interesse do conhecimento. Não há um objeto definido, ainda está em discussão se o pensamento é produto do cérebro ou da mente (*psiquê*); ainda não foi definido um conjunto de princípios e procedimentos considerados os mais adequados para tratar do problema. Por exemplo, para a *psicologia cognitiva* não é suficiente dispor-se de uma tecnologia que dê suporte à construção de uma máquina inteligente, é necessário, isto sim, que sejamos capazes de compreender como é produzida uma ação inteligente, o que demandaria um conhecimento profundo da *psiquê* humana. Os *físicalistas*, por sua vez, afirmam que o pensamento não passa do resultado das interações da matéria.

Assim que as duas possibilidades de abordagem do projeto de IA foram indicadas, inúmeras pesquisas foram apresentadas oriundas de diferentes áreas do conhecimento e com objetivos específicos derivados de uma ou de outra das duas abordagens. Cada uma, por sua vez, tentando dar conta do entendimento de parte da complexidade que é a produção de pensamento. Os objetivos derivados dessas duas formas de abordagem dos objetivos do projeto de IA compreendem áreas de conhecimento que vão desde a eletrônica, informática, química, física e cibernética, passando pela biologia, neurociências e ciências médicas, até a antropologia, filosofia, ciências sociais e psicologia.

Não existe uma teoria unificadora que seja capaz de articular todas essas distintas especialidades. A efetivação de tal tarefa poderá promover a otimização dos esforços que visam atingir o objetivo do projeto de IA. Esse é um campo de investigação que merece

ser melhor definido, cujos propósitos, objeto e métodos ainda se encontram em estado de latência. A IA percorre um caminho que pode ser descrito como o limbo entre o espiritual e o concreto, um espaço em que a imprecisão das nossas observações é o principal aspecto daquilo que nele acontece.

A influência da dicotomia e sua origem cartesiana

Em nossa opinião, esta discussão em torno da admissibilidade ou não do projeto de IA não avança porque ela está delimitada por arranjos dicotômicos, enquanto um dos membros que forma o par-de-contradição afirma ser o pensamento um fenômeno imaterial, o outro afirma o contrário. Pares dicotômicos como ‘natural/artificial’ e ‘físico/não-físico’ parecem nos revelar que, no nível macro ontológico, a estrutura existente é a mesma que no nível micro ontológico, como se fosse um cristal. O que podemos ver, quando tomamos conhecimento do debate em torno da IA, são posições que chegam, às vezes, a se negarem de forma totalmente incompatível uma com a outra. As razões apresentadas a favor de uma são exatamente aquelas que se contrapõem à possibilidade de aceitação da outra. Entendemos que a saída para esse tipo de situação é não aceitarmos a condição de que a discussão se circunscreva aos limites impostos pelo par antagônico.

Esse cenário de antagonismo é fortemente determinado pela forma cartesiana de conceber o homem. Conforme a metafísica cartesiana, os seres humanos são constituídos por dois tipos de substâncias que realizam atividades não intercambiáveis e nem relacionáveis.

De acordo com Paul M. Churchland (*The Ontological Problem*, 2013), a tese do *dualismo de substâncias* afirma que o pensamento é uma atividade não-física, imaterial, não realizada pelo corpo, mas pelo espírito, pela alma racional, pelo intelecto.

Descarte argumentou que a realidade se divide em dois tipos básicos de substância. A primeira é a matéria física, cuja característica essencial desse tipo de substância foi dito ser a *extensão no espaço*: qualquer instância da substância material possui altura, peso e largura, e ocupa

uma determinada posição no espaço. Ela a chamou de *res extensa* (substância extensa).

Descarte não tentou diminuir a importância deste tipo de matéria. Pelo contrário, ele foi um dos físicos mais imaginativos de seu tempo, sendo, também, um entusiástico advogado do que foi então chamado de ‘mecanicismo’. Contudo, havia um isolado canto da realidade que ele pensava não ser possível de ser abordado nos termos mecanicistas: a razão consciente dos homens. Esse foi o motivo para propor um segundo e radicalmente diferente tipo de substância, uma substância que não possui extensão ou posição espaciais, uma substância cuja característica essencial é a atividade do pensamento. Ele chamou este segundo tipo de matéria de *res cogitans* (substância pensante). Esta visão é conhecida como ‘dualismo cartesiano’. (CHURCHLAND, 2013, Kindle: posição 60, tradução nossa)

Seguindo essa linha de raciocínio, somos levados a concluir que a estrutura do mundo é ordenada por meio de relações de oposição e negação. Ao postular a existência de dois tipos de substâncias constitutivas do mundo, Descartes criou um mundo dividido, cuja integração das partes é impossível de ser concebida. Como sabemos, fica difícil a esta doutrina explicar fenômenos que exibem uma harmonia entre si, cujas origens são fontes que não se comunicam. Acreditamos que muitas das dificuldades enfrentadas por aqueles que tentam elaborar um modo de reconhecimento de inteligência artificial tem origem nesse tipo de pensamento, que se caracteriza por conceber o mundo como o resultado de uma divisão.

Como dizia Descartes, o *seu* real não é o seu corpo material, pelo contrário, é uma substância pensante não-espacial, uma unidade individual de matéria mental bastante distinto do seu corpo material. (CHURCHLAND, 2013, Kindle: posição 60, tradução nossa)

O dualismo de Descartes propiciou o surgimento de duas doutrinas antagônicas acerca da produção de pensamento, o *dualismo de propriedades* e o *materialismo* (Searle, 1997). Para os adeptos da primeira doutrina, o problema mente-corpo é praticamente impossível de ser solucionado por tratar-se de algo muito complexo e difícil. Já para os adeptos da segunda doutrina, embora não deixe de considerar o problema mente-corpo difícil e complexo; vê, na redução da *consciência* e da *intencionalidade* a fenômenos físicos, a possibilidade de solução para o mesmo.

Para Searle, esse tipo de contraposição não auxilia em nada a pavimentação de uma saída deste atoleiro conceitual que é a *dicotomia mente-corpo*.

Acredito que os dois lados estão completamente enganados. Ambos aceitam um determinado vocabulário e, com ele, um conjunto de hipóteses. Pretendo mostrar que o vocabulário é obsoleto, e que as hipóteses são falsas. É fundamental mostrar que tanto o dualismo quanto o monismo são falsos porque em geral se supõe que esgotam o campo, não deixando outras opções. (SEARLE, 1997, p. 9)

A posição de Searle a respeito da dicotomia *materialismo/dualismo* é a de desconsiderar a forma como são relacionados os dois elementos que formam o par dicotômico, pois os julga como sendo falsos e errados. Pensamos haver um exagero na sua desconsideração por essas hipóteses de compreensão da atividade do pensamento. Embora seja difícil obter-se uma demonstração da imaterialidade do pensar, isso não pode ser tomado como uma prova de que ele não seja dessa natureza. Contudo, concordamos com a avaliação de que o *materialismo* e o *dualismo de substâncias* não são as únicas alternativas teóricas possíveis para compreender-se a atividade intelectual humana.

Talvez o erro do *naturalismo biológico*, como é chamada a opção concebida por Searle para superar a dicotomia acima apontada e que consiste em afirmar que “os processos e fatos mentais fazem parte de nossa história biológica tanto quanto a digestão, a mitose, a meiose ou a secreção enzimática” (Searle, 1997, p. 7), seja o de se colocar como um dos elementos formadores de um par dicotômico, pois exclui como falsas as outras possibilidades de compreensão que se apresentam como alternativas.

A tentativa de ruptura com a forma dicotômica de organização de ideias não consiste em se propor uma terceira opção que não só desfaz o par, como, também, elimina seus membros. Isso seria incorremos no equívoco que pretendemos evitar.

A busca do incorrigível é um dos mais veneráveis pesadelos da história da filosofia. Invadindo toda a filosofia antiga, notoriamente Platão, foi poderosamente revigorada por Descartes, e por ele legada a uma extensa linhagem de sucessores. Não há dúvida de que possui muitos motivos e assume múltiplas formas, mas, naturalmente, não podemos nos estender aqui sobre esta questão. Em alguns casos o motivo parece ser um anseio relativamente simples por algo *absolutamente certo* – anseio que pode ser de difícil satisfação se o manipularmos de tal modo

que a certeza se torne absolutamente inatingível; em outros casos, como em Platão talvez, o que se procura aparentemente é algo que será sempre verdadeiro. (AUSTIN, 1997, p. 139)

Nesta tese, procuramos sustentar a ideia de que não é necessária a adoção de uma abordagem que elimine o que for divergente. Propomos um tratamento da questão em que as diferentes perspectivas de tratamento do problema são levadas em consideração, dando-nos, assim, condições de superar o impasse gerado pelo arranjo dicotômico das mesmas. Entendemos que uma abordagem teórica que vise a explicar o fenômeno do pensamento, da consciência ou da atividade intelectual, só poderá representar algum avanço se pressupor a contribuição das diferentes formas de tratamento do problema. A atitude reducionista somente reforça o impasse produzido pelo modo dicotômico de tratamento das questões.

A proposta turingeana de superação da dicotomia

Ao propor o *jogo da imitação*, Turing apresentou um critério que serviria de distinção entre uma ação inteligente e outras formas de ação. O critério proposto seria a capacidade de uso competente de uma linguagem natural no desenvolvimento de uma conversação.

O jogo da imitação consiste numa disputa em que participam cinco jogadores, dos quais dois devem responder a perguntas feitas pelos demais três jogadores, chamados juízes. Um dos jogadores que responde é um computador digital programado para participar do jogo como se fosse um jogador humano. O propósito é conseguir ter uma atuação tão humanamente convincente que os juízes não conseguiram distinguir a atuação do jogador máquina da do jogador humano.

Turing, ao nomear o jogo, de certo modo, delimitou o escopo do projeto de IA, ou o que ele julgava ser o alcance dessa nova área do conhecimento. Talvez devêssemos levar em consideração sua prudência ao escolher o jogo concebido para o reconhecimento de inteligência e, descartar, de antemão, a versão forte do projeto de IA.

Não é isso que nos propomos, ao contrário, pretendemos mostrar com este trabalho que a tese da não admissibilidade da versão forte do projeto de IA enfrenta dificuldades de tal ordem, que a impedem de ser tomada como a maneira correta de entendimento da questão.

Na língua inglesa, segundo o *Dictionary Cambridge*, o termo *imitation* é um adjetivo que significa: a ação de fazer com que algo pareça com outra coisa que ele não é; gerar uma aparência que não é real, como no exemplo: *an imitation leather jacket* (uma jaqueta de couro de imitação). Não há um sentido de *imitação* que nos dê a entendê-la como possibilidade de uma reprodução original, na verdade, parece tratar-se mais de um procedimento de simulação do que propriamente uma produção original. *Imitar* não significa a realização de algo genuíno, já que possui um caráter secundário ao derivar de algo, em função do qual procura assemelhar-se.

O termo “imitar” também possui o sentido de plágio, dando-nos a entender que não há grande mérito na execução dessa atividade. Pode-se cogitar, a partir desse sentido, que Turing não pensou que as máquinas seriam capazes, algum dia, de pensar exatamente como um ser humano, mas sim de realizar algo que, no máximo, teria alguma semelhança com o pensar humano, a ponto de não ser fácil reconhecer quando se trataria de algo realizado por um ou pelo outro.

De acordo com o *Dicionário Houaiss*, na língua portuguesa encontramos, pelos menos, cinco significados distintos de “imitar”:

- 1 reproduzir ou tentar reproduzir fielmente (o que foi feito por outrem ou as características de alguém ou algo); arremediar
Ex.: i. um estilo, um sotaque
- 2 procurar reproduzir o estilo de; inspirar-se em; ter como exemplo; copiar
Ex.: <o aprendiz imita o mestre><as figuras cubistas de Picasso imitam as máscaras africanas>
- 3 produzir com intenção de fazer passar a cópia por verdadeira, como original; falsificar, plagiar
Ex.: <i. uma assinatura><ele, além de se inspirar na obra do outro, imitou-a>
- 4 ter falsa aparência de; assemelhar-se a; lembrar
Ex.: tecido plástico imitando couro
- 5 Derivação: por metáfora.

Traduzir com fidelidade; reproduzir, repetir
Ex.: a arte imita a vida

“Arremedar” é um sentido que não é totalmente incompatível com o objetivo pretendido por Turing ao elaborar o jogo da imitação, visto que “arremedar” significa: “tentar reproduzir; ter algo de; lembrar; assemelhar-se”. Contudo, “arremedar” possui, ainda, significados que, podemos afirmar com segurança, não são compatíveis com o objetivo do jogo, como: “imitar de forma caricatural; simular; fingir; executar mal uma determinada atividade”. Isto faz com que esse significado de “imitação” não seja o mais apropriado de ser atribuído às pretensões de Turing na elaboração do jogo, pois o arremedo é uma forma de repetição que se caracteriza como plágio, um ato inautêntico copiado de outro.

Falsificar e *plagiar* parecem ser, dentre todos os significados os menos adequados de ser atribuído à proposta do jogo da imitação. Em várias passagens do artigo em que faz tal proposição, Turing nos adverte que o objetivo último do jogo não é a produção de um simulacro, mas, isto sim, de desenvolvimento de um procedimento de teste que servisse para avaliar e aferir o quanto o desempenho da máquina de estado discreto pode ser, genuinamente, indistinto do humano. Além disso, esse significado possui uma conotação negativa, pois é inerente ao ato de falsificar o procedimento, de fraudar e de adulterar, o que, certamente, não é o que Turing pretende alcançar ao propor o projeto da máquina de estado discreto e sua participação no jogo da imitação.

O sentido de “falsa aparência” não é compatível com o jogo de Turing, considerando o que ele mesmo afirma:

o método de pergunta e resposta parece ser adequado para ser usado em quase todos os campos de atividade que desejemos abranger. Não queremos punir a máquina por sua inabilidade de brilhar em concursos de beleza, nem punir um homem por perder uma corrida contra um aeroplano. As condições de nosso jogo tornam essas inaptidões irrelevantes. (TURING In: TEIXEIRA, 1996, p. 23)

O objetivo do jogo não é criar condições que concorram para a produção de um erro de julgamento ou engano, como seria no caso de uma performance ilusionista. O que

Turing propõe é o estabelecimento de um procedimento de identidade para reconhecer a portabilidade de uma mesma razão. Ele não nos apresenta um conjunto de conceitos, princípios, método e descrição do que ele entende como algo inteligente e que sirvam para ser aplicados universalmente. Seu critério é empírico, não é explicativo. Se não formos capazes de distinguir o desempenho do homem e o do computador na prática de uma atividade, isto significa dizer que ambos são portadores das mesmas capacidades inerentes ao desempenho dessa atividade.

O sentido metafórico de *imitar*, significando *traduzir com fidelidade; reproduzir; repetir*, possui compatibilidade com o sentido do jogo. Contudo, tal compatibilidade não é plena, por tratar-se de uma metáfora. Ao propor o jogo da imitação, Turing determina um âmbito prático da resposta à pergunta sobre a possibilidade de máquinas virem a pensar, ou seja, ele não pretende meramente apelar ao recurso de *por assim dizer*.

No nosso entendimento, o sentido mais apropriado de ser aplicado ao objetivo do jogo da imitação é o de *copiar; inspirar-se em; reproduzir o estilo; ter como exemplo*. Entendemos que Turing não está propondo uma forma de reconhecimento do homem, mas, simplesmente, do exercício de uma das suas tantas faculdades, nesse caso, a capacidade de se comunicar. O sentido de *inspirar-se em* pode ser observado quando Turing afirma que “... [quer] excluir da categoria de máquina todos os homens de carne e osso”. (TURING In: TEIXEIRA, 1996, p. 24). Está claro que, na concepção de Turing, há uma distinção entre homem e máquina que insere esta última numa condição de dependência do primeiro. Dessa maneira, a participação do computador no jogo da imitação não seria outra coisa que não a repetição daquilo que um ser humano é capaz de produzir de forma genuína. Não devemos, portanto, esperar do procedimento apresentado por Turing uma explicação do que seja o pensamento e a inteligência, a qual nos assegure a validade do resultado da sua aplicação.

Turing não deixou de levar em consideração e responder às várias objeções levantadas contra o jogo proposto por ele como procedimento eficaz de reconhecimento de inteligência. Entretanto, duas respostas merecem ser destacadas, por mostrarem que Turing estava mais interessado em avaliar o desempenho do computador digital no desenvolvimento de atividades inteligentes, do que propriamente desvendar os mistérios

acerca do funcionamento da alma, da mente e do intelecto humanos. Ele possui um parâmetro para reconhecer o uso de inteligência, porém, não possui uma explicação de como isso se dá.

A primeira resposta, aqui destacada, é dirigida à crítica ao jogo chamada *objeção matemática*. Essa crítica é um corolário do *Teorema de Gödel* ao afirmar que todo sistema simbólico possui proposições acerca das quais o próprio sistema não tem como aferir seu valor de verdade sem que caia em contradição.

Turing responde à *objeção matemática* em duas etapas. Na primeira ele afirma que nós humanos somos incoerentes, pois cobramos da máquina aquilo que não somos capazes de cumprir. O fato das máquinas possuírem limites no que diz respeito à capacidade de responder satisfatoriamente a perguntas de um determinado tipo, não nos permite inferir que elas não sejam inteligentes. É uma espécie de preconceito ou arrogância assumirmos que os computadores digitais sejam dotados de uma ignorância insuperável. Turing atribui isso ao fato dos computadores digitais responderem a questões cujas respostas possíveis nós sabemos, de antemão, serem erradas ou nulas. De acordo com Turing, isso faz com que nos sintamos superiores à máquina.

No entanto, esse sentimento de superioridade pode não passar de um engano, pois, o fato de considerarmos o computador digital “ingênuo”, já que sabemos que eles irão errar em determinada situação, não nos permite inferir que sejamos superiores a qualquer um dos computadores digitais. Nossa juízo só poderá se estender àquelas máquinas para as quais a pergunta capciosa for formulada.

[...] só nos podemos sentir superiores, nessa ocasião, relativamente à máquina específica sobre a qual conquistamos nosso insignificante triunfo. Não se trata simultaneamente de triunfar sobre todas as máquinas. Em resumo, haverá talvez homens mais inteligentes do que qualquer máquina dada, mas mesmo assim poderão existir outras máquinas mais inteligentes, e assim por diante. (TURING In: TEIXEIRA, 1996, p. 39)

Assim como há diferença no grau de inteligência entre indivíduos humanos, Turing acredita que o mesmo poderá ocorrer entre as máquinas.

Na segunda etapa da resposta à *objeção matemática*, Turing faz uma crítica contundente à ideia de que todo ser humano possui inteligência, ao afirmar que, sem prova alguma, julgamos não estarmos circunscritos a esse mesmo limite que se impõe às máquinas, vencemos o solipsismo entre nós, seres humanos, não por uma comprovação da invalidade dessa postura, mas por um acordo de convivência. Ou seja, não somos tão exigentes conosco mesmos quando se trata da aceitação de que nós, os seres humanos, somos portadores de inteligência.

Uma segunda crítica dirigida ao jogo da imitação é o chamado *argumento da consciência*. Esse argumento Turing atribui a Geoffrey Jefferson (1886-1961), importante neurologista e pioneiro neurocirurgião formado em Manchester, que viveu no Canadá e teve atuação destacada nos esforços britânicos da I Guerra. Professor Jefferson, como é citado por Turing, publicou um artigo no *British Medical Journal*, em 1949, intitulado *The Mind Of Mechanical Man*. Jefferson parte do pressuposto de que a característica marcante do ser humano é o seu desejo por conhecimento, que se expressa de forma incisiva na busca pelo conhecimento da natureza da relação entre a mente e o cérebro. Seu artigo pretende mostrar que é essa peculiaridade da estrutura ontológica humana que as mentes das máquinas não alcançariam.

Somente quando uma máquina puder escrever sonetos ou compor concertos como resultado de pensamentos ou emoções sentidas, e não por via de ocorrência casual de símbolos, é que concordaríamos que a máquina se iguala ao cérebro – isto é, que não apenas os escreveu ou compôs como também sabia que os escrevera. Nenhum mecanismo poderia experimentar (e não meramente assinalar de modo artificial, por meio de uma engenhoca fácil) prazer pelos seus êxitos, tristeza quando suas válvulas queimam; deleite ante a lisonja; sentir-se infeliz por causa de seus erros, encantar-se com o sexo, ficar irritado ou deprimido por não poder alcançar o que deseja. (TURING In: TEIXEIRA, 1996, p. 39)

Embora Turing diga não considerar o professor Jefferson um solipsista, não deixa de imputar esse predicado ao argumento apresentado por este, pois a prova aceita para atribuir inteligência a uma máquina consiste em experimentar ser esta própria máquina, o que não é possível. Mais uma vez, não adotamos o mesmo grau de exigência para reconhecermos inteligência em outros seres humanos. O tratamento que Turing dá ao

problema do reconhecimento de IA é totalmente pragmático. É na medida em que as máquinas se mostrarem capazes de realizar determinadas coisas que nós, seres humanos, fazemos no dia-a-dia, que teremos condições de reconhecer se essas máquinas são portadoras de capacidades semelhantes às nossas. Ele não argumenta que é necessário, primeiro, desenvolver-se um conceitual e métodos que possibilitem a construção de teorias mais robustas para daí, então, termos condições de fazer tal reconhecimento. O jogo da imitação é suficiente para alcançarmos essa condição. Mesmo que não disponhamos de uma justificativa teórica que sustente a afirmação de que algo possui inteligência, isso não nos incapacita de fazê-lo, pois o reconhecimento de inteligência deverá se dar pela interação entre seres inteligentes.

No entanto, Turing não deixa de reconhecer o caráter de mistério da consciência, devido à forma paradoxal como é apresentada e pela dificuldade em defini-la. Isso pode nos servir de indício do quanto ainda precisamos nos desenvolver teoricamente para superarmos tais dificuldades.

Não quero dar a impressão de que penso não existir nenhum mistério no que diz respeito à consciência. Existe, por exemplo, algo assim como que um paradoxo vinculado às tentativas de localizá-la. Mas não acredito que tais mistérios tenham de ser necessariamente resolvidos antes de podermos responder à pergunta que nos preocupa neste artigo. (TURING In: TEIXEIRA, 1996, p. 41)

Ao reconhecer o caráter misterioso da consciência, Turing acentua a importância do seu procedimento de reconhecimento de inteligência. Onde não existe teorização suficiente o procedimento empírico é a alternativa de uso.

A crítica de John Searle ao pragmatismo de Alan Turing

Ainda que se estabeleça uma espécie de debate entre Turing e Searle sobre o reconhecimento de inteligência artificial, ele, de fato, não houve. Turing cometeu suicídio em 1954 após sofrer dura violência por parte do estado inglês que considerava crime a homoafetividade, obrigando-o a escolher entre a prisão ou a castração química. Ele teve

a coragem de não dar preferência a nenhuma das duas alternativas que lhe foram apresentadas. Searle, por sua vez, publicou *Minds, Brains and Programs* (1984), livro em que apresenta sua crítica à concepção forte da IA, 30 anos após a morte de Turing. Esse debate possui intervalos diferentes e interlocutores distintos conforme a época. A posição de Turing não deixou de ser defendida, mesmo após a sua morte. Embora seja uma argumentação com características próprias e distintas das de Turing.

Se a primeira resposta dada por Searle foi dirigida de forma crítica ao *jogo da imitação* como teste suficiente de inteligência, a réplica foi apresentada por Daniel Dennett. Este, por sua vez, foi interpelado por Searle, resultando num debate acalorado e intenso. Se entre a publicação do texto de Turing e a resposta de Searle houve um intervalo de mais de três décadas, no debate que se seguiu entre Searle e Dennett o intervalo entre réplica e réplica foi de meses, culminando com a interrupção abrupta do mesmo.

Searle apresenta três argumentos que criticam a eficiência do teste de Turing em avaliar a presença de inteligência artificial. De acordo com ele, a IA produziu a falsa aparência de conseguir atingir seu objetivo, devendo “...ter parecido – e para muitas pessoas ainda parece – reproduzir e explicar fenômenos mentais, e, acredito que não conseguiremos remover estas ilusões até que tenhamos exposto as razões que as originaram”. (TURING In: TEIXEIRA, 1996, p. 89)

A primeira razão que auxilia na produção dessa ilusão, segundo Searle, deve-se à confusão no uso dos significados da expressão *processamento de informação*. Essa confusão ocorre porque, em sua opinião, é estabelecida uma correlação entre o que é feito pelo cérebro humano e a sua mente com o computador e a sua programação. Assim como a mente e o cérebro humanos realizam *processamento de informação*, também o fazem o computador e seu programa. Dois significados distintos dessa expressão são apontados por Searle: o primeiro, quando essa expressão é referida a pessoas; o segundo, quando ela é aplicada a computadores.

A diferença entre um processamento e outro está no fato de que nós, seres humanos, compreendemos os significados dos símbolos empregados num processamento de

informações, as máquinas, não. Embora as máquinas sejam capazes de realizar operações com símbolos de forma eficiente, elas nada sabem a respeito dos seus significados.

[...] o computador tem sintaxe mas não tem semântica. Dessa forma, se você digita: $2+2$ igual?, ele vai apresentar “4”. Mas ele não tem ideia de que “4” significa 4, ou que isto signifique alguma coisa. O ponto não é que ele não tenha uma informação de segunda ordem acerca da interpretação de seus símbolos de primeira ordem, mas o fato é que estes símbolos de primeira ordem não têm nenhuma interpretação no que diz respeito ao computador. Tudo o que ele tem são mais símbolos. Assim sendo a introdução da noção de “processamento de informação” produz um dilema: ou bem construímos a noção de “processamento de informação” de tal maneira que ela implique a intencionalidade como parte do processo, ou bem nós não o fazemos. (SEARLE In: TEIXEIRA, 1996, p. 89)

Esse argumento servirá de fundamento para a proposição do *argumento do quarto chinês*, experimento mental proposto por Searle para demonstrar que a falta de compreensão do significado dos símbolos empregados num determinado processamento de informações não impede que esse processamento seja realizado de modo eficiente.

A segunda razão que concorre para a produção da ilusão gerada pela IA é o que Searle chamou de *behaviorismo residual ou operacionalismo*. O Behaviorismo faz parte da história da IA, como vimos o capítulo anterior, ele é um método de investigação que consiste na interpretação dos diferentes comportamentos, tendo como ponto de partida um conjunto robusto de observações. Esse método procura explicar o comportamento humano sem presumir a existência de entidades *não observáveis*. De acordo com tal método, não haveria a possibilidade de compreender-se corretamente o comportamento humano sem levarmos em consideração os *dados observáveis* que nos são apresentados durante o seu desempenho.

O Behaviorismo foi uma das primeiras contribuições da ciência no esforço de desenvolvimento de IA. Como a IA visa, dentre outras coisas, construir inteligência tendo como modelo o comportamento de agentes inteligentes, o behaviorismo forneceu alguns ingredientes básicos necessários para que tal objetivo seja atingido, ao apontar um meio de acesso aos fenômenos internos responsáveis pelo comportamento.

O terceiro elemento de sustentação da ilusão de explicar e reproduzir os fenômenos mentais através do uso de computadores digitais é o que ele chama de *forma residual de dualismo*. Para Searle a afirmação de que máquinas podem ser dotadas de inteligência, pressupõe uma distinção fundamental entre mente e cérebro. É baseada nessa distinção que a IA forte sustenta a possibilidade de alcançar o seu objetivo. De acordo com ele, a IA forte não dá importância ao suporte material que realiza as operações de pensamento, já que essa concepção entende que está nos programas a causa da inteligência. Nesse caso, as máquinas só teriam importância na medida em que fossem capazes de rodar programas inteligentes. Ainda segundo essa concepção, é possível que uma mesma operação de pensamento possa ser realizada por diferentes dispositivos computacionais, "... de fato, no que diz respeito a IA um mesmo programa pode ser realizado por uma máquina eletrônica, uma substância mental cartesiana ou o espírito do mundo hegeliano". (SEARLE In: TEIXEIRA, 1996, p. 91)

Para Searle a IA forte nada mais é do que o resultado da confusão de sentidos de uma expressão quando aplicados a diferentes significantes, da aplicação de um modo equivocado de explicação da atividade do pensamento e pela presença de um vício ontológico, o dualismo. Embora tenha apresentado uma longa argumentação para justificar sua descrença na capacidade de computadores digitais serem portadores de inteligência, não deixou de fazer um mergulho profundo no jogo da imitação de Turing para mostrar que não há pensamento algum empregado na realização das jogadas feitas pelo jogador máquina.

O quarto chinês descreve o interior do jogador máquina que participa do jogo da imitação. Trata-se de uma alegoria, criada por Searle, para mostrar o que ocorreria no interior do jogador máquina quando da realização de uma jogada. Sua estratégia é destruir de dentro para fora, não questiona a forma de funcionamento do jogo nem as condições em que ele deve ser jogado, Searle simplesmente converge toda a atenção para o jogador máquina, o jogo em si perde relevância. Seu argumento consiste em afirmar a tese de que a atividade de pensamento, que é a causa do agir inteligente, é produzida pelo cérebro, que é dotado de *intencionalidade*, isto é, a ação do homem dirigida a qualquer objeto distinto dele. Conforme esse argumento, o importante é sabermos o que torna nosso

cérebro capaz de produzir pensamentos e ações dessa natureza. Nessa concepção, duas tarefas seriam necessárias realizar-se se pretendemos construir uma máquina dotada de inteligência. A primeira seria o conhecimento daquilo que dá ao cérebro a capacidade de pensar e possuir intencionalidade; a segunda, sermos capazes de construir uma máquina que realize essas mesmas capacidades.

“Pode uma máquina pensar?” Meu ponto de vista é que somente uma máquina pode pensar, e de fato apenas máquinas de um tipo muito especial, a saber, cérebros e máquinas que têm os mesmos poderes causais do cérebro. E esta é a principal razão pela qual a IA no sentido forte tem pouco a dizer acerca do pensamento: ela não tem nada a dizer acerca de máquinas. Por definição, ela trata de programas, e programas não são máquinas. (SEARLE In: TEIXEIRA, 1996, p. 92)

Searle desconsidera a condição de Turing para aceitar uma máquina como participante do jogo da imitação, ou seja, a de que seres humanos não devem ser considerados máquinas.

A conclusão de Searle é que a IA forte não está credenciada a participar do debate sobre a produção de inteligência artificial, visto que os pesquisadores dessa área não possuem conhecimento sobre o cérebro e as máquinas capazes de realizar sua atividade, indispensáveis à busca do propósito de produção de inteligência. A IA forte não trata com os elementos causadores da ação inteligente, porque programas não são a causa de pensamentos. Isso ele deixa claro desde o primeiro momento em que trata do problema, “argumento que a IA no sentido forte deve ser falsa, uma vez que um agente humano poderia instanciar um programa e, mesmo assim, não ter estados mentais”. (SEARLE In: TEIXEIRA, 1996, p. 64)

O argumento do quarto chinês procura demonstrar que o computador que participa do jogo da imitação, respondendo a questões elaboradas pelos jogadores que perguntam, joga sem saber o significado das jogadas que realiza. O experimento montado por Searle nos mostra o funcionamento interno do computador digital que, mesmo sem saber o significado das expressões que apresenta como respostas adequadas às perguntas formuladas, as faz como se as soubesse.

No quarto chinês o processador tem à sua disposição todos os elementos que lhe permitem identificar qual a pergunta que deve ser respondida e qual a resposta mais adequada a ela. Assim, a tarefa do processador é bastante simples e não envolve nenhum tipo de mistério, pois seguir regras que relacionam conteúdos de diferentes listas, embora possa ser considerada uma atividade que envolve um grau de dificuldade e complexidade, não é nada de excepcional.

Embora Searle critique a forma proposta por Turing de reconhecimento de IA, ele não é céptico quanto ao propósito de criação de IA. Para ele, o pensamento e a consciência são fenômenos naturais que ocorrem nos cérebros humanos. Sua crítica se dirige à concepção de que o que ocorre nos cérebros humanos é passível de conhecimento científico, cuja forma de conhecer dá-se, dentre outras coisas, pela definição de um objeto observável. Searle considera que os estados mentais são distintos dos objetos de interesse das ciências empíricas, pois eles se dão em primeira pessoa, são essencialmente subjetivos. Ele não acredita que os estados mentais que ocorrem nos nossos cérebros sejam o mesmo que as operações realizadas pelos programas quando executados numa máquina digital. É necessário, portanto, o desenvolvimento de uma abordagem rigorosa que seja capaz de atender a novos requisitos epistêmicos. Essa nova abordagem deverá levar em consideração o caráter subjetivo daquilo que é tratado, se a pretensão for a elaboração de uma explicação aceitável da inteligência.

A resposta de Daniel Dennett a John Searle

O diálogo em torno do jogo da imitação entra em seu segundo momento quando, após a morte de Turing, Daniel Dennett responde, como se estivesse em seu lugar, à crítica de Searle de que o jogo da imitação não serve como critério seguro de reconhecimento de inteligência artificial. A entrada de Dennett na discussão adensa a influência filosófica no debate, pois se no primeiro momento o diálogo se deu entre um cientista/matemático e um filósofo, no segundo, o debate acontece exclusivamente entre filósofos.

O que ambos discutem é se os estados mentais expressam algo que é subjetivo e particular ou algo que pode ser redutível à objetividade da ciência, as neurociências. Ambos concordam que estados mentais não podem ser considerados entidades resultantes da ignorância e do senso comum. Estados mentais existem e sua explicação é fundamental e necessária.

Se para Searle, o teste de Turing não consegue aferir se uma determina ação produzida por uma máquina é inteligente ou não, isto se deve, segundo Dennett, a uma espécie de preconceito dos seres humanos para com as demais coisas. Julgamos ser os únicos portadores de *intencionalidade intrínseca ou original*. O conceito de *intencionalidade* é empregado por Searle para distinguir entre os fins objetivados por uma estrutura dotada de intencionalidade original e os fins alcançados por uma estrutura dotada de intencionalidade derivada. A intencionalidade original se caracteriza como uma ação que se dá em função de nenhuma outra coisa que não seja o seu próprio agente. Searle não acredita que o teste de Turing seja capaz de aferir se uma ação praticada por uma máquina possui esse tipo de origem.

Dennett não discorda de Searle de que está no cérebro aquilo que é responsável pelo agir inteligente. A discordância entre ambos é que se para o primeiro é possível alcançar-se um conhecimento científico que nos permitirá reproduzir o agir inteligente de modo artificial, para o segundo, os fenômenos responsáveis pela inteligência não se encontram ao alcance da ciência cognitiva. Para Dennett, há uma distinção entre filosofia na ciência cognitiva e filosofia da ciência cognitiva, ambas podem ser feitas independentes uma da outra. Mas a filosofia somente contribuirá com a ciência cognitiva na medida em que for capaz de desenvolver essas duas abordagens. Ele é um filósofo que trabalha na perspectiva de uma aproximação entre o trabalho do filósofo e o do cientista, pois acredita que há um enorme campo de trabalho a ser feito pelos filósofos na área da ciência cognitiva. Em sua opinião, o trabalho filosófico adquire características do trabalho científico nessa área, pois deverá proceder por meio do emprego de hipóteses testáveis e pelo esclarecimento da relação entre os conceitos da ciência cognitiva e os conceitos usualmente empregados no dia-a-dia nas tomadas de decisão. Esse grande esforço epistêmico a ser realizado só alcançará sucesso se houver, de fato, uma aproximação entre

filósofos e cientistas. Um indício da necessidade da realização dessa tarefa se traduz na forma como os filósofos e leigos veem a filosofia da ciência cognitiva. Para ele, “é um erro pensar que, a filosofia da ciência cognitiva seja “tanto para filósofos, quanto para leigos” apenas uma coleção de antídotos para inferências erradas sobre o que é importante para o trabalho científico contemporâneo”. (DENNETT, 2009, p. 231)

A dificuldade de trabalho conjunto de filósofos e cientistas é antiga, enquanto os cientistas consideram os filósofos de um criticismo absoluto que os impede de pensar soluções para os problemas; os filósofos qualificam os cientistas de um pragmatismo total, que os impede de adotar uma postura crítica diante dos seus esforços para o encontro de soluções.

Uma das principais lições que aprendi dos cientistas é que eles têm uma justificável impaciência, constantemente desprezada, para com os tipos de criticismo puramente destrutivos que os filósofos trazem da filosofia. Os filósofos podem subir nas árvores e atirar para todos os lados alegremente - mas isso não é construtivo. A suposição de fundo dos filósofos sempre deve ser: se um problema chama a nossa atenção, é porque pessoas inteligentes realmente o acham difícil; deveríamos humildemente tentar ajudá-los a organizar tudo, solicitando, não ordenando, devendo em seu criticismo agir por tentativa e não por antecipações. Me encolho quando vejo jovens filósofos fazerem uma chamativa demonstração de demolição número na frente de cientistas, uma conversa que escorreria como o mel em uma sala cheia de filósofos, mas que apenas faz com que cientistas sacudam a cabeça em consternação. (DENNETT, 2009, p. 233, tradução nossa).

Se, por um lado, os cientistas não suportam a índole crítica dos filósofos, esses, por sua vez, não concordam com a disposição dos cientistas em não esclarecer suficientemente os significados de uma expressão que passam a empregar sem maiores problemas.

A comunicação com os cientistas às vezes é difícil porque os cientistas usam o princípio da caridade. Eles ouvem falar sobre *qualia*, por exemplo, e decidem que sabem o que os filósofos querem dizer com isso (e, é claro, *qualia* existe - qualquer um que já tenha tido uma dor de dente sabe disso!). E quando eles aprendem sobre as controvérsias sobre *qualia*, eles pensam que podem resolvê-los ou, pelo menos, apontar para as suas soluções, com algumas observações extraídas de

trabalhos experimentais recentes. (DENNETT, 2009, p. 233, tradução nossa)

A crítica de Dennett a Searle diz respeito ao modo como deve ser feita a abordagem *intencionalista* no esforço de compreender-se o agir inteligente. Para Dennett, a abordagem searleana está equivocada porque atribui ao homem a *intencionalidade intrínseca ou original*. Ele não crê que nós, seres humanos, sejamos portadores de intencionalidade original porque somos produtos da natureza.

“[...] o cérebro é um artefato e obtém sua intencionalidade, seja qual for a intencionalidade que suas partes possuam, a partir do papel que desempenha na economia em funcionamento do sistema maior de que faz parte – ou, em outras palavras, das intenções de seu criador, a Mãe Natureza (também conhecida por processo de evolução por seleção natural). (DENNETT, 1997, p. 53)

Em *The Consciousness Explained* (1992), Dennett propõe-se a apresentar um novo tipo de explicação para a *consciência*. O primeiro resultado alcançado por esta nova forma de explicar a consciência é negativo, pois pretende romper com a representação cartesiana hegemônica que entende a consciência como um tipo de teatro cujos atores são as ideias. Sua explicação é um contraponto à alegoria cartesiana. Ele parte de uma pergunta: como a *consciência* passou a residir no cérebro humano?

Iniciaremos a substituição daquele modelo por uma outra imagem positiva... Para avançarmos deveremos mudar de terreno e abordar a complexidade da consciência de um ponto de vista diferente: o da evolução. Como a consciência humana nem sempre existiu, ela deve ter se originado de fenômenos anteriores que não são, eles mesmos, conscientes. (DENNETT, 1992, p. 171, tradução nossa)

Ele não acredita que a consciência seja uma caixa-preta inexpugnável que não permita uma investigação dos fenômenos que nela ocorrem e que são os responsáveis pela produção de algo tão extraordinário. Dennett prefere adotar um modo de abordar a questão que a torne suscetível de uma investigação de natureza científica. Sua abordagem é tributária da biologia ao admitir que os acontecimentos que se situam nessa esfera possuem uma orientação, indo do mais simples para o mais complexo. O que torna

possível a ele traçar um paralelo entre eventos de naturezas distintas reforçando, com isso, a ideia de que a vida está inserida num ambiente cuja característica principal é o constante aumento de velocidade da complexidade.

A questão da origem da sexualidade e a da consciência são paralelas. A vida sexual das flores, ostras e outras formas simples de vida, não tem quase nada de erótico (em termos humanos), mas podemos reconhecer em suas rotinas de reprodução, aparentemente sem graça, mecanismos que são a base e os princípios de nosso mundo sexual, o qual é muito mais excitante. Da mesma forma, não há nada particularmente egoísta nos precursores primitivos dos seres humanos conscientes, mas lançam as bases das nossas inovações e complicações particularmente humanas. A concepção de nossas mentes conscientes é o resultado de três processos evolutivos sucessivos, empilhados uns sobre os outros, cada qual sendo mais rápido e poderoso que seu predecessor. (DENNETT, 1991, p. 173)

Por estar submetida ao processo evolutivo, a origem da consciência implica num agir por interesse. O agir por interesse ocorre quando uma entidade pauta sua ação por aquilo que esta ação representa de *bom* ou *ruim* para si. De acordo com Dennett (*La Peligrosa Idea de Darwin*, 1999), a capacidade de produzir cópias implica, simultaneamente, na possibilidade da produção de erros no processo de multiplicação, podendo resultar em algo inesperado, como a própria extinção da entidade replicante, por exemplo. Torna-se necessário, portanto, a criação de dispositivos que impeçam que isso ocorra. Não basta a capacidade de replicação, é necessário que a replicação seja correta. Trata-se de um nível mais sofisticado de atividade. Os organismos, além de serem capazes de empregar aquilo que Dennett chama de *dispositivos pré-instalados* (*wired-in*), precisam desenvolver novos tipos de dispositivos que tornem suas respostas à pressão de seleção mais eficientes. Um exemplo de dispositivo pré-instalado (*wired-in*) é a capacidade humana de reconhecer características dos seus predadores através de padrões de simetria.

Em *Tipos de Mentes* (1996), Dennett apresenta quatro tipos de configurações cerebrais por meio da alegoria de uma torre com quatro andares, com o nível mais básico abaixo do mais desenvolvido. Essa graduação tem relação com o processo evolutivo que

ele apresentou em *The Consciousness Explained* (1992) que é dividido em três etapas até chegar ao momento em que a consciência passa a residir nos cérebros humanos.

De acordo com ele (*The Consciousness Explained*, 1992), os cérebros e o sistema nervoso foram estimulados a dar respostas às pressões de seleção exercidas pelo meio ambiente. As primeiras respostas eram reações imediatas e sem planejamento, sendo produzidas pelos mecanismos pré-instalados (*wired-in*).

No começo havia a evolução darwiniana das espécies por seleção natural. Uma grande variedade de organismos candidatos foram gerados às cegas, por processos mais ou menos arbitrário de recombinação e mutação de genes. Estes organismos foram testados na prática e apenas os melhores projetos sobreviveram. Este é o andar térreo das torres. Chamemos seus habitantes de *criaturas darwinianas*. (DENNETT, 1996, p. 79)

Um segundo nível de respostas apresenta um incremento em relação ao primeiro, ao demonstrarem possuir um certo grau de antecipação diante do que poderá ocorrer, embora ainda seja uma resposta produzida pelos dispositivos pré-instalados da estrutura primitiva dos cérebros. O comportamento do organismo não é meramente reativo, ele procura se antecipar ao que poderá acontecer. No processo evolutivo, as criaturas darwinianas sobreviviam em razão de uma configuração rígida que as obrigava a responder aos estímulos do meio ambiente sempre da mesma forma. As criaturas de segundo nível não estão mais submetidas a uma configuração rígida como as de primeiro nível, elas são portadoras de uma novidade biológica chamada de *plasticidade fenotípica*, isso é, a capacidade que alguns organismos possuem de, mesmo sem estarem completos ao nascer, completarem o que falta com a experiência do viver.

Esses indivíduos confrontavam o meio ambiente gerando uma variedade de ações, as quais testavam, uma por uma, até que encontrassem uma que funcionasse. Eles detectavam que essa ação funcionava apenas pela obtenção de um sinal positivo ou negativo do meio ambiente, o que ajustava a probabilidade de aquela ação ser produzida em outra ocasião. Quaisquer criaturas dotadas com as estruturas erradas – com o reforço positivo e o negativo invertido – estariam condenadas, é claro. Apenas aqueles afortunados por nascer com os reforçadores apropriados teriam uma vantagem. Podemos

chamar esse sub-conjunto de criaturas darwinianas de *criaturas skinnerianas...* (DENNETT, 1996, p. 80)

O terceiro nível de respostas não é produzido em razão de um dispositivo pré-instalado, mas da capacidade de aprendizagem com a experiência adquirida em vida pelo organismo. Ele apresenta um refinamento em relação ao nível anterior com o incremento da pré-seleção de hipóteses, descartando aquelas que produziriam um conhecimento nulo, uma estratégia econômica no emprego dos esforços dos organismos dotados com este tipo de cérebro.

O condicionamento skinneriano é uma coisa boa desde que você não seja morto por um dos seus erros anteriores. Um sistema melhor envolve a *pré-seleção* entre todos os comportamentos possíveis ou ações, de modo que os movimentos verdadeiramente estúpidos são descartados antes de serem arriscados na “vida real”. Nós, seres humanos, somos criaturas capazes deste refinamento particular, mas não estamos sozinhos, podemos chamar os que se beneficiam deste terceiro andar da torre das criaturas popperianas, já que, como o filósofo Sir Karl Popper uma vez refinadamente colocou, este reforço do projeto “permitem que nossas hipóteses morram em nosso lugar”. (DENNETT, 1996, p. 83)

O último nível de resposta, além de incorporar todas as estratégias dos níveis anteriores apresenta uma inovação muito importante, a capacidade de um organismo comunicar a outros organismos quais dispositivos de produção de respostas são úteis à sobrevivência, mesmo que não sejam pré-instalados.

Uma das visões intuitivas fundamentais de Darwin é que um projeto é muito caro, mas copiar projetos é barato; isto é, produzir um projeto inteiramente novo é muito difícil, mas reprojetar velhos projetos é relativamente fácil. Poucos dentre nós poderiam reinventar a roda, mas não temos de fazer isto, já que possuímos o projeto da roda (e uma variedade de outros projetos) das culturas nas quais fomos criados. Podemos chamar este sub-sub-subconjunto de criaturas gregorianas... (DENNETT, 1996, p. 92)

Portanto, a gênese da ação por interesse presente nos seres inteligentes se dá a partir de um mecanismo simples de replicação e manutenção dos primeiros seres vivos, é neles que encontramos as origens mais profundas de todo o processo que resulta no surgimento

da consciência, que não é somente um acontecimento biológico, mas também, histórico e cultural.

Bennett e Hacker: uma proposta de trabalho convergente

Em *Searle e Dennett: duas perspectivas de estudo da mente* (PIRES DA SILVEIRA, 2013), argumentamos que há um papel relevante da filosofia no debate sobre os objetivos da IA, dado que a necessidade de uma explicação do que seja a *consciência* e o *agir inteligente* dependem do esclarecimento dos conceitos e suas definições empregados na descrição desses fenômenos. Nesse artigo, procuramos mostrar que há um antagonismo entre as propostas de investigação da *consciência* apresentadas por Searle e Dennett. Esse antagonismo se dá em razão das diferentes concepções de *consciência* por eles assumidas e, consequentemente, dos diferentes modos de investigá-la.

Contudo, esse antagonismo não é uma forma de contrariedade absoluta, pois ambos são considerados, por diferentes razões, como filósofos que não duvidam da possibilidade de, algum dia, alcançarmos conhecimento e condições práticas suficientes para produzirmos IA.

Numa demonstração de que o desafio do trabalho em colaboração pode ser superado, Maxwell Bennett (1939) e Michael Stephan Hacker (1939) publicaram *Fundamentos Filosóficos da Neurociência* (2005), sendo o primeiro um neurocientista e o segundo um filósofo. Suas observações não são consideradas muito alentadoras para aqueles que pretendem alcançar o objetivo principal do projeto de IA. Nesse livro, seus autores propõem-se a fazer uma revisão rigorosa dos conceitos empregados na elaboração das explicações dos resultados alcançados pela investigação neurocientífica. Eles partem do pressuposto de que a tarefa do cientista consiste em testar hipóteses e a do filósofo em esclarecer conceitos, sendo ambas complementares.

É preferível considerar que ambas tratam de tipos de questões diferentes. Todavia, visto que é necessário um esquema conceptual para se fazer uma experimentação proveitosa, não podemos deixar de por esses dois tipos de questões. (BENNETT & HACKER, 2005, p. 11)

Essa é a maneira mais adequada para tratar a questão de como se dá o agir consciente, isto é, através de perguntas de natureza empírica e perguntas de natureza teórica, cada uma devendo ser respondida de acordo com a sua origem epistêmica.

Daí que as páginas introdutórias sejam dedicadas a esclarecer seu campo de estudo e, sobretudo, a delimitar a forma de acesso ao conhecimento em cada uma delas, distinguindo dois tipos logicamente diferentes de indagação intelectual: por um lado, as perguntas empíricas (próprias da investigação em neurociência) e, por outro lado, as conceptuais (próprias da investigação filosófica), já que cada campo e, portanto, cada tipo de pergunta contarão com suas próprias vias de abordagem e solução. Desta maneira, as primeiras tentarão determinar o que é verdadeiro ou falso (ou correto e incorreto, no caso das teorias científicas), enquanto que as segundas o que possui ou não possui sentido. (BOTTO, 2009, p. 164, tradução nossa)

Entretanto, o que os fez levar adiante o projeto conjunto de *Fundamentos Filosóficos da Neurociência* (2006) foi uma necessidade que tem profunda relação com a tarefa filosófica, a de esclarecer os conceitos empregados pelos neurocientistas quando apresentam suas explicações sobre a *consciência* baseados nos experimentos por eles realizados.

Se quisermos compreender as estruturas e dinâmicas neurais que tornam possível a percepção, o pensamento, a memória, a emoção e o comportamento intencional, é essencial que haja clareza quanto a estes conceitos e categorias. Ambos os autores chegando a esta investigação proveniente de direções muito diferentes, ficaram perplexos, e por vezes embaraçados com o uso de conceitos psicológicos na neurociência contemporânea. A perplexidade ficou a dever-se, muitas vezes, ao significado possível de uma dada afirmação dos neurocientistas em relação ao cérebro e à mente, ou ao motivo que levava o neurocientista a pensar que as experiências que levava à cabo esclareciam a capacidade psicológica que estava a ser estudada, ou os pressupostos conceptuais das questões postas. O embaraço produzido pela suspeita de que em alguns casos os conceitos estavam mal construídos, ou mal aplicados ou alargados para além das suas condições específicas de aplicação. E quanto mais pesquisávamos, mais convencidos ficávamos de que, a despeito dos notáveis progressos da neurociência cognitiva, nem tudo ia bem quanto à teoria geral. (BENNETT & HACKER, 2005, p. 15)

Bennett & Hacker (2005) identificaram diferentes tipos de equívocos produzidos pelos usos errados e inadequados dos conceitos nas explicações neurocientíficas, das quais destacaremos duas: a primeira por ter relação com a tarefa do cientista, a segunda por nos remeter à tarefa do filósofo. Na opinião desses dois pesquisadores, um dos grandes erros que os investigadores envolvidos na tarefa de explicar a “consciência” cometem, foi o de atribuir ao cérebro predicados que só fazem sentido quando aplicados ao organismo como um todo, a chamada *falácia mereológica*.

O termo *mereologia* foi originado pela conjunção de dois termos do grego (<https://it.wikipedia.org/wiki/Mereologia>) e consiste na atitude de atribuir às partes propriedades que só podem ser atribuídas ao todo. Em *Fundamentos Filosóficos da Neurociência* (BENNETT & HACKER, 2005) este tipo de atitude é considerado um dos equívocos cometidos pelos neurocientistas na elaboração das suas explicações. Eles atribuem o cometimento desse equívoco à herança cartesiana que influencia a reflexão e a pesquisa em torno do *problema da consciência*. De acordo com eles, os primeiros investigadores dedicados ao problema da consciência eram cartesianos, isto é, adotavam o *dualismo de substâncias* como modelo metafísico na elaboração de suas explicações, entendiam que a mente e o corpo eram substâncias distintas e com suas respectivas realidades.

Segundo os autores, ainda hoje, após pelo menos três gerações de pensadores modernos dedicados à investigação do problema da consciência, vários neurocientistas como Francis Crick, Gerald Edelman, Antonio Damasio e outros, continuam a incorrer no mesmo erro dos pioneiros dessa área de investigação ao não questionarem o dualismo de substâncias. Várias afirmações feitas por esses pensadores são apresentadas como provas que nos mostram uma variedade de formas de cometimento desse equívoco. Nessas citações, os autores destacam que a atitude consiste sempre em afirmar que é o cérebro que sente, vê, sente, categoriza, etc.

De forma que o cérebro tem experiências, acredita em coisas, interpreta pistas com base em informação que foi posta à sua disposição, e faz conjecturas. (BENNETT & HACKER, 2005, p. 83)

De forma que o cérebro conhece coisas, raciocina indutivamente e constrói hipóteses com base em argumentos, e os neurônios que o

constituem são inteligentes, podem calcular probabilidades e apresentar argumentos. (BENNETT & HACKER, 2005, p. 84)

Antonio Damasio afirma que “os nossos cérebros podem muitas vezes decidir corretamente, em segundos, ou minutos, dependendo da duração de tempo que consideramos adequada para o objetivo que queremos alcançar..”. De modo que o cérebro decide ou, pelo menos “decide”, e inicia a ação voluntária. (BENNETT & HACKER, 2005, p. 84)

Para eles, não faz sentido afirmar que o cérebro seja o agente daquelas ações, pois uma decisão, por exemplo, não é tomada por um cérebro, mas por um agente que, dentre outras coisas, possui um cérebro.

Não é que o cérebro não pense nem deixe de pensar, é que não faz sentido atribuir-lhe esse predicado. Por exemplo, quando dizemos que o falcão voa e que o leopardo corre, atribuímos o vôo e a corrida ao falcão e ao leopardo em sua totalidade. Não teria sentido dizer que as asas do falcão são o que voa e as patas do leopardo o que corre.

(<https://vonnewmannmachine.wordpress.com/2010/02/28/?mi-cerebro>)

A tarefa de tentar esclarecer este cenário confuso cabe, na opinião de Bennett & Hacker (2005), à filosofia. A filosofia, a despeito de ser considerada uma área de conhecimento que pouca ou nenhuma contribuição traz para o avanço das pesquisas na área de IA, possui um papel relevante em todo esse debate, pois “a questão com que nos defrontamos é filosófica, e não científica. Exige esclarecimentos conceptuais e não investigações experimentais”. (BENNETT & HACKER, 2005, p. 86)

Esse quadro de confusão nos usos de termos relevantes na produção das explicações sobre a “consciência” eles creditam ao que chamam de “forma mutável de cartesianismo”. Em três gerações de neurocientistas, o cartesianismo não deixou de influenciar o trabalho nessa área de pesquisa; visto que serviu de modelo metafísico, tanto para ser adotado quanto para ser rejeitado, pelos investigadores. O cartesianismo é a questão de fundo que, de alguma forma ou de outra, constitui o horizonte de pensamento dessa área de pesquisa e, de certo modo, de todo o debate em torno da IA. Ao adotarem o *dualismo de substâncias*, os pesquisadores atuais acabam por atribuir predicados a qualquer coisa, indiscriminadamente, sem levar em consideração a qual das duas categorias pertence.

Era um traço característico do dualismo cartesiano imputar predicados psicológicos à mente, e só por derivação ao ser humano. Sherrington e seus discípulos Eccles e Penfield aderiram a uma forma de dualismo nas suas reflexões sobre a relação das suas descobertas neurológicas com as capacidades humanas perceptivas e cognitivas. Os sucessores rejeitaram o dualismo – e com razão. Mas a terceira geração de neurocientistas cerebrais, por seu lado, aplicou irrefletidamente ao cérebro os predicados que os dualistas imputam à mente imaterial. (BENNETT & HACKER, 2005, p. 87)

Na opinião deles, não há *intercambialidade* de predicados entre os elementos pertencentes às diferentes substâncias. Esse tipo de atitude é um vício de pensamento, só ocorre porque podemos combinar livremente termos, mesmo que essa combinação não faça sentido algum. Isso indicaria um uso descuidado dos conceitos e dos termos de linguagem. No melhor das hipóteses, essa *intercambialidade* de predicados somente deveria ser levada em consideração se a compreendêssemos como uma construção metafórica, a elaboração de uma imagem com vistas a auxiliar a compreensão de algo.

Colocação do problema: o projeto de IA é admissível?

O contexto de surgimento e desenvolvimento do projeto de IA é bastante agitado epistemicamente, com muitas concepções, doutrinas, formas de tratamento dos problemas e influências advindas das mais variadas áreas de conhecimento.

Devido ao fato desse projeto possuir uma face mais destacada que outras, somos induzidos a pensar que a questão de se devemos ou não aceitar como realizável o objetivo de construção de um *novo homem* deve ficar circunscrito a ela, qual seja, a face científico-tecnológica. Não há dúvida que essa área de conhecimento tem um papel de suma relevância no alcance do objetivo em questão, já que não cabe às formas de conhecimento estritamente teóricas a tarefa de construção do *novo homem*.

Entretanto, como pudemos ver no capítulo anterior, foi a interposição de formas de conhecimento especulativo-filosóficas com as empírico-tecnológicas que tornou possível a concepção, construção e o desenvolvimento das máquinas de estado discreto de Turing. Foi do cruzamento de modelos metafísico-lógico-matemáticos, com o modo de

funcionamento de componentes eletro-eletrônicos, equiparados ao funcionamento de células do cérebro humano, que o projeto de IA surgiu. Foi de um modelo metafísico de explicação do mundo e das coisas que se derivou a compreensão de que o pensamento é um fenômeno que envolve um determinado tipo especial de realização, o cálculo computacional.

Não é de causar surpresa a constatação do caráter divergente dos projetos que tem em vista alcançar o objetivo último do projeto de IA. O caminho de descoberta desse continente parece estar em fase inicial, ele é mais intuitivo do que planejado. Ainda não temos uma rota que nos leve até ele, nem temos muita certeza se os meios de que dispomos nos permitirão avistar seus rochedos e praias. Tais condições estimulam a propositura de diferentes modos de tratar a questão. Em determinado ponto da linha temporal de desenvolvimento da explicação do *agir inteligente* a ênfase foi dada a definição metafísica do “homem”; noutro, na atribuição de uma interioridade ontológica inerente a este; num terceiro, na observação do seu comportamento; num quarto, ao conhecimento do cérebro e dos neurônios e da estrutura constitutiva básica do mundo.

Outra característica marcante desse contexto de intenso debate são as previsões de realizações práticas na busca por inteligência artificial, como se as pesquisas e problemas relativos a ela estivessem submetidos a um cronograma de construção. Isso pode ser visto na previsão de Hans Moravec em que ele compara a quantidade de cálculos realizáveis por segundo com a quantidade de dólares necessários para se adquirir uma máquina com determinada capacidade de processamento.

A quantidade de capacidade computacional que um dólar pode comprar aumentou mil vezes a cada duas décadas desde o início do século. Em oitenta anos houve uma diminuição do custo da computação. Se essa taxa continuar no próximo século, os 10 teraops requeridos por um computador humano estará disponível num super-computador de dez milhões de dólares antes de 2010 e, num computador pessoal de mil dólares até 2030. (MORAVEC, 1988, p. 68, tradução nossa)

As mesmas expectativas e suposições são compartilhadas por Raymond Kurzweil (*A era das máquinas espirituais*, 2007) ao se referir ao ritmo do progresso alcançado na construção de computadores. Kurzweil aponta que a característica dos acontecimentos

que se dão no tempo é a de terem um início lento de mudança que aumenta sua velocidade com o passar do tempo. Tal compreensão está de acordo com o *princípio de entropia*, segundo o qual o grau de desordem de um sistema fechado só tende a aumentar com o passar do tempo, gerando calor e maior intensidade de movimentos.

Para Kurzweil é uma questão de tempo para o pensamento humano ser substituído pelo realizado por outra entidade pensante, neste caso, os computadores. O cérebro humano possui uma capacidade limitada de processamento, o que o torna um obstáculo para o processo evolutivo.

Mas a Lei dos Retornos Acelerados nos diz que a evolução não permanecerá presa num beco sem saída por muito tempo. E, de fato, a evolução encontrou um jeito de tangenciar as limitações computacionais dos circuitos neurais. De modo inteligente, ela criou organismos que, por sua vez, inventaram uma tecnologia computacional um milhão de vezes mais rápida do que neurônios com base em carbono (que estão continuando a ficar ainda mais rápidos. A computação realizada em circuitos neurais mamíferos extremamente lentos acabará por ser transportada para uma versão equivalente eletrônica (e fotônica) bem mais versátil e veloz. (KURZWEIL, 2007, p. 146)

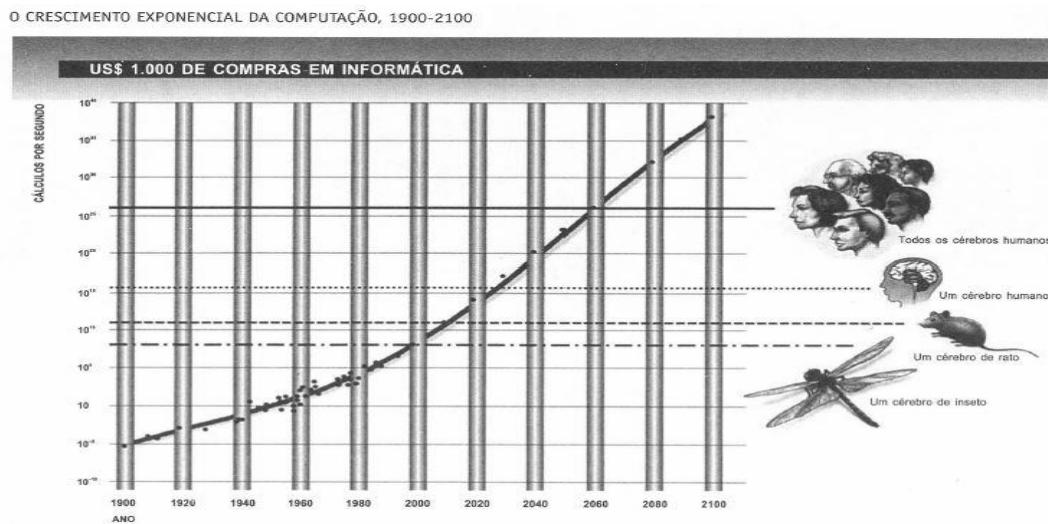


Figura 1: Kurzweil, 2007

Como podemos ver na figura acima, a expectativa é de que antes do final deste século o cérebro humano será superado pelos processadores de redes neurais.

O cérebro humano possui cerca de 100 bilhões de neurônios. Com uma média estimada de mil conexões entre cada neurônio e seus vizinhos, temos cerca de 100 trilhões de conexões, cada uma capaz de um cálculo simultâneo. Este é um processamento paralelo um tanto maciço, e uma das chaves para a força do pensamento humano. Entretanto, uma profunda fraqueza é a velocidade excruciantemente lenta dos circuitos neurais, de apenas 200 cálculos por segundo. Para problemas que se beneficiam de paralelismo maciço, como reconhecimento de padrões com base em redes neurais, o cérebro humano faz um grande trabalho. Para problemas que exijam extenso pensamento sequencial, o cérebro humano é apenas medíocre. (KURZWEIL, 2007, p. 147)

Desse modo, o projeto de IA não se propõe a desenvolver e construir um dispositivo que seja similar ao cérebro humano, ele coloca no horizonte de suas realizações a superação da própria capacidade humana de pensar. O arrojo das afirmações feitas pelos pesquisadores do projeto de IA soa como um exagero, mais ainda, apresenta-se como uma atitude heterodoxa que vai contra princípios e postulados de várias áreas de conhecimento como a antropologia e a própria biologia, ao presumir a ideia de que o homem poderá melhorar a si mesmo por meio de um processo evolutivo artificial, isto é, criado por ele mesmo. Estamos presenciando a aurora de uma nova forma de conhecimento que reformulará paradigmas até aqui consagrados?

Há um episódio na história da ciência que serve de elemento de comparação com o contexto de surgimento e desenvolvimento do projeto de IA, o de surgimento da Química.

A Alquimia nem sempre foi considerada como precursora da Química, pelo contrário, era considerada sua antípoda. Essa diferença é apontada por Patisson-Muir (1848-1931) em *The story of Alchemy and the beginnings of Chemistry*, (2004), cuja primeira edição é de 1902. No prefácio o autor adverte ao leitor de que o livro pretende lhe mostrar a diferença entre um modo científico e um modo emocional de olhar para a natureza, ou seja, entre a Química, empregadora do modo científico, e a Alquimia. Elas não pertenceriam ao mesmo veio de conhecimento. Contudo, ele não deixa de reconhecer um determinado âmbito da atividade dos alquimistas que se identifica com a do químico na sua essência, embora hoje os químicos já não trabalhem mais em porões, sendo suas pesquisas são financiadas pelo Estado, além de andarem limpos em suas roupas de laboratório.

Os alquimistas não podem ser acusados de preguiça ou aversão ao trabalho em seus laboratórios. Paracelso (sec. XVI) diz deles: “Eles não são dados à ociosidade, nem se orgulham do hábito de vestir roupas de pelúcia e de veludo, de usar anéis nos dedos ou espadas de prata na cintura e de calçar luvas finas nas mãos, pelo contrário, eles seguem diligentemente seus trabalhos, suando dias e noites inteiras em seus fornos. Eles não passam seu tempo recreando-se ao ar livre, mas deleitando-se em seus laboratórios. Eles colocam seus dedos nas brasas, no barro e na imundície, não em anéis de ouro. Eles estão sempre cobertos de fuligem e carvão, como os ferreiros e os mineiros, não se orgulhando de terem os rostos limpos e bonitos. (PATTISON-MUIR, 2004, p. 12, tradução nossa)

Hoje a compreensão da relação entre a Química e a Alquimia é distinta, em *Da Alquimia à Química* (Alfonso-Goldfarb, 1988) a Química é apresentada como o resultado da mudança de paradigma do vitalismo para o mecanicismo. De acordo com essa compreensão, a passagem da Alquimia para a Química não pode ser compreendida como um processo de ruptura completa que se consolidou no período entre as publicações de Boyle (*O químico cético*, 1661) e A. Lavoisier (*Tratado Elementar de química*, 1789).

[...] nesse pouco mais de um século, como distinguir o alquimista do químico, igualmente envolvidos pela quietude cheia de fumos do laboratório; ao pé da fornalha por dias, meses e anos? Talvez porque os dois tenham sido um só e o mesmo, existe uma história, bem menos conhecida e com duração bem mais longa, que parece contar de forma mais suave e coerente como se deu essa passagem. (ALFONSO-GOLDFARB, 2011, <http://www.quimica.seed.pr.gov.br/modules/noticias/article.php?storyid=595>)

Embora sejam apontadas diferenças entre ambas, esse novo modo de abordar a relação entre a Alquimia e a Química não estabelece uma comparação com vistas a mostrar o caráter primitivo da primeira com relação a segunda. O contexto de surgimento da Química é visto como um desdobramento da Alquimia que, por milhares de anos, desenvolveu procedimentos e buscou respostas para questões que hoje são objeto de investigação da primeira. Os alquimistas lançaram as bases de um conhecimento que tinha por objetivo compreender as mudanças observadas nos materiais.

Longe de algo sensacional como a descoberta do oxigênio ou o elixir da longa vida, o ponto alto dessa história hoje pode parecer

insignificante: a padronização dos modos de pensar e operar no laboratório, traduzida modernamente por “materiais e métodos”. (ALFONSO-GOLDFARB, 2011, <http://www.quimica.seed.pr.gov.br/modules/noticias/article.php?storyid=595>)

A despeito de sua busca pela *pedra filosofal* e pelo *elixir da longa vida*, os alquimistas implementaram práticas que ainda hoje são empregadas nos laboratórios, fruto da sua dedicação no desenvolvimento de técnicas que lhes permitissem compreender as mudanças por meio dos processos químicos que lhes subjazem.

Houve um resgate da Alquimia no final do século passado pela realização de pesquisas que se utilizaram de novas tecnologias e equipamentos, o que deu a possibilidade de chegar a resultados que sustentam algumas das teses da Alquimia consideradas irrealizáveis.

Em 1980, cientistas da Universidade da Califórnia em Berkeley usaram um acelerador de partículas para transformar uma amostra inimaginavelmente pequena de bismuto em ouro. Isso saiu ao custo de 10 mil dólares para fazer um bilionésimo do valor de um centavo de ouro. Eles mostraram que a transmutação – a transformação de um elemento químico em outro – hoje é possível, mas claramente não é uma forma de fazer dinheiro. No começo do século XX, a transmutação não estava ao alcance dos cientistas. Antes, no século XIX, poucos químicos tinham algum interesse numa ideia tão louca, a ponto de a considerarem como parte de uma Alquimia desacreditada. A maioria dos historiadores tiveram pouco interesse na Alquimia, exceto para mostrar seus erros e o quanto ela não era científica. Apenas alguns excêntricos continuaram a procurar pela pedra filosofal, a fabulosa substância que transformaria os metais comuns em ouro. (LEVERE, 2001, p. 1, tradução nossa)

Parece que estamos diante de um caso raro na história do conhecimento humano, a reabilitação e a recuperação do prestígio de uma forma de conhecimento que foi desprezada por sua falta de científicidade.

Assim como a Alquimia sofreu a crítica de que não seria capaz de atingir os propósitos que almejava, o projeto de IA também sofre críticas semelhantes.

Hubert Dreyfus (*O que os computadores não podem fazer*, 1975 e *Mind over Machines*, 1986) afirma que a versão forte do projeto de IA não tem como desenvolver

máquinas que atinjam a mesma capacidade que a inteligência humana. Sua argumentação, com vistas a sustentar tal afirmação é apresentada em dois momentos.

No primeiro momento, a premissa forte do argumento é a afirmação de que o uso competente de regras não é suficiente para gerar uma determinada compreensão situacional do mundo.

Em *O que os computadores não podem fazer*, escrito em 1969, a principal objeção era a impossibilidade do uso de regras para selecionar somente aqueles fatos sobre o mundo real que são relevantes numa determinada situação. (DREYFUS, 1986, p. 102, tradução nossa)

A evolução no desenvolvimento de máquinas e programas, desde a publicação do seu livro (*O que os computadores não podem fazer*, 1975), como os programas capazes de jogar xadrez, fizeram com que ele tivesse de rever seu argumento inicial.

No segundo momento da sua argumentação (*Mind over Machines*, 1986), a premissa forte do argumento passa a ser a afirmação de que os computadores não possuem e não possuirão, algum dia, a capacidade de intuir. Ele considera a *intuição* uma habilidade da mente humana que a torna capaz de desenvolver um nível superior de inteligência. As máquinas digitais possuiriam limites que impedem que seu desenvolvimento alcance os mesmos níveis de inteligência que a humana.

Para sustentar essa afirmação, ele divide o aprendizado humano em cinco níveis: noviço, iniciante avançado, competente, proficiente e perito. O aprendiz está imerso num contexto que vai desde o aprendizado do reconhecimento de situações e aquisição de regras aplicáveis a elas, até o nível de saber solucionar problemas, passando pelo preparo para lidar com situações reais, pela capacidade de reconhecer um maior número de diferentes contextos e situações no mundo e pelo hábito de tomar decisões seguindo um método que o torna apto a fazer escolhas conscientes a partir da consideração de várias alternativas.

Após a exposição do que consiste cada um dos cinco níveis, Dreyfus (1986) conclui que os dois últimos não estão ao alcance das máquinas digitais.

A mudança para os dois últimos níveis de aprendizado pressiona os limites das máquinas lógicas. Os aprendizes dos níveis proficiente e perito não buscam conscientemente por fatos para inferir metas ou ações, eles escolhem inconscientemente qualquer meta ou ação seja qual for. ...A impressão que os peritos dão, e seu comportamento confirma isso, é de que são capazes de lidar com situações de um modo rápido e flexível. Eles sabem que é necessária uma considerável carga de experiência para realizar o salto para a intuição, mas eles não têm consciência do que acontece nos seus cérebros quando isso acontece. As pessoas falam, às vezes, em acalmar a mente analítica e deixar o emocional e holístico lado direito do cérebro, envolvido, fazer o seu trabalho. (DREYFUS, 1986, p. 65, tradução nossa)

As máquinas digitais, de acordo com a argumentação de Dreyfus não teriam a capacidade de dar esse salto para um modo intuitivo de proceder, o que as impediria de alcançar o nível mais elevado de inteligência. Podemos perguntar se a impossibilidade das máquinas digitais serem dotadas de inteligência se deve aos limites que esse tipo de dispositivo possui ou aos próprios limites da capacidade de conhecimento dos seres humanos.

Mas, se a argumentação de Dreyfus de que as máquinas digitais jamais serão capazes de atingir os mesmos níveis de inteligência que os alcançados pelos seres humanos estiver correta, então a defesa da versão forte do projeto de IA não faz sentido, já que, atualmente, o dispositivo julgado capaz de fazer o mesmo que o cérebro humano são as máquinas digitais.

Dito isso, colocamos nossa hipótese de trabalho, qual seja: a versão forte do projeto de IA é, ao menos, admissível no que diz respeito ao seu objetivo de construir uma máquina capaz de realizar algo que até então é exclusividade dos seres humanos e que lhe foi engendrado pela natureza, ou tais objetivos estabelecem uma meta que está fora da capacidade humana de realização?

CAPÍTULO 3

COMO TRATAR O PROBLEMA DA IA

A maneira como o problema desta tese será tratada é a resultante da interposição dos pensamentos de três filósofos do século XX, John Langshaw Austin (1911-1960), Reinhard Koselleck (1923-2006) e Douglas Hofstadter (1945).

Austin por nos advertir dos falsos problemas com os quais poderemos nos defrontar, se a configuração metafísica com a qual os abordarmos for dualista do tipo dicotômica. A reflexão em torno dos objetivos do projeto de IA, como vimos no capítulo anterior, está profundamente influenciada por modelos metafísicos dicotômicos, o que torna a advertência de Austin altamente relevante para quem se propõe a produzir uma reflexão em torno de tal problemática.

Koselleck por apresentar uma possibilidade de superação do impasse dicotômico com sua visão estratificada de *tempo histórico*. Na concepção koselleckiana, o *tempo histórico* constitui-se em camadas que não formam pares de oposição, pois todas produzem efeitos umas sobre as outras. Seu modelo de tempo pressupõe outro arranjo metafísico, em que as perspectivas de compreensão do que acontece são muito mais vastas e diversificadas do que as constituintes que formam um par dicotômico. Ele pode ser considerado um pensador que se propôs a superar uma relação dicotômica, pois ao conceber o tempo como uma espécie de estratificação ele nos apresenta uma nova concepção de tempo que não é nem linear, nem circular.

Hofstadter, que ao tratar da noção de *volta estranha*, expõe-nos um movimento de pensamento que adquire expressão nas artes e na matemática. Um arranjo de pensamento que, em nossa opinião, reúne condições de superar os limites impostos pelo modo de pensar dicotômico.

Entendemos que a discussão envolvendo o projeto de IA é permeada por teses filosófico-metafísicas acerca das concepções e definições de *homem, inteligência, mundo, conhecimento e tecnologia* que influenciam fortemente a compreensão geral do debate inerente a esse conjunto de questões.

A discussão em torno do projeto de IA é bastante influenciada por uma infinidade de ideias que de modo algum são compatíveis com os propósitos deste projeto. Muitas críticas quanto ao sucesso desse empreendimento geralmente se apoiam em concepções

filosóficas tão antigas, oriundas de um tempo em que ainda não era cogitada a capacidade do homem construir, com o seu conhecimento, algo que fosse, de certa forma, uma cópia de si. Entretanto, embora distantes no tempo, elas não deixam de produzir efeitos quando defrontadas com algumas das realizações da ciência e da tecnologia atuais, influenciando o ambiente do debate em torno da tentativa de construção de um dispositivo dotado de inteligência como a humana. Tais efeitos não são irrelevantes. Pressupomos ser exatamente eles um dos principais responsáveis pelo impasse na discussão sobre a capacidade de o homem poder, um dia, produzir algo tão inteligente quanto ele próprio.

Em nossa opinião, as críticas dirigidas ao projeto de IA, às quais sustentam que o objetivo desse projeto não tem como ser atingido, só chegam a esta conclusão, porque pressupõem uma concepção dicotômica da metafísica no tratamento das questões relacionadas a ele.

Destacaremos, nos pensamentos de Austin, Koselleck e Hofstadter, os elementos de suas análises que servem de sustentação ao modo como, aqui, abordamos a reflexão em torno do projeto de IA. Do cruzamento dos pensamentos desses filósofos pretendemos mostrar que: 1) metafísicas de caráter dicotômico se fazem presentes no debate atual sobre IA, e produzem como resultado um impasse, uma situação sem saída, que se apresenta como um limite impossível de ser superado; 2) concepções dicotômicas do mundo não conseguem apreender e explicar a complexidade que este possui como um modelo *estratificado* e um modelo de *voltas estranhas* são capazes; 3) a maior parte das críticas dirigidas ao projeto de IA podem ser consideradas uma modalidade de *monismo* que se nutre de uma relação dicotômica com o seu par de negação (natural/artificial, humano/máquina, mente/cérebro, corpo/alma); 4) algumas perspectivas de superação dos impasses em torno da viabilidade do projeto de IA, oriundos do dualismo metafísico que lhes subjaz, são frustrantes. Em geral, elas incorporam elementos das concepções que formam o par dicotômico, o que faz com que não consigam romper com o modelo da metafísica tradicional.

A dicotomia como entrave do debate filosófico

John Langshaw Austin foi um pensador do séc. XX que viu nos usos da linguagem fatores que desencadearam entraves no avanço da reflexão filosófica. Ao comparar a linguagem do homem comum com a dos filósofos identificou diferenças entre elas que o levaram a não aceitar a substituição da linguagem ordinária, nas discussões filosóficas, por uma própria da filosofia. Os termos que fazem parte do vocabulário ordinário além de terem sido submetidos a um exaustivo teste de uso, marcam distinções em seus significados que a *nova linguagem* não leva em consideração.

Ao assinalar tais diferenças, Austin destacou a necessidade de recuperar-se o emprego do vocabulário ordinário, por este ser portador de uma riqueza de distinções sobre a experiência perceptiva jamais alcançada pelos vocabulários que “[...] adotam um ponto de vista **acadêmico**” (Austin, 1993, p. 9, negrito nosso). A abordagem austiniana do problema da percepção é uma reflexão em torno dos efeitos gerados pelo *esquecimento* da linguagem ordinária ao tratarmos de alguns problemas filosóficos. Ao considerar-se a linguagem ordinária inepta para o tratamento dos problemas filosóficos nos tornamos indiferentes ao seu vocabulário e riqueza semântica.

Embora seja reconhecido como o criador da *teoria dos atos de fala*, o que destacaremos no pensamento de Austin são as duras críticas dirigidas à atividade filosófica apresentadas em *Sentido e Percepção* (1993). Nessa obra, ele examina alguns problemas relacionados às teorias da percepção que procuram responder à pergunta pelo conteúdo desta e, as compara com as possíveis respostas formuladas pelo *homem comum* para a este mesmo conjunto de questões. Este domínio da reflexão filosófica é marcado pelo enfrentamento do *realismo* com o *idealismo*, que se confrontam de tal modo que ambos expressam o maior grau de antagonismo de um para o outro, onde um propõe a anulação do outro.

Tais teorias partem de um mesmo princípio, de que aquilo que percebemos se trata sempre de uma e mesma coisa. Para o realismo o que percebemos são *coisas materiais*, enquanto para o idealismo são *dados dos sentidos*. Essa discussão é tão antiga que, para Austin, já não é possível acessar-se aos textos originais que lhe deram início, “[...] ao

contrário, por exemplo, de doutrinas acerca dos “universais” – já eram bastante antigas no tempo de Platão”. (AUSTIN, 1993, p. 8). Por essa razão, Austin escolhe pensadores contemporâneos seus para buscar identificar as premissas básicas que sustentam ambas posições. Para ele, estes lhe:

[...] parecem fornecer as melhores exposições disponíveis das razões aceitas para sustentar doutrinas que remontam pelo menos a Heráclito, mais completas, coerentes e exatas em sua terminologia do que, por exemplo, as que se encontram em Descartes e Berkeley. (AUSTIN, 1993, p. 7)

A impossibilidade de acesso aos textos que deram origem às doutrinas da percepção não representa um impedimento que torna a reflexão sobre o seu conteúdo uma tarefa sem nenhuma possibilidade de alcançar algum resultado positivo. Austin não vê problema com a falta de acesso às fontes originais dessas doutrinas, pelo contrário; as fontes das quais ele lança mão representam um aprimoramento daquelas, garantindo uma vantagem àquele que se propõe a analisá-las.

De acordo com este filósofo, não há um acordo sobre as razões desta diferença entre elas, pois “[...] parecem divergir quanto a saber se a distinção principal que estabelecem é entre duas linguagens ou entre duas classes de entidades”. (AUSTIN, 1993, p. 8)

Não seria a mera adoção de um vocabulário mais recente a solução para o impasse, como também não seria a eliminação de todo um vocabulário já disponível. Austin não era contrário à adoção de novos termos ao longo que temos a nossa disposição para pensarmos os problemas filosóficos, pelo contrário. O que não é aceitável, para ele, é o abandono de uma linguagem em detrimento de outra por se considerar essa mais moderna do que aquela. Em várias ocasiões ele enfatiza que uma das principais características da linguagem ordinária é a de ser constantemente renovada com a incorporação de novos termos e significados após estes serem submetidos a um exaustivo teste de uso.

Mas, mesmo assim, é aconselhável ter sempre em mente (*a*) que as distinções incorporadas ao nosso cabedal de palavras comuns – vasto e, em grande parte, relativamente antigo – não são nem poucas, nem sempre muito óbvias e, quase nunca, tão-somente arbitrárias; (*b*) que, seja como for, antes de introduzir qualquer alteração por nossa conta,

precisamos descobrir com que estamos a lidar; e (c) que essa alteração das palavras, que se passa naquilo que vemos como um pequeno recesso do campo semântico, é sempre passível de repercussões imprevistas no território adjacente. A alteração não é, de fato, tão simples como se supõe, não se justifica nem é necessária com a frequências que se lhe atribui, e em geral é vista como necessária exatamente porque aquilo de que dispomos já se acha erroneamente representado. (AUSTIN, 1993, p. 89)

Seu método de abordagem dos problemas filosóficos dá relevância aos diversos significados atribuídos e atribuíveis aos termos que constituem o vocabulário com o qual expressamos nossos pensamentos. A adoção de um novo vocabulário com o consequente esquecimento do antigo representa, antes de tudo, uma perda.

No entanto, isso não significa dizer que tudo aquilo que faz parte do nosso acervo de linguagem deve ser mantido em uso. Sua abordagem do antagonismo entre realismo e idealismo visa mostrar que o mesmo não expressa uma relação de oposição genuína. Por essa razão, um dos principais objetivos do seu trabalho é mostrar que alguns usos de termos e significados não passam de ilusões de linguagem. Torna-se necessária, portanto, a remoção do uso dos mesmos, visto que não auxiliam a reflexão filosófica na superação do impasse gerado por esse tipo de engano.

[...] dissolver as inquietações filosóficas (*algumas* espécies de inquietação filosófica, não a filosofia inteira), e, também, algo acerca dos significados de algumas palavras (“realidade”, “parece”, “tem aparência de”, etc), as quais, além de filosoficamente escorregadias, são interessantes por si mesmas. Além disso, não há nada tão obviamente maçante quanto a constante repetição de asserções que não são verdadeiras, e que, às vezes, são desprovidas da mínima sensatez; se pudermos reduzir um pouco esse estado de coisas, tanto melhor. (AUSTIN, 1993, p. 12)

Austin está convencido de que o espaço do debate filosófico está cheio de hábitos, argumentos e usos de linguagem que permanecem sem uma razão que justifique a manutenção dos mesmos. Em *Sentido e Percepção* são várias as referências à necessidade de se levar em consideração o maior número possível das diversas situações de uso de uma palavra, como, também, da atenção que deve ser dada aos avanços nas pesquisas relativas aos fenômenos perceptivos alcançados pelas várias áreas do conhecimento.

Austin não acredita que possa haver um avanço na discussão enquanto ela ficar submetida a um arranjo dicotômico. Sua crítica à relação realismo/idealismo enfatiza que o debate, neste caso em torno do conteúdo das nossas percepções, perde qualquer perspectiva de avanço se não for feito um movimento de saída deste círculo, um movimento que ao ampliar as perspectivas de abordagens nos permitam, ao mesmo tempo, ver o quanto limitado é o território demarcado pelos muros daquelas duas doutrinas. Sua crítica à dicotomia é apresentada por meio de uma análise da linguagem. Nessa análise ele busca mostrar que quando o jogo de linguagem é jogado de forma dicotômica, produz-se a falsa constatação de que o debate esteja limitado pelas teses que nele se antagonizam. No máximo poderá haver uma miríade de outras sub-doutrinas que estariam dentro do espectro de ideias delimitado pela relação dicotômica em questão. Seja qual for a pergunta dirigida aos realistas ou, aos idealistas, a respeito do conteúdo das representações sensíveis, eles sempre responderão se tratar ou de “coisas materiais”, para os realistas, ou de “dados dos sentidos”, para os idealistas. Para ambas doutrinas, todas as questões e considerações se dirigem para uma só resposta. Esta convergência é, na opinião de Austin, um empobrecimento das abordagens que poderíamos empreender no debate.

Austin constata que o procedimento convergente de organização do pensamento dá-se, no meio filosófico, desde os seus primórdios. As razões para esta forma de proceder parecem dever-se à clareza e simplicidade, que dão ao esquema uma aparência de maior rigor e precisão de abordagem. E, também, a uma preferência pela criação e desenvolvimento de um vocabulário que não esteja influenciado por ideias que representassem uma certa antiguidade nociva que prejudicasse o avanço da discussão.

O que acontece, como procurarei mostrar, é que as palavras correntes são muito mais sutis em seus usos, e marcam muito mais distinções do que as vislumbradas pelos filósofos, e que os fatos da percepção, tal como descobertos, por exemplo, pelos psicólogos, mas também pelo comum dos mortais, são muito mais diversos e complexos do que se tem pensado. (AUSTIN, 1993, p. 10)

Quando o procedimento de convergência é amplamente empregado, é preenchida uma das condições que favorece a aceitação das dicotomias. Não são somente os

significados dos termos de linguagem que são submetidos a essa forma de ordenamento do pensamento, mas também as próprias experiências perceptivas. Importante destacar que essa influência não se dá exclusivamente sobre o homem comum, mas, sobretudo, com os próprios estudiosos dos fenômenos da percepção.

Em Lógica a *convergência* é uma estratégia de montagem de um procedimento de sustentação em que os elementos que exercem a função de sustentação são de conteúdos e origens distintos, dando à construção uma solidez maior, dada a possibilidade de ser sustentada por diferentes razões. Contudo, não é este tipo de convergência que conferimos à crítica de Austin dirigida à dicotomia. A convergência dicotômica é uma raiz invertida, ela reduz a pluralidade das origens das razões como também, elimina a pluralidade de consequências que podem ser advindas de um conjunto de premissas. Em nossa opinião, Austin nos mostra a convergência como um arranjo que envolve o movimento de uma unidade para outra, enquanto na Lógica ela envolve o movimento de uma pluralidade referindo-se a uma mesma unidade. Na Lógica, mostra-se que de uma pluralidade de razões poderemos chegar a uma mesma consequência. No modo como o criticado por Austin o que se dá é o inverso, não é a pluralidade o ponto de partida, mas a progressiva rejeição da possibilidade de muitas outras razões.

Não existe uma espécie de coisa que nós percebemos, mas muitas espécies diferentes, cujo número pode ser reduzido (se é que pode) pela investigação científica, e não pela filosofia: sob muitos aspectos, mas não sob todos, canetas são muito diferentes de arco-íris, e estes são diferentes sob muitos, mas não sob todos os aspectos, das imagens consecutivas que, por sua vez, sob muitos aspectos, mas não todos, são diferentes de imagens na tela do cinema – e assim por diante, sem que se lhe possa determinar um limite. (AUSTIN, 1993, p. 11)

A perspectiva austiniana é de abertura, de um conjunto de premissas poderemos inferir muitas conclusões distintas, o modelo que ela emprega é divergente. Ao invés de restringir o conjunto de possibilidades é exatamente o oposto disso que garantirá uma compreensão mais robusta de mundo.

Isso pode ser visto na análise dos significados de dois termos bastante empregados nas discussões travadas entre o realismo e o idealismo. O propósito de Austin ao fazer a análise dos diferentes significados de expressões como *parecer* e *real* é o de nos alertar

para a perda ou abandono de significações que essas expressões possuem. Vale lembrar que Austin, com sua perspectiva divergente, empreende um esforço para que seja recuperada a característica polissêmica dessas expressões. Não há razão suficiente para que se escolha um dos seus significados como sendo capaz de cobrir todo campo semântico que uma expressão possui. Esta situação de empobrecimento semântico é fundamental para a construção de dicotomias, pois essas se nutrem de uma eliminação de possibilidades de ser para se imporem enquanto tal.

A análise dos usos de *parecer*, por exemplo, dá-se em função da crítica que Austin dirige à obra de Alfred Jules Ayer (1910-1989) *The Foundations of Empirical Knowledge*, na qual este se posiciona em favor do idealismo, afirmando que Ayer

us[ou] com muita liberdade as expressões “parecer”, “ter aparência de” (*look*), “parecer”, “afigurar-se” (*appear*), “parecer”, “dar a impressão de” (*seem*) - aparentemente como faz a maioria dos outros filósofos, **sem atribuir grande importância à questão de saber qual expressão é usada onde**; na verdade deixa implícito, devido à rapidez do seu vôo filosófico, que tais expressões podem ser usadas de modo intercambiável, que não há muito a escolher entre elas. (AUSTIN, 1993, p. 48, grifo nosso)

É bastante evidente que a construção de dicotomias depende da adoção de uma determinada maneira de lidar com palavras e ideias, onde as diferentes forças que aí operam se dirigem para um mesmo ponto, tendo por consequência a eliminação de outros pontos possíveis. No caso de Ayer, a não distinção dos diferentes significados do termo “parecer” servirá de ponto de apoio para que se posicione a favor de um dos membros que forma o par dicotômico “dados dos sentidos/coisas materiais”.

[...] não é assim; na verdade, as expressões em questão têm usos inteiramente diferentes, e o uso de uma ou outra implica uma grande diferença. Nem sempre é verdade – pois há casos, como veremos, em que usar uma ou outra dá praticamente no mesmo, e contextos em que as expressões têm um uso mais ou menos intercambiável. Mas, devido à existência de tais casos, seria um erro concluir que não existe nenhuma diferença específica no uso das palavras; essa diferença existe, e há um grande número de contextos e construções que o demonstram. (AUSTIN, 1993, p. 48)

Austin, por outro lado, parte da ideia que há relação de sinonímia entre termos da linguagem, situações de uso em que o emprego de uma ou outra não gera alteração do significado que está em uso. Ele identifica três níveis de intercambialidade entre expressões. No primeiro não há intercambialidade entre os termos; no segundo há total intercambialidade e, no terceiro, a intercambialidade é parcial. Iremos nos deter nos casos em que não há intercambialidade visando mostrar que, pelo menos nesses, a perspectiva de convergência semântica gera equívocos. Ao dar relevância à abertura semântica, Austin expõe, assim, a fragilidade do procedimento dicotômico cuja característica principal é a simplificação.

Ao analisar as expressões *look*, *appear* e *seem*, Austin leva em consideração aspectos de natureza sintática para apontar os casos em que não é possível o uso intercambiável destas expressões de linguagem. Tais características servem de indícios para apontar diferenças semânticas sutis que, somente por meio de uma análise criteriosa, como a realizada por Austin, podem ser percebidas. A comparação, entre as expressões acima mencionadas, mostra-nos que umas possuem mais riqueza de participação em construções de linguagem do que as outras. Austin nos expõe uma assimetria semântica, por exemplo, entre *look* e *appear*, o que torna temerária a substituição de uma pela outra. Ele identifica quatro situações de uso para *look* e sete para *appear*. As diferentes situações de uso das expressões devem ser levadas em consideração se não quisermos cair no erro de intercambiarmos livremente expressões que não possuem simetria semântica.

Este procedimento, de levar em conta as diferentes situações de uso e significados, salienta a característica de riqueza semântica que as expressões possuem. A forma austíniana de organizar o cenário desta discussão é diametralmente oposto àquele que sustenta a construção de dicotomias. Enquanto, para Austin, uma expressão abre um leque de opções significativas, por mais simples que seja sua carga semântica; para a concepção de uso da linguagem criticada por ele, o que se dá é exatamente o oposto, em havendo um grande número de possibilidades significativas, elimina-se ou desconsidera-se suas diferenças, fazendo-as, com isso, convergir para uma só. Portanto,

[...] mesmo sem entrar em grandes detalhes, fica bastante claro que as idéias básicas por trás dos usos de *look*, *appear* e *seem* não são as

mesmas; e, com muita freqüência, onde se usa uma palavra não cabe usar a outra". (AUSTIN, 1993, p. 54)

Mesmo em se tratando de um conjunto tão restrito de expressões o que se afere é que há uma riqueza de significados tal que se torna um desafio empreender qualquer tipo de substituição de uma pela outra.

Na argumentação de defesa da tese do idealismo formulada por Ayer, é apresentado o “argumento da ilusão, em que são empregados termos como *look* e *appear*. Tal argumento tinha por propósito inferir que o conteúdo das nossas percepções são os “dados dos sentidos” e,

Baseia-se no fato de que as coisas materiais podem apresentar **aparências diferentes** a diferentes observadores, ou ao mesmo observador em diferentes condições, e que a natureza destas aparências é, até certo ponto, casualmente determinada pelo estado das condições e do observador. (AUSTIN, 1993, p. 31, negrito nosso)

A princípio, a intenção de Ayer é mostrar que algumas percepções não são seguras o suficiente para que tenhamos certeza de que elas estão nos fornecendo informações sobre as qualidades genuínas de uma determinada coisa. Colocando-se ao lado do pensamento cartesiano, que põe sob suspeição os indícios gerados pela experiência sensível, Ayer estende sua constatação, isto é, não são somente algumas percepções que podem ser classificadas como ilusórias, mas todas elas. Por essa razão Austin faz a seguinte advertência:

Desejo agora chamar a atenção, antes de mais nada, para o nome desse argumento – “argumento da *ilusão*” – e para o fato de ser apresentado como se estabelecesse a conclusão de que pelo menos algumas de nossas “percepções” são *enganosas*, pois há nessa conclusão duas claras implicações – (a) a de que todos os casos citados no argumento são casos de ilusões, e (b) a de que ilusão e delusão são a mesma coisa. É evidente, porém, que essas duas implicações são totalmente errôneas, e é de grande importância chamar a atenção para isso, pois como veremos, o argumento tira partido de uma confusão exatamente nesse ponto. (AUSTIN, 1993, p. 34)

Ao criticar o argumento que serve de fundamento para sustentar a tese de que o

conteúdo das nossas percepções são “dados dos sentidos”, Austin emprega a mesma estratégia da crítica dirigida por ele à intercambialidade dos termos *look* e *appear*, como vimos acima. Sua crítica enfatiza que as estratégias empregadas pelos defensores, tanto do idealismo quanto do realismo, caracterizam-se por serem procedimentos convergentes. Procedimentos esses que, à medida que são executados, eliminam a diversidade de experiências perceptivas, resultando no surgimento de um novo elemento amalgamado, possibilitado pela junção dos subprodutos, oriundos do processo de eliminação.

Dois fatores concorrem para tal realização, o procedimento de generalização e a falta de atenção aos significados de expressões importantes na formulação da demonstração. Na avaliação que faz da argumentação de Ayer, Austin aponta o procedimento de generalização sendo empregado ao Ayer considerar todos os exemplos apresentados por ele como se tratando de casos de ilusão. Isso é consequência da falta de distinção entre *ilusão* e *delusão*. A crítica de Austin é apresentada em duas etapas: na primeira é mostrado que Ayer não é cauteloso ao escolher os casos a serem analisados para sustentar que o conteúdo das nossas percepções são *dados dos sentidos*, pois trata alguns casos como se fossem de ilusão, quando não o são. O ponto de partida de Ayer é um erro de sentido. Na segunda etapa, Austin justifica sua afirmação do erro de Ayer ao apontar a desconsideração que este faz da distinção entre ilusão e delusão.

O argumento da ilusão se apoia em casos em que a experiência perceptiva mostra-se *fallha*, pois não apresenta as qualidades *reais* dos objetos percebidos nestas condições. Austin salienta a adoção do procedimento de generalização ao mostrar que nem todos os casos apresentados por Ayer, como sendo casos de ilusão, podem ser considerados como tais. Dos três casos selecionados por Ayer, um deles não pode ser considerado um caso de ilusão, mas de delusão.

Austin compara dois exemplos, dados por Ayer como sendo casos de ilusão, o famoso caso do *bastão dentro d'água* e o caso das miragens. A primeira constatação é de que em ambos Ayer pressupõe que alguma coisa é percebida nos dois tipos de experiências. De acordo com Austin, isto faz parte da estratégia argumentativa para nos colocar em condições de aceitar a tese de que tudo que percebemos são, na verdade, dados dos sentidos e que toda experiência perceptiva é percepção de alguma coisa. Não haveria

experiência perceptiva sem algo que pudesse ser percebido.

No caso do bastão dentro d'água a explicação de Ayer explora uma contradição. O bastão ora se mostra reto, ora se mostra curvo. Sua explicação afirma que isso só é possível porque há um tipo de coisa, que não é material, que serve de base para a percepção que não se constitui como uma qualidade genuína do bastão. Vale lembrar, são considerados como ilusões porque não propiciam acesso às qualidades reais daquilo que esteja sendo percebido.

No caso das miragens, a argumentação de Ayer apresenta-se como a explicação capaz de dar conta de algumas experiências perceptivas que causam embaraço às explicações que pressupõem as “coisas materiais” como sendo os conteúdos das experiências perceptivas, já que fica difícil afirmar que seja esse tipo de entidade o constituinte do conteúdo de uma miragem. A vantagem dos “dados dos sentidos” é que esses não necessitam possuir um correlato que se faça presente numa outra estrutura que não a perceptiva.

Uma pessoa que vê uma miragem, diz Ayer, “não percebe qualquer coisa material, pois o oásis que pensa perceber *não existe*”. Mas “a sua *experiência* não é uma experiência de nada”; assim, diz-se que apreende dados dos sentidos que são semelhantes, por sua natureza, àquilo que estaria apreendendo se estivesse vendo um oásis de verdade, mas que são enganosos no sentido de que a *coisa material que parecem apresentar* não está *realmente lá*. (AUSTIN, 1993, p. 33) (negrito nosso)

A argumentação de Ayer estabelece uma correlação entre fenômenos, quando, de fato, não o são. Ver um bastão dentro d'água e ver uma miragem não são, ambos, casos de ilusão, destaca Austin. O procedimento de convergência de possibilidades é empregado por Ayer, que tipifica todos os casos de conteúdo das nossas percepções como sendo de um só tipo. Parece que Ayer escolhe dentre duas premissas a mais forte. Em sua opinião, os dados dos sentidos se constituem como premissas mais fortes porque podemos constatar que, em muitos casos, não há uma *coisa material* da qual parecíamos ter percebido algo, pois a coisa percebida não existe, o que então foi percebido? De acordo com o pensamento de Ayer, mesmo nos casos em que há percepção de uma *coisa material* o conteúdo desta percepção tem de ser um *dado dos sentidos*, já que ele está presente

inclusive naquelas percepções em que não há “coisa material” sendo percebida. Como uma *coisa material* não está presente em todas as experiências perceptivas, então, este conceito não pode ser considerado o conteúdo básico das nossas representações.

Contudo, ao Ayer desconsiderar a distinção entre *ilusão* e *delusão*, lança por terra todo seu esforço de explicação ao reduzir seu horizonte de significação. Há uma nítida diferença de significado entre esses dois termos que não nos permite desconsiderá-la. A distinção é simples, num caso o sujeito de observação está em pleno funcionamento, não há problemas com sua estrutura perceptiva, nem com o seu intelecto; enquanto no outro, a percepção é produzida pelo mau funcionamento das estruturas perceptivas e cognitivas que constituem o sujeito de observação.

Aqui, as diferenças mais importantes são que a expressão “uma ilusão” (num contexto perceptivo) não sugere que uma coisa totalmente **irreal** seja *produzida por um passe de mágica* – pelo contrário, o que ali está é uma disposição de linhas e setas na página, a mulher no palco com a cabeça num saco preto, as rodas giratórias; ao passo que o termo “delusão” **realmente** sugere algo de totalmente **irreal**, que ali não está de modo algum. (As convicções da pessoa que tem delírios persecutórios podem ser *completamente* desprovidas de fundamento.) Por esse motivo é que as delusões são um caso muito mais grave – algo está **realmente** mal, e, o que é pior, mal quanto à pessoa que os tem. Mas, quando tenho uma ilusão ótica, por mais perfeita que seja, nada há de errado comigo pessoalmente, a ilusão não é uma pequena (ou grande) peculiaridade ou idiossincrasia de minha parte; é completamente pública, todos a podem ver, e, em muitos casos, podem estabelecer-se procedimentos clássicos para produzi-la. Além disso, se não queremos nos deixar levar pela ilusão, devemos estar de *sobreaviso*; de nada adiante, porém, dizer ao que sofre de delusões que esteja de sobreaviso. O que ele precisa é ser curado. (AUSTIN, 1993, p. 35) (negritos nossos)

A crítica de Austin ao argumento da ilusão afirma que este argumento gera uma rede de implicações que, em três etapas, chega à conclusão de que os casos ilusórios geram enganos a partir de algo que é considerado irreal ou imaterial. Isto se dá porque não é feita a distinção entre ilusão e delusão. Por não fazer tal distinção, Ayer infere que, do mesmo modo que nos casos delusivos, também nos casos de ilusão, algo que não pode ser considerado uma *coisa material* está sendo percebido.

O modo como o “argumento da ilusão” explora a falta de distinção entre ilusões e delusões é, segundo penso, esse. Enquanto se sugere que os casos para os quais nos chamam a atenção são casos de “ilusão”, existe a implicação (derivada do uso comum da palavra) de que ali existe, **realmente** algo que nós percebemos. Mas então, quando esses casos começam tranquilamente a ser chamados de enganosos, insinua-se a sugestão muito diversa de algo sendo produzido magicamente, algo **irreal** ou de qualquer forma, “imaterial” (AUSTIN, 1993, p. 37, negritos nossos)

O uso da palavra *real* possui grande relevância em toda esta argumentação e é o uso descuidado dela que faz com que Ayer incorra em erros de avaliação que colocam seu esforço por terra. Na análise deste termo, Austin faz muitas considerações, algumas parecendo entrar em choque com as outras. Por exemplo, ao mesmo tempo em que afirma ser *real* uma palavra normal, ele também a considera “[...] de forma alguma, uma palavra normal; mas sim extremamente excepcional”. (AUSTIN, 1993, p. 90) Ele considera *real* uma palavra normal, porque ela já possui um significado de tal forma estabelecido que não permite aos seus usuários mudá-lo de acordo com as suas vontades, “ao sabor dos caprichos individuais” (Austin, 1993, p. 88) e, por ser antiga o suficiente para que não a consideremos uma palavra inovadora, recente, ou atual. Por sua vez, também, considera *real* uma palavra fora do comum, porque “não possui um significado único, especificável e sempre o mesmo”. (AUSTIN, 1993, p. 90) Esse duplo caráter do significado de “real” indica o grau de dificuldade que seu uso impõe. Portanto, se não for usado com a precisão que o seu significado exige, a consequência será um desastre semântico. Confundir o uso destes dois significados de *real*; um claro e preciso, o outro sutil e variável, implicará a falta de sentido da proposição

A pretensão de se introduzir novas distinções de significado neste tipo de palavra requer que se dê atenção às distinções já presentes no vocabulário disponível, sendo necessário um conhecimento aprofundado do tema que nos impele a querer marcar uma nova distinção e das alterações que estas mudanças produzirão nos *campos semânticos* adjacentes. Austin não julga ser necessária uma alta frequência de alteração nos significados dos termos do vocabulário disponível, sendo. A linguagem necessita de uma certa estabilidade para exercer com eficiência sua função. Ele está convencido que não devemos descartar as riquezas de significados produzidas ao longo da história de uso de

um determinado vocabulário. Entretanto, sua concepção de linguagem se caracteriza por entendê-la como abertura significativa, ou seja, é impossível impedir a introdução de novos elementos e distinções num vocabulário, porém, devemos observar que toda e qualquer tentativa de introdução de uma inovação só deverá ser aceita, se e somente se, de fato ela for pertinente. Na verdade, o que acontece é que, na maior parte das vezes, o sentimento de necessidade de marcar uma nova distinção revela-se mais um desconhecimento de distinções já marcadas.

O uso de *real* é considerado complexo por Austin. Sua complexidade se deve ao fato de que *real* se apresenta combinado com outros termos que não possuem nada em comum uns com os outros. Ele distingue dois contextos em que empregamos o termo em questão, um em que não sabemos com clareza qual significado deve ser levado em consideração se pretendemos apreendê-lo adequadamente e, outro, em que as condições de apreensão são sabidas. Na sua análise são apresentados casos de usos do termo *real* em que não temos como saber exatamente qual o significado. Embora, num primeiro momento, pareçam fazer sentido, determinados empregos desta expressão se mostram embaraçosos quando submetidos a uma análise mais minuciosa. Ao mostrar e analisar alguns usos de expressões formadas pela combinação de *real* com a palavra *cor*, Austin identifica sete diferentes situações de uso em que esta mesma combinação possui significados distintos em cada uma delas.

Podemos observar que a perspectiva austiniana é de enfatizar o caráter divergente dos significados dos termos empregados nas explicações sobre o conteúdo das percepções. Ao se fazer uso de uma expressão, provocamos uma abertura de sentido que se apresenta como múltipla. Estes casos são classificados por ele como aqueles em que não sabemos exatamente como devemos proceder para termos clareza de quais sejam os significados que lhes são atribuídos.

Dos sete destacamos o terceiro tipo em que a expressão *cor real* não nos permite compreender qual o seu significado adequado. Quando pergunta pelo significado da pergunta: qual a cor real de um peixe multicolorido que numa determinada profundidade parece ser brilhante, mas, quando pescado, podemos perceber, à luz do dia, que sua coloração é branco-acinzentado. A resposta de Austin é cuidadosa, mas não deixa de ser

espantosa, porque afirma que dificilmente saberemos responder sobre a cor real de um ser assim. Ou seja, parece que, embora empregado, o uso de *real* não faz muito sentido num contexto como este, pois não auxilia em nada a compreensão sobre este aspecto, possuir uma cor. Em comparação com outros usos deste termo, a situação é totalmente diferente, devido ao fato de podermos dizer, com razoável clareza, qual o significado que ele possui ao fazer-se um determinado uso dele.

“Estes diamantes são reais?”, “É um pato verdadeiro?” Objetos de joalheria que se assemelham mais ou menos fielmente a diamantes podem não ser diamantes reais por serem de massa vítreou de vidro; o pato pode não ser real porque se trata de uma imitação, ou de um pato de brinquedo, ou de uma espécie de ganso muito parecida com pato, ou porque estou tendo uma alucinação. (AUSTIN, 1993, p. 94)

O uso descuidado de “real” produziu como consequência a sinonímia entre “ilusão” e “delusão”, entre o ato de engano dos sentidos e um devaneio ou alucinação. Isto porque o objetivo do argumento da ilusão é mostrar que existem casos em que não há uma coisa material sendo observada, mas sim algo de outra natureza. Numa percepção ilusória nada vai mal, nem com o observador, nem com as condições de observação e, nem com o que é observado. A “normalidade” de uma ilusão reside no fato de que quem é submetido a ela pode tomar consciência de que está sendo iludido, sem que isso altere a experiência perceptiva. O que produz uma delusão são alterações ocorridas com o observador, cuja capacidade de tomar consciência do que se passa está totalmente prejudicada.

A análise austiniana do argumento da ilusão nos mostra que algumas palavras são mais exigentes do que outras quando lhes fazemos uso, “real” é uma delas. Ao não levar em consideração a distinção entre “ilusão” e “delusão” e os múltiplos e variáveis significados de “real”, Ayer acreditou ser possível posicionar-se com segurança em favor de um dos pares da dicotomia “dados dos sentidos/coisas materiais”.

Em suma, a contribuição do pensamento de Austin na realização deste trabalho se dá através de duas advertências feitas por ele em “Sentido e Percepção”: Primeiro, que as dicotomias não são um modo de arranjo de ideias que nos possibilita a apreensão e compreensão adequadas dos fenômenos; segundo, que devemos dar a devida atenção à riqueza semântica que o vocabulário comum nos põe à disposição. Se levarmos em

consideração essas advertências, estaremos menos propensos a cometer equívocos que alguns usos de linguagem suscitam.

A estratificação como modelo de compreensão

Reinhard Koselleck aplica o modelo de estratos, termo derivado da geologia, para definir em que consiste o trabalho do historiador, cuja principal tarefa é procurar elaborar um enredo extraído de um aparente caos. A teoria geológica na qual Koselleck se inspira foi formulada pelo dinamarquês Nicolas Steno o qual no século XVII afirmou que as camadas de rochas mais antigas são sobrepostas pelas mais recentes, seguindo uma ordem cronológica e que só é alterada pelo movimento da crosta terrestre.

De acordo com o seu proposito, uma concepção estratificada do tempo apresenta vantagens sobre as concepções tradicionais que concebem o tempo como linear ou circular, porque não pressupõe o tempo como algo ao qual se possa atribuir um e somente um único tipo de movimento. Essas duas concepções, que são antagônicas e excludentes, compreendem o tempo como algo no qual só se realiza um tipo de movimento; se para a primeira o tempo não possibilita retorno algum, para a segunda é exatamente o retorno a marca da temporalidade.

Esse modelo [circular], que destaca o retorno do tempo, é freqüentemente atribuído aos gregos; em contraposição, judeus e cristãos teriam desenvolvido o modelo linear. Momigliano já demonstrou que essa oposição é ideologicamente enviesada. Ambos os modelos são insuficientes, pois toda a seqüência histórica contém elementos lineares e elementos recorrentes. A circularidade também deve ser pensada em termos teleológicos, pois o fim do movimento é o destino previsto desde o início: o discurso circular é uma linha que remete a si mesma. (KOSELLECK, 2014, p. 19)

Assim como para Aristóteles na *Física* (IV, 219 b1), o “tempo” é considerado a percepção das diferenças que distinguem um antes de um depois, para Koselleck o “tempo” se caracteriza pela mudança, seja pela diferença entre os seus constituintes, seja pela diferença de ordenamento destes mesmos constituintes e por ser uma categoria

hermenêutica que se altera conforme a interpretação dada ao relato daquilo que nele se desenvolve. Para o discípulo de Gadamer, a atividade do historiador só se torna possível a partir de um relato do que acontece e de uma reflexão sobre este relato. O “tempo” não é o próprio movimento, mas só pode ser percebido através deste.

A abordagem estratificada pretende dar conta das lacunas interpretativas originadas pelo procedimento de exclusão empregado pelas concepções tradicionais do tempo. O conceito de estratificação nos permite relacionar elementos temporais que não parecem, à primeira vista, ter nexo algum. Contudo, isso não implica haver uma relação de total dependência, entre si, dos elementos que formam o arranjo estratificado.

Tentarei em seguida compreender os resultados históricos por meio de uma estratificação temporal que pretende solapar a oposição entre o linear e o circular. Os tempos históricos consistem em vários estratos que remetem uns aos outros, mas que não dependem completamente uns dos outros. (KOSELLECK, 2014, p. 19)

A concepção estratificada do tempo histórico deve ser recebida como um elemento de linguagem. Koselleck a apresenta como metáfora, isto é, uma representação de linguagem que nos permite a compreensão de algo daquilo que tomamos como um objeto de interesse. Assim, os esforços do historiador não se confundem com os de outras formas de conhecimento que se debruçam sobre o desenvolvimento dos acontecimentos. A atividade do historiador aproxima-se da atividade do poeta, pois ambos precisam construir imagens como representação dos fatos históricos, recurso sem o qual não há a possibilidade de compreensão de algo que não possui forma e cujo conteúdo é difuso.

Na condição de historiador, devo adiantar que não tenho capacidade para fazer afirmações de natureza física ou biológica. Situo-me no campo das metáforas: a expressão “estratos do tempo” remete a formações geológicas que remontam a tempos profundamente diferentes, que se transformaram e se diferenciaram umas das outras em velocidades distintas no decurso da chamada história geológica... Sua transposição para a história humana, política ou social, permite separar analiticamente os diversos planos temporais em que as pessoas se movimentam, os acontecimentos se desenrolam e os pressupostos de duração mais longa são investigados. (KOSELLECK, 2014, p. 19)

Ao afirmar que se situa no campo das metáforas, a perspectiva de Koselleck também se aproxima do pensamento de Giambatista Vico que, na *Ciência Nova* (1725), defende a tese de que a compreensão do direito natural, “[...] tal como os juri-consultos romanos o definem: Direito ordenado pela Providência divina, com princípios sobre todas as necessidades ou utilidades humanas, igualmente observado em todas as nações” (VICO, 2006, posição 362, Kindle) só pode ser alcançada pelo estabelecimento de uma relação das leis com Deus. Por não estabelecerem esta relação com o divino é que Grocio, Selden e Pufendorf malograram nas suas reflexões sobre o direito natural. Segundo Vico, nenhum conjunto de indivíduos, ao longo da história, decidiu formar uma nação prescindindo de uma religião que venere uma divindade. Para que haja um conhecimento rigoroso do direito dos povos é necessário o conhecimento das verdades metafísicas que demonstram a existência de Deus. Se, por um lado, não é possível haver conhecimento válido das coisas físicas sem o respaldo das verdades abstratas da matemática, por outro, não há possibilidade de conhecimento científico das coisas morais sem o amparo das verdades abstratas da metafísica.

Há, contudo, uma relação problemática de Vico com a matemática; esta é uma ciência, porém não serve como base metodológica para as outras ciências, como propunha Descartes. Para o filósofo napolitano, nem todos os feitos humanos (como a política, a economia ou a poesia) podem ser compreendidos por métodos matemáticos ou pela pura abstração racional, filosófica, da qual vem o conhecimento da verdade, como no caso do raciocínio exato. Entretanto, estes feitos humanos também não podem ser considerados apenas como fatos, como a física fazia com os fenômenos naturais, pois esses fatos são também feitos. Um objeto de estudo dessa “natureza” requer uma ciência diferente, a qual não teria um grau de racionalidade menor, pois também se utilizaria da filosofia, e nem seria menos verdadeira, pois investigaria feitos humanos. Esta ciência diferente seria a Ciência Nova, ou seja, seria uma terceira forma de cognição, cujo objeto são os *fatos feitos* pelos homens e que, por conta disso, parte do certo para se atingir a verdade. (LENZI e VICENTINI, 2002, p. 203)

Embora se inspire num modelo de conhecimento oriundo das “coisas físicas”, Koselleck não vai além da transposição, não adotando um método físico-matemático para compreender o tempo. A abordagem do tempo histórico exige um esforço de

compreensão dos sentidos possíveis das narrativas das ações humanas. Assim como pensava Vico, tal esforço situa-se no plano do discurso, da linguagem, e não da matemática.

Para que seja levado a bom termo o esforço de atribuição de sentido aos acontecimentos que se dão no tempo, Koselleck propõe três tipos de estratos temporais, os quais se relacionam às experiências dos indivíduos e das gerações. O primeiro é a “singularidade” ou a experiência dos “[...] acontecimentos como ocorrências surpreendentes e irreversíveis[...]” (KOSELLECK, 2014, p. 21), o tempo histórico é concebido como uma experiência de indivíduos e gerações dos acontecimentos dados no tempo, dados de forma especial, pois servem de marcadores das etapas temporais. O segundo são as “estruturas de repetição” que tornam possíveis as singularidades. Tais estruturas garantem as condições de aparecimento dos acontecimentos singulares, pois “[...] sem o retorno do mesmo – ou, pelo menos, de algo semelhante, conforme um planejamento – e sem organização seria impossível que eventos singulares se realizassem”. (KOSELLECK, 2014, p. 21). O terceiro é a “transcendência” ou permanência de uma determinada experiência do tempo histórico,

[...] não no sentido de um além, mas no sentido de que se estendem por várias gerações. Todas as unidades de experiência precisam de um mínimo de transcendência: sem ela não haveria uma explicação última – por mais provisória que esta possa ser – e nenhuma experiência poderia ser convertida em ciência. (KOSELLECK, 2014, p. 25)

O cenário no qual se movimenta o historiador é descrito por Koselleck como um contexto de descoberta. Nessas condições, o historiador deverá realizar três ações se pretende alcançar um conhecimento rigoroso e necessário. Entretanto, este conjunto de procedimentos não é suficiente por si só para dar o caráter de ‘verdade histórica’ das experiências no tempo.

O que é *buscado, encontrado e apresentado* como *verdade histórica* nunca depende somente das experiências feitas pelo historiador, tampouco somente dos métodos que ele usa. (KOSELLECK, 2014, p. 27, itálicos nossos)

O historiador necessita de um ímpeto que o lance ao encontro do que é buscado. Sem esse impulso inicial não é possível a realização dos estágios subsequentes. Há uma sequência de acontecimentos que deve ser seguida para que o historiador consiga executar sua tarefa. A proposta de estratificação do tempo configura os próprios estágios de realização da atividade do historiador.

A ‘verdade histórica’ é algo que se situa num contexto de descoberta, em que o objeto a ser descoberto independe de qualquer outra atividade da parte de quem realiza a busca que não a disposição de realizá-la, mesmo não sabendo o que encontrará pela frente. Ele estabelece um paralelo entre o conceito grego de *historie* e o alemão de *experiência* (*Erfahrung*) para mostrar que a atividade do historiador indica a intenção de se produzir um movimento no sentido de se saber alguma coisa de algo. Isto é, ao mesmo tempo em que uma ação é produzida por um agente, um efeito desta ação é produzido neste. Algo é produzido em nós em decorrência de algo produzido por nós mesmos.

Outra observação etimológica, importante para a orientação da narrativa histórica [*Historie*]: “historia”, em grego, significava o que nós, alemães, denominamos “experiência”. “Ter uma experiência” significa ir daqui para lá a fim de experimentar e conhecer algo: tratar-se, de certo modo, de uma viagem de descoberta. Mas a narrativa histórica só surge como ciência a partir do relato dessa viagem e da reflexão sobre esse relato. Ela é, por definição, a expressão mais pura de uma ciência da experiência. (KOSELLECK, 2014, p. 20)

Nesse sentido, a atividade do historiador consistirá na busca pelos “vestígios da experiência do tempo histórico”, condição necessária para que o historiador consiga elaborar sua própria reflexão. O “tempo histórico” deixa vestígios que se fazem presentes no momento atual. Ideias, representações, compreensões, interpretações, toda atividade e esforço humanos cujo propósito último é o de entender o que acontece, continuam a produzir efeitos mesmo sendo tão antigas quanto a ideia de deus(es) e outras.

Contudo, existem numerosas repetibilidades que se estendem bem além de uma geração, além até daquela sucessão empiricamente evidente das gerações que ainda conseguem se comunicar oralmente. Esses fenômenos de repetição que ultrapassam os limites do cotidiano podem ser chamados de “transcendentais”. Situam-se aqui verdades religiosas ou metafísicas que se fundamentam em enunciados básicos que só

muito lentamente se modificam ao longo dos séculos e permanecem disponíveis, mesmo que nem todos os compartilhem. (KOSELLECK, 2014, p. 25)

As mudanças ocorridas no tempo histórico são alterações da “transcendência” das experiências históricas possibilitadas pelas “estruturas de repetição”. A tarefa do historiador de encontrar sentido na sucessão dos acontecimentos para, daí, entender como se dão as mudanças deve levar em consideração um duplo cujos membros compartilham entre si a capacidade de influenciar um ao outro, esta é, segundo Koselleck, a forma de proceder adequada ao propósito de entender as mudanças que ocorrem no tempo histórico. Koselleck acredita que a tarefa do historiador está compreendida entre dois fatores “[...] de um lado, os *pragmata*, as *res gestae*, os acontecimentos e, de outro, as histórias ou a disciplina da história”. (KOSELLECK, 2014, p. 30). As mudanças no tempo histórico decorrem das interações entre esses dois fatores que, como foi dito acima, produzem efeitos um no outro. A tarefa do historiador consiste em identificar os estratos do tempo histórico e aplicar a estes o método de análise adequado e capaz de lhe possibilitar uma atribuição de sentido aos acontecimentos que estão sob seu escopo. Isto implica que as mudanças nos acontecimentos repercutem na maneira como nós os compreendemos, assim como nossa compreensão não deixa de ser um acontecimento.

Oferecem-se então duas possibilidades de analisar autonomamente, também partindo de uma perspectiva temporal, as alterações de experiências ou de métodos, de modo a considerá-las como fator primário das mudanças. Normalmente o historiador tenderá a conceder primazia à mutação da experiência, definindo-se como mero analista ou como narrador, aquele que somente registra. Mas não pode haver dúvida de que uma experiência histórica, quando enquadrada metodologicamente, se transforma em um fator autônomo, causando efeitos muito significativos. Sem a interpretação teológica e escatológica do mundo, feita pela igreja cristã, nem a controvérsia das investiduras, como todas as suas consequências políticas, nem as Cruzadas teriam sido possíveis; tampouco as conquistas ultramarinas pela navegação cristã, nem, evidentemente, a história das guerras civis religiosas no início da era moderna. É possível menosprezar a influência direta de Maquiavel na história, mesmo que ela seja onipresente de forma indireta, mas é impossível negar que a visão histórica da Marx, elaborada com método, rigoroso ou não, tenha tido consequências históricas de caráter mundial, inimagináveis sem Marx.

Assim, em primeiro lugar, seria possível elaborar uma história imanente

dos métodos que se alimente primordialmente de suas inovações, as quais, apesar dos pressupostos que as condicionam, não seriam completamente dedutíveis deles...

A segunda possibilidade consistiria em fazer a alteração no método decorrer das alterações parciais na experiência. É fácil verificar essa hipótese, pois modificações comprováveis no ambiente social ou político coincidem com inovações metodológicas. (KOSELLECK, 2014, p. 31)

Embora pareça haver uma precedência da experiência sobre o método, ela só o é analiticamente, pois a compreensão de uma determinada experiência do tempo histórico somente será possível se houver um método com o qual se realize esta compreensão. Assim como a experiência nos impele a buscar uma forma da sua compreensão, também esta forma de compreensão produz novas experiências. Essa dupla produção de efeitos entre os acontecimentos e nossa compreensão sobre eles é a forma que Koselleck encontrou para superar a tentação de entender essa relação como sendo de natureza dicotômica. Os elementos que formam esse dualismo não interagem de forma a um eliminar o outro, pelo contrário, de acordo com esta maneira de pensar, somente poderemos compreender adequadamente o sentido de tempo histórico se formos capazes de perceber as influências das experiências sobre os métodos e destes sobre aquelas. Como o próprio Koselleck constata, há aí uma circularidade em que ambos componentes se retroalimentam. Mudanças nos acontecimentos muitas vezes são produzidas por nossa compreensão de como eles se deram. Do mesmo modo, nossa compreensão é alterada em razão dos próprios acontecimentos. A tarefa do historiador se desenvolverá na medida em que ele for capaz de perceber as sutilezas das mudanças que ora são produzidas pelos fatos históricos, ora pelos métodos empregados para compreendê-los. Embora ele deixe a possibilidade de identificar outras fontes de mudança de sentido do tempo histórico que não os próprios acontecimentos e os métodos empregados pelos historiadores.

Lançando mão de diferenciações antropológicas nos conceitos de experiência e métodos, tentaremos possibilitar a articulação entre ambos e estabelecer correlações baseadas na premissa de que a história (*Geschichte*) e a narrativa histórica (*Historie*), a realidade e seu processamento consciente, sempre remetem uma à outra e se constituem reciprocamente, mas não são completamente derivadas uma da outra. (KOSELLECK, 2014, p. 32)

Procuramos lançar mão desse tipo de abordagem ao tratarmos aqui da viabilidade do projeto de IA. Vemos nesta abordagem duas razões que nos permitem o vislumbre de uma saída para o impasse epistemológico que trava o avanço do debate em torno das capacidades humanas e suas realizações. Primeiro, por não tomar como ponto de partida uma concepção dicotômica dos elementos envolvidos na reflexão sobre a capacidade e limites do conhecimento humano; segundo, por não identificar nas mudanças a manifestação de uma e somente uma causa. Ao afirmar que a experiência no tempo e os métodos para compreendê-la provocam alterações um no outro, Koselleck não deixa espaço para que seja aceita a ideia de que cada efeito é produzido por uma e somente uma causa. A estrutura do tempo e seus momentos, bem como a tarefa do historiador, formam um complexo que não permite a adoção de modelos simples para a sua compreensão.

Para evidenciar o alto grau de complexidade da tarefa de buscar um sentido no aparente caos dos acontecimentos, não bastando a simples adoção de uma categoria para o desempenho de tal tarefa, Koselleck aborda a importância do conceito de “espaço” no desenvolvimento da pesquisa histórica. Segundo ele, há uma preferência geral dos historiadores em dar relevância ao tempo como categoria especial da história. É mais comum o estabelecimento de relação entre o espaço e o tempo do que entre o espaço e a história. Esta predominância da relação entre espaço e tempo deve-se à preferência que os historiadores dão à formulação de explicações acerca das mudanças nos acontecimentos e à influência do desenvolvimento da indústria e tecnologia ocorridas na Europa do séc. XVIII.

À primeira vista, é plausível a preferência pelo tempo, em detrimento do espaço, nas teorias da história. Primeiro, por uma razão geral: desde sempre, o historiador, ao perguntar como chegamos à situação atual, diferente da anterior, se interessa por novidades, pela mudança, por alterações. Junta-se a isso uma razão particular: a experiência da singularidade, que se impôs na Europa a partir de 1770, com o horizonte do progresso técnico-industrial e de sua veemente pressão pela mudança (KOSELLECK, 2014, p. 76)

Contudo, para Koselleck, é imprescindível que o “espaço” receba a relevância devida na elaboração de teorias históricas, pois ele também sofre alterações ao longo do

tempo.

Minha tese inicial é a seguinte: espaço e tempo representam, como categorias, as condições de possibilidade da história. Mas também o “espaço” tem uma história. O espaço é algo que precisamos pressupor meta-historicamente para qualquer história possível e, ao mesmo tempo, é historicizado, pois se modifica social, econômica e politicamente. (KOSELLECK, 2014, p. 77)

A simplificação no uso de categorias não é a estratégia adequada para a formulação de teorias históricas, pois esta forma de procedimento metodológico determina um empobrecimento e consequente perda de rigor e capacidade explicativa. O destaque dado por Koselleck à distinção entre as condições espaciais meta-históricas e os espaços históricos humanos enfatiza a necessidade de ser feita uma abordagem que não empobreça o horizonte de reflexão em torno dos acontecimentos que interferem na ação humana.

Há condições espaciais naturais que, por sua disponibilidade técnica, econômica ou política devem ser levadas em conta como condições de possibilidade da ação. O fato de a África do Sul não ter um litoral europeu à sua frente, como a Argélia, modifica fundamentalmente o status da guerra civil que ameaça o sul da África. A solução política que De Gaulle adotou não poderia ser obtida tão facilmente na África do Sul. Portanto, a situação geográfica é uma das condições do desesperado domínio da minoria branca, diferentemente do que ocorreu com o domínio dos franceses na Argélia. (KOSELLECK, 2014, p. 80)

Uma investigação histórica que pretenda oferecer uma explicação rigorosa dos acontecimentos da história humana não deve levar em consideração somente aspectos humanos propriamente ditos, mas, também, fatores que afetam a própria ação do homem no tempo. Isolar a ação humana dos fatores que lhe propiciam e causam modificações não é uma alternativa metodológica viável, pois a pesquisa histórica não deve se limitar àquilo que pertence exclusivamente ao humano. A pesquisa histórica deve ir além do humano para entender o seu agir na história.

Tucídides explica a longa duração da Guerra de Tróia pelo número insuficiente de embarcações capazes de atravessar o Mar Egeu para

conquistar a cidade.

Como se vê, as condições geográficas meta-históricas dos espaços de ação humanos se modificam, dependendo de como estão sujeitos ao domínio econômico, político ou militar. Ou, formulando-se de modo teórico: a repentina transformação de situações meta-históricas (para usar uma expressão de Ratzel) em espaços históricos faz parte da investigação de uma teoria da história. Sua utilização, implícita ou explícita, se manifesta em toda a historiografia. (KOSELLECK, 2014, p. 81)

Há um processo de historicização de todos os elementos que produzem influência na ação humana. Tal perspectiva impõe um modelo explicativo que não permite a simplificação dos componentes que possibilitam a produção de uma determinada representação dos acontecimentos. A proposta de Koselleck de tomar emprestada da geologia a concepção de estratos nos permite compreender as inconsistências dos esforços humanos de entendimento dos acontecimentos ao longo do tempo. Não estamos diante de um acontecer único que se desenrola inexoravelmente. O desenvolvimento da história parece seguir o mesmo padrão que a “pressão de seleção” na biologia. Não há um *telos* a ser buscado, um propósito último a ser alcançado, uma orientação de rumo. Os acontecimentos históricos, assim como as mudanças nos organismos biológicos, não se submetem a objetivos alcançáveis e preestabelecidos. Podemos perceber as mudanças e produzir explicações para um determinado conjunto de fenômenos que se dão no tempo, mas essa condição não nos assegura a compreensão do que realmente se passa. Essa é uma das razões que explica as diferenças de rumo das teorias que visam entender historicamente o sentido dos acontecimentos. Conforme nossa existência se dá no tempo, a necessidade de mudanças nas teorias que tratam do desenvolvimento deste acontecer se apresenta de tal forma que nossos esforços parecem produzir poucos resultados úteis, a não ser a esperança de sermos menos suscetíveis a eles.

Comparados com os 5 bilhões de anos da nossa crista planetária e comparados com o 1 bilhão de anos de vida orgânica nessa crosta, os cerca de 10 milhões de anos de existência do homem simiano representam um período curto, e os 2 milhões de anos dos quais temos prova da produção de ferramentas são ainda mais curtos. Sem dúvida, as condições geológicas e geográficas, biológicas e zoológicas, ou seja, todas as condições meta-históricas de determinação do espaço humano, exercem uma influência muito mais marcante durante essa época do que

em fases posteriores da nossa história. O ser humano sabe se aproveitar do ambiente mesmo sem poder controlá-lo. (KOSELLECK, 2014, p. 84).

Essa condição de seres que possuem consciência do tempo e dos acontecimentos que nele se dão não nos possibilita ir além da mera tomada de consciência, pois toda tentativa de ir além sofre alterações provocadas pela interferência de fatores que não têm como ser antecipados. Tais interferências aumentam à medida que avançamos no tempo. Ao mesmo tempo que aumentamos nossa compreensão dos acontecimentos, também há um aumento na necessidade de mudança da compreensão dos mesmos. Entretanto, essa situação não nos desestimula a perseguir o entendimento dos acontecimentos no tempo. Quando trata da hermenêutica de Gadamer, Koselleck afirma que

[...] assim como a teologia, a jurisprudência, a poesia e suas interpretações, também a história (*Geschichte*) é vista como uma subdivisão da compreensão da existência. O homem, projetado para compreender, não tem outra opção senão conferir sentido à experiência da história, ou, por assim dizer, assimilá-la hermeneuticamente. (KOSELLECK, 2014, p. 92)

Embora nossos esforços não alcancem o resultado desejado, isso não parece ser suficiente para os abandonarmos e vivermos nossas vidas sem perguntar pelo sentido do que acontece. Sem dúvida, parece que estamos condenados a compreender. Contudo, isso não nos permite deduzir que haja um progresso na compreensão dos acontecimentos dados no tempo, já que não há um modelo ideal que sirva de parâmetro de comparação para avaliar-se o grau de proximidade do que acontece com o que é idealizado. Há a necessidade de incrementar-se nossas categorias para que compreendamos de forma mais adequada os acontecimentos. Ao relacionar a hermenêutica às condições de possibilidade de elaboração de uma teoria geral da história Koselleck distingue, remetendo-se a Gadamer, os fatos da linguagem empregada no relato destes. Haverá sempre uma oposição, uma fronteira que marca uma diferença fundamental que torna o objetivo da compreensão do que acontece um objetivo inalcançável, pois “[...] a superioridade daquilo que deve ser compreendido nunca pode ser completamente alcançada por qualquer interpretação. A temática da teoria da história é um caso especial, que toda

linguagem tenta, em vão, esgotar". (KOSELLECK, 2014, p. 106)

As teorias da história devem buscar ir além dos textos e relatos dos fatos a que se remetem, a tarefa do historiador, difere das do teólogo, do jurista e do filólogo que se apoiam na linguagem para desempenhar suas tarefas.

O mesmo não ocorre com o historiador: basicamente, ele se serve dos textos apenas como testemunhas, para extrair uma realidade que existe além deles. Portanto, mais do que todos os outros exegetas de textos, ele destaca um fato extratextual, apesar de reconstituí-lo por meios lingüísticos. Parece até uma ironia. Na comparação com as ciências do espírito, o historiador depende menos de textos do que o jurista, o teólogo ou o filólogo. Ao serem transformados em fontes pelas perguntas que ele formula, os textos possuem apenas um valor indicativo para as histórias que ele deseja conhecer. (KOSELLECK, 2014, p. 107)

Com isso, Koselleck expõe a necessidade de uma constante ampliação e modificação dos elementos empregados pelo esforço de compreensão dos acontecimentos dados no tempo. Uma vez que esse esforço se mostra como expressão da linguagem, e os fatos não se resumem à narrativa que deles fazemos, a atitude do historiador deve ser a de não colocar a linguagem como ponto de chegada. A narrativa dos fatos históricos é meramente um ponto de partida para a compreensão da história, não a própria história. Nossas origens culturais, como afirma Umberto Eco, atribuem a um ato de fala o começo de toda a história.

A nossa história, em comparação com numerosíssimas outras histórias, tem a vantagem de poder começar desde o Início.

Antes de mais nada, quem fala é Deus, que, ao criar o céu e a terra, diz: "Faça-se a luz". E logo a seguir desta palavra divina, "Fez-se a luz" (Gênesis 1, 3-4). A criação aconteceu por um ato de palavra, e somente nomeando as coisas que via, cria Deus sucessivamente, conferindo-lhes um estatuto ontológico: "E Deus chamou a luz 'dia' e as trevas 'noite' [...] (e) declarou o firmamento 'céu'". (ECO, 2001, p. 25)

A concepção estratificada do tempo histórico proposta por Koselleck apresenta-se como alternativa teórica que nos permite ver o emaranhado de inconsistências gerado pelo esforço de interpretação que visa dar sentido ao que acontece como uma totalidade.

Os elementos envolvidos nesta trama adquirem performances que se alteram conforme o enredo se desenvolve, a ponto de o próprio esforço de compreensão ter diferentes interpretações quanto ao seu significado e propósitos. A concepção de Koselleck nos mostra que é justamente a inconsistência a característica que se sobressai quando diferentes interpretações dos acontecimentos históricos são relacionadas. Tais interpretações se apresentam com diferenças suficientes que nos permitem ver seus contornos e limites. A abordagem koselleckiana nos possibilita compreender e relacionar os diferentes resultados a que chegaram os historiadores no percurso da história, mas não resolve as inconsistências propriamente ditas. Elas são consideradas como inerentes ao processo. Não há perspectiva para uma proposta que vise a elaboração de uma teoria unificadora geral capaz de solucionar todas as inconsistências geradas pelo esforço de encontrar sentido no que acontece. Tal abordagem evidencia que, mesmo distantes no tempo, os estratos gerados pelas interações dos elementos históricos, não deixam de produzir seus efeitos posteriormente. Os novos estratos na verdade se devem exatamente àqueles que se sobrepõem. O esforço teórico do historiador apresenta-se como uma rede de inconsistências que se acumulam umas sobre as outras, interligando-se e constituindo um todo que se mantém pela constância das suas forças de oposição e coesão.

Essa arquitetura da inconsistência, em que diferentes contribuições do pensamento se sobrepõem umas às outras, sem que isto represente a substituição da camada mais antiga pela mais nova, é aplicada à reflexão aqui empreendida. Entendemos que a discussão em torno da possibilidade de obtenção de inteligência de modo artificial possui várias camadas de compreensão que se deram ao longo da história do pensamento. Esses estratos de pensamento, independente da época em que surgiram, continuam a produzir efeitos uns sobre os outros, influenciando a reflexão atual. O resultado disso é uma elaboração em que as partes não se articulam de forma a mantê-la erguida. Contudo, não devemos nos deixar seduzir pela busca de uma saída redentora, que remova do nosso horizonte de reflexão todas as inconsistências que levam ao colapso dos nossos esforços e nos apresente, assim, um modelo em que todas as explicações sejam complementares umas às outras, dando um alento à ideia de que é possível a elaboração de uma teoria que seja capaz de unificar todas as perspectivas de tratamento da questão.

A noção de “volta estranha” como superação da dicotomia

Como pretendemos mostrar que os objetivos da versão forte do projeto de IA são, ao menos, admissíveis e que para nos certificarmos disso basta superarmos a dicotomia subjacente às formas de tratamento dessa questão ao longo do tempo, lançaremos mão da noção de *volta estranha* apresentada por Douglas Hofstadter em *Gödel, Escher, Bach: um entrelaçamento de gênios brilhantes* (2001), como podemos ver na imagem de Escher (Figura 1), a fim de aplicá-la às diferentes concepções e definições de *homem, inteligência e artifício*, o que será feito no capítulo 4 da presente tese.

A noção de *volta estranha*, de acordo com o registro de Hofstadter (2001),

[...] não foi inventada por Escher, mas sim por Roger Penrose, matemático britânico, em 1958. Contudo, o tema da volta estranha já se encontrava presente na obra de Escher em 1948, ano em que desenhou *Drawing Hands*. (HOFSTADTER, 2001, p. 16)

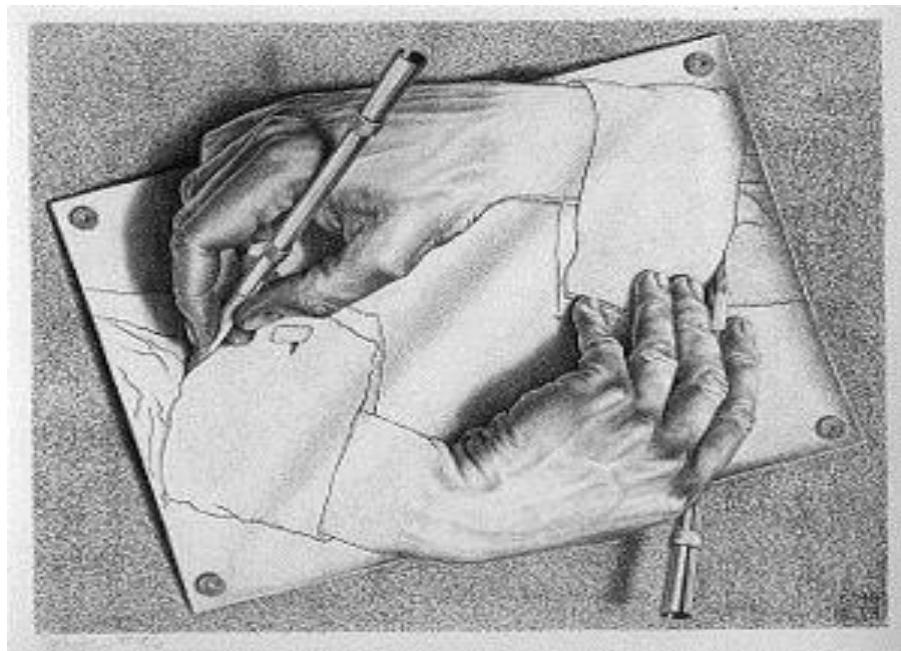


Figura 2: *Drawing Hands*, M. C. Escher, 1948

Na verdade, como podemos ver na exposição de Hofstadter, o emprego da noção

de *volta estranha* é bem mais antigo que a sua invenção por Roger Penrose (1931). O autor deixa isso claro quando optou por colocar no título de sua obra uma menção explícita e direta às obras artísticas de Escher e Bach, e matemática de Gödel, produzidas antes da publicação do trabalho de Penrose.

Penrose se ocupou com o problema de as máquinas serem capazes de ter consciência e sobre as dificuldades de explicação de como essa se dá em *O grande, o pequeno e a mente humana* (1998). Nela ele se refere à noção de *volta estranha* quando aponta os limites da explicação da consciência baseada nos princípios da física atual.

De acordo com o seu pensamento, os procedimentos não computáveis da inteligência humana não podem ser descritos corretamente pela física contemporânea, eles só seriam corretamente descritos por um novo tipo de física que fosse capaz de elaborar novas explicações com maior grau de exatidão ao dispor de um conceitual mais robusto e distinto do atual. Para ele, a mente é uma espécie de ponto de encontro da mecânica clássica com a mecânica quântica.

Na referida obra (1998), Penrose coloca em discussão a relação do mundo físico com o mundo mental. A relação entre as escalas abordadas pela física não gera complementaridade, pelo contrário, as leis que as descrevem são tão distintas ao ponto de serem consideradas incompatíveis, parecem descrever mundos completamente distintos, quando na verdade se tratam apenas de escalas diferentes de um mesmo mundo.

Uma das razões pelas quais optei por dividir minhas descrições das leis físicas entre dois capítulos, a saber, o Grande e o Pequeno, é que as leis que governariam o comportamento em grande escala do mundo e aquelas que o fazem em pequena escala parecem ser muito diferentes (PENROSE, 1998, p. 18)

Penrose possui uma representação tripartite de mundo constituída de três instâncias. É da relação entre esses elementos que se origina a forma como o mundo se apresenta a nós. As três instâncias de mundo relacionadas são: o mundo das leis matemáticas ou mundo platônico, o mundo físico e o mundo mental. Como observado na figura 2

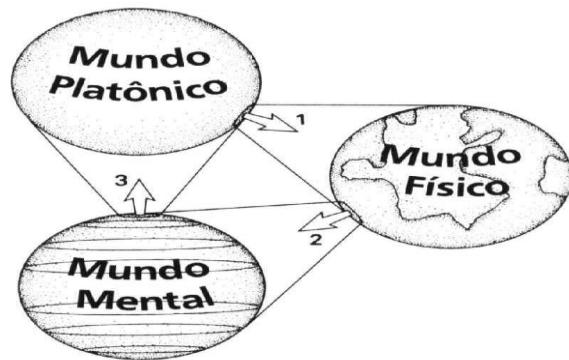


Figura 3: Penrose (1998, p. 109)

Tal posicionamento é considerado inusitado, pois inverte a relação de compreensão entre o mundo e o conhecimento matemático.

Sua posição geral é algo surpreendente, pois ele defende a existência de um mundo matemático platônico que gera, por assim dizer, o mundo físico. Isso se opõe ao que normalmente se crê, quando se mantém que as investigações sobre o universo é que nos induzem a construir determinadas teorias matemáticas (cálculo diferencial e integral, teoria do potencial, análise vetorial, etc.); ou seja, que a matemática nos ajuda a compreender o real, mas que as estruturas matemáticas não são os componentes últimos do universo. (NEWTON DA COSTA, <http://resenhasbrasil.blogspot.com.br/2008/10/o-grande-o-pequeno-e-mente-humana.html>)

As interações entre essas instâncias do mundo são o que ele denomina “mistérios”.

Acho que existe um problema fundamental com a ideia de que a mentalidade nasça da fisicalidade isso é algo com que os filósofos se preocupam, por muito boas razões. As coisas de que falamos na física são matéria, coisas físicas, objetos massivos, partículas, espaço, tempo, energia etc. Como poderiam os nossos sentimentos, a nossa percepção do vermelho ou da felicidade ter algo a ver com a física? Vejo isso como um mistério. (PENROSE, 1998, p. 107)

Para ele ainda se faz necessária a elaboração de uma explicação sobre como os acontecimentos do mundo físico correspondem de forma exata àquilo que as leis da matemática descrevem, assim como a relação entre o mundo físico e o mundo mental.

Para indicar a necessidade de um novo modelo de explicação, Penrose dá como

exemplo a experiência de Libet e Kornhuber que trata da relação entre o nosso comportamento e a atividade cerebral. Essa experiência visava mensurar a diferença entre o início de uma atividade cerebral e a tomada de consciência de uma ação sofrida ou praticada. De acordo com a interpretação dos seus resultados, há um lapso de tempo entre o início da atividade no cérebro e o momento em que as pessoas julgam tomar consciência do objeto dessa atividade.

Precisamos de meio segundo até nos tornarmos conscientes de algum acontecimento. Então, tentamos fazer a consciência funcionar, para fazer alguma coisa com ele. Precisamos de outro segundo para que o livre-arbítrio faça alguma coisa - ou seja, precisamos, no total, de um segundo e meio. Assim, se algo exigir uma resposta conscientemente deliberada, precisamos de cerca de um segundo e meio antes de podermos fazer realmente uso dela. (PENROSE, 1998, p. 147)

No seu entendimento, isso não faz sentido, pois necessitariam de um tempo demasiado longo para deflagrarmos uma ação consciente. Ele não confia que levemos tanto tempo para que uma resposta seja dada.

Ele desconfia da medição de Libet e Kornhuber ao cogitar a hipótese de que a interpretação da experiência realizada por eles possui um erro. Este erro estaria na escala de interpretação das atividades envolvidas na tomada de consciência que empregaria um modelo físico inadequado para esse fim, neste caso, o da física clássica. Como para ele a física clássica não é capaz de descrever os fenômenos que se encontram no nível das subpartículas, se faz necessária uma *nova física*, que se anuncia através da mecânica quântica. Libet e Kornhuber empregam um modelo de compreensão que não se ajusta de modo adequado ao fenômeno que pretende descrever. É essencial, portanto, o surgimento de uma nova física.

Ele não chega a afirmar categoricamente que a mecânica quântica seja a expressão exata de uma nova física, mas ela é, ao menos, um pensamento que impulsiona o surgimento desse novo campo de conhecimento.

Minha maneira de encarar isso é que pode muito bem haver algo na maneira como interpretamos essas experiências que faça alguma suposição de que a física que estamos usando seja basicamente a física

clássica. Lembremo-nos do problema do teste de bombas, quando falamos acerca de contrafactuals e do fato de que eventos contrafactuals poderiam ter uma influência sobre as coisas, ainda que não ocorram realmente. O tipo comum de lógica que se usa tende a levar ao erro se não se toma cuidado. Temos de ter em mente como se comportam os sistemas quânticos, e assim pode ser que algo estranho esteja acontecendo nessas contagens de tempo, por causa da não-localidade quântica e dos contrafactuals quânticos. E muito difícil entender a não-localidade quântica dentro do quadro da relatividade restrita. Minha interpretação é que, para entender a não-localidade quântica, vamos precisar de uma teoria radicalmente nova. Essa nova teoria não será apenas uma leve modificação da mecânica quântica, mas sim algo tão diferente da mecânica quântica-padrão quanto a relatividade geral é diferente da gravidade newtoniana. Teria de ser algo com um quadro conceitual completamente diferente. Nessa interpretação, a não-localidade quântica estaria incorporada à teoria. (PENROSE, 1998, p. 147)

A noção de não-localidade quântica é fundamental quando se tem por propósito a realização de uma medição mais precisa desses fenômenos. Essa noção é que nos permitirá encontrar sentido para aquilo que, de acordo com o conhecimento atual, faz sentido algum.

A compreensão da consciência, do pensamento e de outras atividades de nível superior depende da compreensão de fenômenos, supostos responsáveis pelo surgimento dessas capacidades, que ocorrem numa escala muito breve do tempo, “essa escala de tempo é de cerca de 10^{-43} de segundo e com frequência é chamada escala de tempo de Planck, ou ‘crônon’”. (PENROSE, 1998, p. 21)

Os fenômenos que nos permitiriam compreender a consciência e o agir inteligente estariam próximos dos limites dessa escala. Para se ter uma ideia, “esta escala de tempo é muito mais breve do que qualquer coisa experimentada na física de partículas. Por exemplo, as partículas de vida mais curta, as chamadas ressonâncias, duram cerca de 10^{-23} de segundo”. (PENROSE, 1998, p. 21).

Aqui ele emprega a noção de *volta estranha* para mostrar a necessidade de elaboração de um novo aparato conceitual que nos dê a possibilidade de compreender os fenômenos que fogem, atualmente, da nossa compreensão. A *volta estranha* é apresentada na forma de uma imagem que ele denominou de “triângulo de impossibilidades”, o qual não parece estar corretamente desenhado ao ser visto na sua totalidade. Porém, quando

ele é decomposto para ser analisado em parte encontra o erro, nenhuma delas apresenta ter algum problema. É necessário que vejamos a figura de uma nova maneira se quisermos sair do paradoxo criado pela comparação da visão geral com a visão parcial do objeto.

Penrose não está convencido de que chegaremos a compreender a relação entre o nosso pensamento e o mundo, mas, também, não nos diz que devamos encarar com resignação nossa ignorância acerca dessa relação. No entanto, não deixa de considerar que “[...]é importante reconhecer quando ocorrem quebra-cabeças e mistério”. Mas o mero fato de estar acontecendo algo muito intrigante não significa que algum dia seremos capazes de entender-lo. Com esse misto de arrojo e ceticismo é que ele afirma a necessidade de conhecermos melhor acontecimentos que se apresentam a nós de forma inusitada, pois, talvez esteja na compreensão destes a explicação do surgimento da consciência, da inteligência e do pensamento.

Newton da Costa (1929) afirma que Hofstadter tratou o tema das *voltas estranhas* de maneira semelhante à de Penrose, pois ele “recorreu a analogias (nem sempre óbvias) entre as realizações de Gödel, Escher e Bach para tentar esclarecer aspectos da atividade cognitiva, em particular do significado da inteligência artificial.” <http://www1.folha.uol.com.br/fsp/resenha/rs1108200103.htm>.

Hofstadter relaciona as intuições de paradoxo (Gödel), de escada (Escher) e de escala musical (Bach) para descrever a noção de *volta estranha* (*strange loop*). Nos deteremos na descrição que ele faz dessa noção presente na obra de Maurits Cornelis Escher. Fizemos tal opção por entendermos que é correto o entendimento de Richard Rorty (*A filosofia e o espelho da natureza*, 1994) de que as imagens proporcionam uma compreensão mais imediata e acessível dos problemas, pois “são as imagens mais do que as proposições, as metáforas mais que as afirmações que determinam a maior parte das nossas convicções filosóficas”. (RORTY, 1994, p. 27)

Hofstadter apresenta uma explicação de como foram possíveis as criações desses gênios que superaram os limites impostos pela maneira de pensar de suas épocas. No entanto, seu interesse principal não é explicar o contexto de criação dos gênios de Gödel, Escher e Bach, sua pretensão é a de abordar o problema da *auto-referência*, movimento de pensamento que possibilita um *ver-se a si mesmo*, um problema que ocupa a filosofia

desde o seu início. Como sair do emaranhado que nos apresenta o paradoxo do mentiroso? A máquina de Turing pode ser comparada, quanto ao seu funcionamento, com o cérebro humano? Nosso cérebro é capaz de compreender completamente o funcionamento de si próprio? Somos capazes de superar os limites demarcados pelo teorema de Gödel de que a lógica a qual estamos submetidos não nos permite comprovar a veracidade de todas as nossas afirmações feitas a partir desse sistema formal?

Hofstadter pressupõe que a consciência é uma emergência do sistema mente/cérebro. Quando se referir a um sistema hierarquizado de regras, ele não deixará de emprega-la fazendo referência ao funcionamento do cérebro humano.

A noção de *volta estranha* é definida como um movimento que revela algo de inusitado e extraordinário.

O fenômeno das *voltas estranhas* ocorre sempre que, quando nos movemos para cima (ou para baixo) através dos níveis de um sistema hierárquico, encontramo-nos, inesperadamente, de volta ao lugar de onde partimos. (HOFSTADTER, 2000, p. 11)

Existem diferentes tipos de *voltas estranhas*, umas sendo mais apertadas do que outras, isso é determinado pelo número de etapas ou dos níveis que um determinado sistema hierárquico possui, podendo a mesma completar o seu percurso cíclico em um só passo ou etapa. A realização desse movimento só é possível se houver uma ruptura com a forma dicotômica de arranjo de ideias. Isso pode ser constatado quando da análise de algumas obras do artista holandês em que diferentes perspectivas da realidade e da imaginação são interpostas para compor uma imagem que é transformada, em várias etapas, até retornar, num salto inusitado, a sua forma inicial.

Escher realizou voltas estranhas de muitas maneiras diferentes e elas podem ser arranjadas segundo a rigidez da volta. A litografia *Ascending and descending (Subindo e descendo)* (figura 6), em que os monges caminham eternamente em volta, e a versão mais aberta, uma vez que tantos são os passos envolvidos antes que o ponto de partida seja de novo alcançado. *Waterfall* apresenta uma volta mais apertada, que, como já observamos, envolve apenas seis passos definidos. Pode-se pensar que haja aqui certa ambiguidade na noção de “passo” - não se poderia dizer, por exemplo, e com igual razão, que *Ascending and*

descending contem quatro níveis (escadas) ou quarenta e cinco níveis (degraus)? E bem certo que existe uma indefinição inerente na contagem dos níveis, não só nos desenhos de Escher, mas também em sistemas hierárquicos de níveis múltiplos. Posteriormente, aprimoraremos nosso entendimento a respeito dessa indefinição. Mas não nos distraímos demais por agora! A medida que apertamos nossas voltas, chegamos ao notável *Drawing hands* (*Mãos que desenham*) (figura 135), no qual cada uma das mãos desenha a outra: uma volta estranha de dois passos. E, finalmente, a mais apertada de todas as voltas estranhas e realizada em *Print gallery* (*Galeria de gravuras*) (figura 142): um quadro de um quadro que contém a si próprio. (HOFSTADTER, 2000, p. 12)

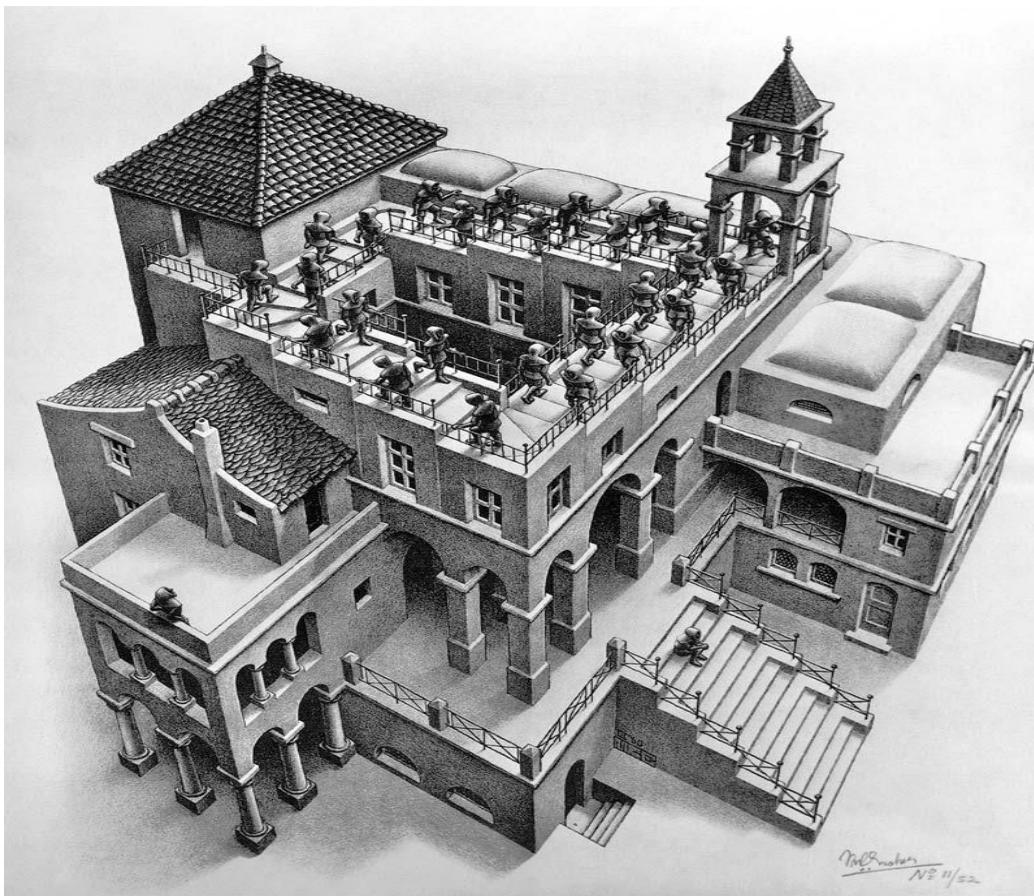


Figura 4: *Ascending and descending*, M. C. Escher, 1960

As *voltas estranhas* surgem de um antagonismo entre finito e infinito, entre real e imaginário, da composição de elementos que, a princípio, não poderiam ser justapostos na elaboração de uma teoria, melodia ou desenho. Hofstadter lança mão de exemplos nas

artes, na matemática e na lógica para abordar uma questão que está estreitamente ligada com os objetivos do projeto de IA, visto que este se fundamenta numa ideia bastante antiga que já encontrava estabelecida entre os gregos, qual seja, a de compreender o pensamento humano como o desenvolvimento de uma atividade como qualquer outra, estando submetida ao uso e seguimento de regras e axiomas.

No entanto, o pensamento humano parece não se submeter a princípios e limites de ordenamento que o impedem de dar prosseguimento a uma determinada linha de raciocínio ou pensamento. Não só a capacidade de raciocinar nos distingue das demais coisas, a própria atividade de pensamento possui algo de peculiar ao se mostrar capaz de transgredir as leis e princípios que regem o seu funcionamento. Isso nos leva a considerar a hipótese de que o pensamento humano pode não ser passível de o compreendermos como um processo mecânico.

Nossa capacidade de raciocinar tem sido considerada, muitas vezes, como aquilo que nos distingue das demais espécies; não deixa assim de parecer paradoxal, à primeira vista, tratar de mecanizar aquilo que é mais humano. Contudo, mesmo os gregos antigos sabiam que o raciocínio é um processo que segue padrões e, pelo menos em parte, comandado por leis enunciáveis. (HOFSTADTER, 2000, p. 21)

Ao mesmo tempo que nossa capacidade de produzir raciocínios e pensamentos segue regras, também é capaz de transgredi-las, o que se apresenta como um paradoxo, pois ela não parece seguir padrões quando realiza um movimento como o de uma *volta estranha*. Para que esse movimento heterodoxo se realize de forma completa, para que seja completada a volta em toda a sua circularidade, é necessário que sejam colocados em suspensão os princípios e regras do seu funcionamento.

As histórias da matemática e da lógica nos revelam sempre que determinados princípios e definições são estabelecidos algo acontece no sentido de mostrar que eles não se aplicam a determinados casos, sempre existem exceções. Conforme Hofstadter, uma das causas para que isso ocorra é a existência de uma assimetria entre a nossa intuição e os sistemas hierarquizados de regras que elaboramos com o objetivo de compreender o funcionamento das nossas próprias faculdades.

O mais famoso é o paradoxo de Russell. Pareceria que, na maioria dos casos, os conjuntos não são membros deles próprios - por exemplo, o conjunto das focas não é uma foca, o conjunto que contém apenas Joana D'Arc não é Joana D'Arc (um conjunto não é uma pessoa) - e assim por diante. Sob esse aspecto, a maioria dos conjuntos é "relativamente comum". Contudo, alguns conjuntos "auto-devoradores" *realmente* incluem-se a si próprios entre seus membros, como o conjunto de todos os conjuntos, ou o conjunto de todas as coisas exceto Joana D'Arc, e assim por diante. Claramente, todo conjunto ou é relativamente comum ou é auto-devorador e nenhum conjunto pode ser ambas as coisas. Ora, nada nos impede de inventar R: *o conjunto de todos os conjuntos relativamente comuns*. À primeira vista, R pareceria ser uma invenção relativamente comum - mas essa opinião deve ser revista quando nos perguntamos: "O próprio R é um conjunto relativamente comum ou auto-devorador?" Verificaremos que a resposta é: "R não é nem relativamente comum, nem auto-devorador, pois ambas as escolhas levam a um paradoxo". Tente! (HOFSTADTER, 2000, p. 21)

De acordo com Hofstadter, os paradoxos como o acima mencionado são originados pela *auto-referência*, fenômeno semântico que produz como resultado *voltas estranhas*.

Embora autores como Stephen Yablo (*A paradox without self-reference*, Analysis, 53,4, 1993) afirmem que os paradoxos gerados pela *auto-referência* sejam apenas um dos tipos de paradoxo existentes, a compreensão no meio matemático e lógico é que ela, a *auto-referência*, é a principal causadora desse tipo de movimento do pensamento. Por essa razão, a forma de superar esse tipo de anomalia, que impede o pensamento de seguir adiante, seria através da sua exclusão do contexto cujo objetivo é o estabelecimento de uma simetria entre nossas intuições e os sistemas de axiomas hierarquizados.

Tal tarefa não envolveria grandes dificuldades se a *auto-referência* fosse um simples procedimento de aplicar o sentido de uma proposição a ela mesma. Mas não é o que nos mostra o exemplo, dado por Hofstadter, da formulação análoga do *Paradoxo de Epimênides*.

A sentença seguinte é falsa.

A sentença anterior é verdadeira.

A diferença com relação ao exemplo original do *Paradoxo de Epimênides* (*Eu estou mentindo*) é que o paradoxo acima é produzido em dois passos, enquanto a formulação original do paradoxo do mentiroso é produzida em um passo somente. Isso ocorre porque, na formulação análoga, o paradoxo não é gerado por um procedimento de auto-referência *imediata* das proposições que o constituem, mas por um procedimento de auto-referência *mediado* pelo estabelecimento de uma relação lógica entre elas,

“a ‘culpa’ dessa volta estranha não pode ser atribuída a nenhuma das sentenças - apenas à maneira como elas ‘apontam’ uma à outra”. Do mesmo modo, cada uma das regiões locais de *Ascending and descending*, de Escher, é bastante legítima, e apenas a maneira pela qual elas são globalmente articuladas que cria uma impossibilidade. (HOFSTATER, 2000, p. 22).

Segundo ele, uma das tentativas de nos livrar dos paradoxos gerados pelo emprego das regras de um determinado sistema de axiomas hierarquizados foi a realizada por Bertrand Russell (1872-1970) e Alfred Whitehead (1861-1947) com a criação da *teoria dos tipos*. Hofstadter afirma que a teoria dos tipos consiste no banimento da *auto-referência*, tal ação se concretiza por meio da elaboração de um sistema de axiomas hierarquizados concebidos de tal forma que as proposições que o constituem devem ser todas do tipo *autoexcludente*, isso é, proposições cujo sentido não podem fazer referência a elas mesmas.

Segundo todas as aparências, portanto, esta *teoria de tipos*, que poderíamos também denominar “teoria da abolição das voltas estranhas”, livra, com êxito, a teoria dos conjuntos de seus paradoxos, mas apenas ao custo de introduzir uma hierarquia de aparência artificial e de desautorizar a formação de certas espécies de conjuntos - tais como o conjunto de todos os conjuntos relativamente comuns. Intuitivamente, essa não é a maneira segundo a qual imaginamos os conjuntos. (HOFSTADTER, 2000, p. 23)

A tarefa de Russell e Whitehead alcançou um êxito restrito porque ela se remeteu exclusivamente aos paradoxos constituídos pelos axiomas da linguagem da teoria dos conjuntos. A *teoria dos tipos* não teve o mesmo sucesso quando aplicada às linguagens

naturais, pois nessas, suas sentenças não estão impedidas de fazer a *auto-referência*, já que elas não guardam entre si uma hierarquia tão rígida quanto a da linguagem da matemática.

A teoria de tipos deu conta do paradoxo de Russell, mas nada conseguiu com relação aos paradoxos de Epimênides e de Grelling. Para as pessoas cujo interesse não ia além da teoria dos conjuntos, isso era suficiente, mas para os que se interessavam pela eliminação dos paradoxos em geral, parecia necessária uma “hierarquização” similar para impedir esse tipo de volta para trás na linguagem. Na base de tal hierarquia estaria uma *linguagem-objeto*. A referência aqui só pode ser feita a um domínio específico - não a aspectos da própria linguagem-objeto (tais como suas regras gramaticais, ou sentenças específicas nessa linguagem). Para esse propósito haveria uma *metalinguagem*. (HOFSTADTER, 2000, p. 23)

Ou seja, os problemas de um nível de linguagem seriam resolvidos por uma linguagem de nível superior *ad infinitum*. Com isso seria criada uma cadeia de resolução de problemas que envolveria a realização de saltos de um nível para o outro, de uma linguagem de primeira ordem para uma metalinguagem. Com a linguagem ordinária não ocorre o mesmo, pois ela parece sanar a necessidade de uma metalinguagem ao não impedir que o fluxo de pensamento siga em curso. Os saltos, nas linguagens naturais, parecem se dar em níveis dentro do próprio sistema. Há uma dificuldade na definição do que seja a *consciência*, pois esta não consiste simplesmente na realização de saltos de um sistema de axiomas para outro. Não é isto que parece ocorrer com os usos que fazemos das linguagens naturais.

Haveria, ainda, uma outra dificuldade. A inteligência humana parece escapar aos esforços que ela mesma realiza para compreender a si própria. Uma das razões para isso estaria nas próprias características do cérebro humano.

[...] pode ser que, ao contrário dos motores de automóveis, nossos cérebros sejam sistemas teimosos e refratários, que não possam ser decompostos com precisão de nenhuma maneira. Atualmente, não temos ideia se nossos cérebros cederão as repelidas tentativas de dividi-los em camadas claras que possam ser explicadas em termos de camadas mais baixas, ou se, ao contrário, eles farão fracassar todas as nossas tentativas nesse sentido (HOFSTADTER, 2000, p. 777)

Além dos supostos limites do cérebro humano que representariam sérios obstáculos na compreensão de como somos capazes de realizar saltos inferenciais que nos permitem sair de uma situação paradoxal, Hofstadter salienta que os saltos inferenciais podem resultar da realização de movimentos mais complexos que o de subida de um nível inferior para um nível superior dentro de um sistema hierarquizado de regras. Para ele, nossa inteligência realiza movimentos complexos de idas e vindas entre os diferentes níveis de regras que o constituem. Esse tipo de recursividade produz as *voltas estranhas*. Quando tal movimento é executado os níveis inferiores não são totalmente passivos na sua transitividade com os demais, eles também exercem influência e produzem efeitos nas regras pertencentes aos níveis que lhes são superiores.

Acredito que as explicações de fenômenos “emergentes” em nossos cérebros - por exemplo, ideias, esperanças, imagens, analogias e, finalmente, a consciência e o livre-arbítrio - baseiam-se em uma espécie de volta estranha, uma interação entre níveis na qual o nível mais alto se volta em direção ao nível mais baixo e o influencia, ao mesmo tempo em que ele próprio é determinado pelo nível mais baixo. (HOFSTADTER, 2000, p. 779)

Hofstadter acredita que a explicação da *consciência* exige que sejamos capazes de descrever com maior precisão os movimentos complexos realizados por nossa inteligência na superação de obstáculos que, à primeira vista, se apresentam como intransponíveis. A compreensão dessas transposições só poderá ser alcançada quando o uso dos axiomas do sistema hierarquizado de regras que rege o funcionamento da nossa inteligência for melhor investigado. Para tanto, será necessário que a fluidez do seu movimento seja captada, o que exige uma flexibilidade na sua abordagem.

Para lidar com toda a riqueza do sistema mente/cérebro, temos de ser capazes de deslizar confortavelmente de um nível para outro. Além disso, temos de admitir vários tipos de causalidade”: maneiras pelas quais um evento em um nível de descrição pode “causar” a ocorrência de eventos em outros níveis. (HOFSTADTER, 2000, p. 779)

Embora Hofstadter considere necessário compreendermos o fluxo de emprego das regras do sistema de axiomas, se quisermos definir em que consiste a *consciência*, ele não deixa de salientar que também devemos levar em consideração o fato de que na base de todo esse sistema e da complexidade de seu funcionamento estão as leis da física.

Existe um famoso hiato entre duas linguagens discursivas: a linguagem subjetiva e a linguagem objetiva. Por exemplo, a sensação “subjetiva” da cor vermelha e o comprimento de onda “objetivo” da luz vermelha. Para muitos, elas parecem ser absolutamente inconciliáveis. Não penso assim. Como não creio inconciliáveis as duas visões das *Mãos que desenham*, de Escher - a do “interior do sistema”, em que as mãos desenham uma a outra, e a do exterior, em que Escher desenha tudo. A sensação subjetiva do vermelho provém do vórtex de autopercepção no cérebro; o comprimento de onda objetivo e a maneira como as coisas são vistas quando você salta fora do sistema. Embora nunca ninguém possa saltar fora do sistema o suficiente para perceber a totalidade, não nos devemos esquecer de que ela existe. Devemos recordar-nos de que as leis físicas são o que faz tudo acontecer - bem lá embaixo, nos cantos e gretas neurais, demasiado remotos para que possamos atingi-los com nossos meios introspectivos de nível alto. (HOFSTADTER, 2000, p. 779)

Tal posição leva ao debate de se as leis da biologia seriam subalternas às leis da física. Sem entrar no mérito dessa discussão, entendemos que a relevância dada por Hofstadter às leis da física nos adverte que assim como há um nível de interação entre os axiomas que constituem um determinado sistema hierarquizado, também há uma interação entre esses axiomas e o ambiente no qual eles se dão.

As *voltas estranhas* são realizações do pensamento que ao serem consideradas de um determinado ponto de vista não parecem fazer sentido, contudo, quando procuramos as razões disso, não as encontramos.

A fim de respondermos à nossa hipótese de trabalho que pergunta pela admissibilidade do projeto de IA, construiremos, no próximo capítulo, uma *volta estranha* envolvendo diferentes significados dos conceitos de *Homem, Inteligência e Artifício*. Para isso, não estabeleceremos uma relação de antagonismo dicotômico entre esses diferentes significados, pois entendemos que eles constituem um histórico de mudanças em

decorrência do longo tempo de uso desses conceitos. Nossa *volta estranha* terá como ponto de partida e de chegada o conceito de *Homem*.

CAPÍTULO 4

UM TRIÂNGULO DE IMPOSSIBILIDADE

HOMEM, INTELIGÊNCIA E ARTIFÍCIO

Neste capítulo abordaremos algumas das diferentes definições dos conceitos de *Homem, Inteligência e Artifício*, importantes na formulação da resposta à nossa hipótese de trabalho.

Como foi dito no capítulo anterior, pretendemos construir um arranjo desses conceitos evitando cair na armadilha do estabelecimento de uma relação dicotômica entre suas diferentes definições. A fim de responder à hipótese de trabalho dessa tese, que pergunta sobre a admissibilidade do projeto de IA, construiremos uma *volta estranha* envolvendo os três conceitos acima mencionados.

Apresentaremos a resposta ao problema colocado pela hipótese de trabalho em duas etapas. A primeira consiste na apresentação de diferentes definições que os conceitos de *Homem, Inteligência e Artifício* receberam ao longo de suas histórias de uso. Faremos esta parte da resposta inspirados na perspectiva de tratamento dos conceitos concebida por Arthur Oncken Lovejoy (1873-1962), por ele denominada de *história das ideias*. Nessa concepção de trabalho filosófico, os conceitos são considerados a partir das mudanças que eles enfrentam ao longo do seu tempo de uso, mudanças essas envolvidas numa variada gama de situações. A segunda etapa compreende a construção de uma *volta estranha* envolvendo os conceitos de *Homem, Inteligência e Artifício*. Essa *volta estranha* será a premissa forte do embasamento da conclusão final deste trabalho.

Lovejoy entende que a perspectiva de abordagem dos conceitos da *história das ideias*, que se dá a partir da observação das mudanças nos seus significados, pode ser compreendida como uma *grande cadeia*, uma teia de significados que uma vez entrelaçados nos permite compreender tais alterações como elementos que constituem uma totalidade.

Em *A grande cadeia do ser* (2001), o tema central não é o problema do *ser* e suas definições, mas o conceito de *grande cadeia*. O autor não elege como objetivo principal dessa obra o tratamento específico do problema do *ser*, mas sim das mudanças ocorridas com o seu significado ao longo de sua história de uso nos mais diferentes contextos.

Lovejoy abordou a temática do *ser* para evidenciar e destacar o conceito de *grande cadeia*, que se desvenda quando levamos em consideração as mudanças observadas no significado de um determinado conceito ao longo do tempo.

Essas palestras, tentam exemplificar, em pequena medida, um tipo de investigação histórico-filosófica da qual eu esbocei apenas os seus objetivos gerais e método. Devemos primeiro especificar, não realmente, uma ideia simples e singular, mas três ideias que têm tido, durante grande parte da história do ocidente, uma associação bastante estreita e constante em que muitas vezes operaram como se fossem uma só, e, quando se juntaram, produziram uma concepção – uma das principais concepções do pensamento ocidental - que pode ser expressada num único termo: *a grande cadeia do ser*, onde devemos observar o funcionamento de ambas, tanto separadamente quanto em conjunto. (LOVEJOY, 2001, p. 20, tradução nossa)

De acordo com Lovejoy, a concepção de trabalho da *história das ideias* é mais ampla do que a da história dos conceitos filosóficos, nem se restringe ao pensamento de um filósofo. A *história das ideias* possui um escopo bastante mais amplo que o abrangido por um determinado campo conceitual. Ela leva em consideração todos os contextos de uso que uma determinada ideia ou conceito *foi, pode ou poderá* ser empregada.

Isso pode nos levar a pensar, erroneamente, que a *história das ideias* seja um projeto de trabalho que gira em torno da linguagem, que represente uma teoria do funcionamento e desenvolvimento desta. Entendemos que a *história das ideias* é um método de trabalho especulativo que tem por objetivo último produzir uma reflexão sobre o homem a partir do *agir significante* desse, cheio de nuances e alterações que só se evidenciam se tomarmos a perspectiva de visada propiciada pela lente da *grande cadeia*.

Embora a história das ideias seja uma história de tentativa e erro, os erros acabam por iluminar a natureza peculiar, o desejo, a dedicação e as limitações da criatura que neles cai, bem como a lógica dos problemas refletidos que lhes subjaz; eles podem ainda servir para nos lembrar que o modo dominante de pensamento de nossa própria época, que alguns consideram como claras, coerentes, solidamente fundamentadas e finalizadas, provavelmente não parecerão merecer, aos olhos da posteridade, tais atributos.(LOVEJOY, 2001, p. 23, tradução nossa)

A *história das ideias* possui um forte caráter crítico, pois ela não se apresenta como defensora de um determinado postulado ou conjunto de ideias. Ela faz um esforço de reflexão que tem por objeto a própria atividade reflexiva do homem ao longo do tempo. Por essa razão destaca que essa atividade possui uma regularidade, a mudança.

É uma das ironias instrutivas da história das ideias que um princípio, introduzido por uma geração à serviço de uma determinada tendência ou disposição filosófica, muitas vezes prova conter em si, de forma insuspeita, o germe de uma tendência contrária – o de ser, em virtude das suas implicações latentes, o destruidor daquele *Zeitgeist* ao qual estava destinado a ministrar. (LOVEJOY, 2001, p. 288), (tradução nossa)

Entender os mecanismos que subjazem às mudanças de significado é o objetivo mais relevante do método da *história das ideias*. Tal método objetiva mais do que a simples constatação de alterações verificadas num determinado conjunto de ideias que constituem a base das nossas crenças e convicções. Ele é concebido para propiciar a compreensão das razões de fundo que provocam tais alterações e mudanças. No entanto, não podemos confundi-lo como consistindo num método estritamente de análise da linguagem. De acordo com tal concepção, a linguagem é uma fornecedora dos indícios que nos levarão àquilo que realmente é o responsável pelas alterações verificadas nos significados dos termos de linguagem.

Temos muito mais razões para acreditar que o método da *história das ideias* se aproxima mais do interesse de uma investigação de naturezas psicológica e sociológica do homem, que têm como ponto de partida as mudanças verificadas nos usos que este faz da linguagem, do que de qualquer outro campo de investigação.

Finalmente, faz parte da tarefa final da história das ideias a aplicação do seu próprio e distintivo método analítico na tentativa de compreender como nossas crenças e preferências intelectuais são introduzidas e difundidas, a fim de auxiliar a elucidar o caráter psicológico dos processos pelos quais se deram as mudanças em voga e a influência que as ideias tiveram sobre elas; esclarecer, se possível, como as concepções dominantes, ou amplamente influentes, perdem, em uma geração, o controle sobre as mentes dos homens cedendo seu lugar a outras. (LOVEJOY, 2001, p. 20), (tradução nossa)

Em *Reflections on human nature* (1961), Lovejoy expõe em que consiste a tarefa de quem se propõe a empregar o método da *história das ideias*. Nesse livro, dentre outras coisas, ele afirma que a investigação sobre a natureza humana deve ser considerada uma das principais tarefas de quem emprega o método da *história das ideias*. Para ele, quem se ocupa em refletir sobre as teorias acerca da natureza humana, sua gênese, seu desenvolvimento e modo como se dão os processos de escolha humanos, deverá buscar subsídios em toda manifestação humana que diga respeito ao significado desse conceito e suas mudanças.

Tais teorias, implícitas ou explícitas, podem ser discernidas se forem procuradas numa parte considerável da literatura, especialmente no drama e no romance, nos grandes poemas didáticos e na poesia lírica; e, muitas vezes, explicitamente nos escritos de teólogos, filósofos, sátiros e teóricos políticos. Existem coisas mais importantes a saber sobre um escritor do que a sua visão expressa ou tácita, como os seus principais motivos e pressuposições concernentes a natureza humana ou, o conhecimento sobre o período em que as ideias sobre esses foram predominantes. A história da teoria da natureza humana - das ideias dos homens sobre o homem – portanto, é ou deveria ser um dos principais campos de investigação para quem estuda a história das ideias. (LOVEJOY, 1961, p. 12), (tradução nossa)

Buscar encontrar nas mais variadas formas de manifestação do pensamento humano o fluxo em que as mudanças ocorrem implica em realizar expedições exploratórias em todos os territórios em que elas acontecem. Não há uma área de significação que seja privilegiada em relação às demais.

O mesmo tipo de atitude vemos no pensamento de John L. Austin apresentado nos *Philosophical Papers* (1961), uma proposta de trabalho que se assemelha muito a de Lovejoy. Na referida obra, Austin enfatiza a importância de se levar em consideração os diferentes significados de um termo e seus *contextos de uso*, se nos propomos a dizer alguma coisa sobre o significado das palavras e dos demais elementos da linguagem, uma proposta de trabalho filosófico que ele chamou de *fenomenologia linguística*.

Quando examinamos o que nós devemos dizer, quando e quais palavras devemos empregar e em que situações, *olhamos não somente para as palavras (ou significados, seja lá o que isso signifique), mas também,*

para as realidades nas quais empregamos as palavras para falar sobre: empregamos uma consciência aguda das palavras para afiar nossa percepção, embora não como árbitro final, dos fenômenos. (AUSTIN, 1961, p. 130, tradução e itálico nossos)

Pensadores como Austin e Lovejoy compreendem que as diferenças de significados dos conceitos ao longo de suas histórias de uso nos permitem pensá-las como um movimento contínuo, não como uma interrupção do mesmo e surgimento de um novo. Assim, as mudanças ocorridas nos significados passam a ser consideradas como acontecimentos característicos e intrínsecos aos campos semânticos dos termos de linguagem.

O conceito de ‘homem’

Definir o que seja o *homem* é um objetivo que a Filosofia busca desde o seu princípio. Mas a busca em nos definirmos como tais não é uma tarefa que somente a Filosofia se propôs a fazer. Praticamente todas as esferas que envolvem a existência humana procuram responder, a seu modo, a pergunta pelo que somos. Essa seria uma das razões que fazem com que a tarefa de definir o *homem* envolva um elevado grau de dificuldade na sua realização.

O esforço de definir em que consiste o *homem* encontra-se envolvido num paradoxo. Ao mesmo tempo em que proliferam tentativas de definição do *homem*, também aumentam, na exata proporção dessa, as dificuldades em o definirmos. Por ser um objeto que desperta o interesse de investigação em muitas áreas de interesse, poderíamos pensar que essa realidade representaria a obtenção de melhores condições para se alcançar a definição do *homem*, que haveriam menos dificuldades dada a grande soma de esforços. No entanto, o que se percebe é um terreno onde existem muitas concepções e poucos acordos.

Ernst Cassirer an obra *An essay on man. An introduction to a philosophy of human culture* de 1944, afirma que o autoconhecimento constitui a principal tarefa da Filosofia. A condição que deve ser cumprida com vistas ao conhecimento do mundo seria o

conhecimento aprofundado do *ente* que se dispõe a tal propósito. Uma prova da preponderância do autoconhecimento sobre todos os demais objetivos da filosofia seria o ceticismo que, na opinião de Cassirer, não passaria de uma estratégia empregada com o propósito de direcionar toda a investigação para o próprio homem.

Parece ser geralmente admitido que o autoconhecimento é o principal objetivo da investigação filosófica. Em todos os conflitos entre as diferentes escolas filosóficas esse objetivo permaneceu invariável e permanente: provou ser o ponto arquimédico, o centro fixo e imóvel, de todo o pensamento. Nem os mais céticos pensadores negaram a possibilidade e necessidade do autoconhecimento. Eles destruíram todos os princípios gerais concernentes à natureza das coisas, no entanto, tal destruição significa somente a abertura de um novo e mais seguro modo de investigação. Na história da filosofia, o ceticismo é frequentemente considerado como a contrapartida de um *humanismo* bem estabelecido. Com a negação e destruição da certeza objetiva do mundo externo, o cético espera que todos os pensamentos do **homem** se voltem para o seu próprio ser. O autoconhecimento – dizem eles – é o primeiro pré-requisito da autorrealização. (CASSIRER, 1944, p. 15, tradução e negrito nossos)

Se é possível estabelecer um acordo mínimo em torno da tese de que a busca pelo autoconhecimento pode ser considerada o principal objetivo da filosofia, o mesmo não ocorre quanto ao modo como devemos proceder na realização da mesma. Ao longo da história da filosofia essa busca foi orientada de muitas e diversas maneiras. Suas variâncias vão desde a necessidade de conhecimento do mundo externo para daí se depreender a posição do homem numa hierarquia cósmica, até a desconsideração total desta por meio de uma concentração absoluta em nós mesmos, como nos mostrou Ernst Cassirer quando se referiu ao ceticismo.

Max Scheler (1874-1928) em *A situação do homem no cosmos* (2008), afirma que a cultura europeia produziu três ideias incompatíveis de *homem*, ideias essas que tiveram origens e compreensões distintas deste, o que serve como evidencia a favor da afirmação de que o propósito de definir o *homem* não é uma tarefa fácil.

Se a um europeu culto se perguntar que entende ele pela palavra ‘homem’, quase sempre começam, na sua cabeça, a defrontar-se *ter* sâmbitos de ideias de todo inconciliáveis. Primeiro, o universo

intelectual da tradição judeo-cristã de Adão e Eva, da criação, do paraíso e da queda. Em segundo lugar, o círculo de ideias da Grécia antiga em que, pela primeira vez no mundo, a autoconsciência do homem se elevou a um conceito da sua situação particular, mediante a tese de que o homem é homem pela posse da “razão”, *logos, phronesis, ratio, mens*, etc. – *Logos* significa aqui tanto o discurso como a aptidão para apreender a ‘quididade’ de todas as coisas. Estreitamente unida a esta ideia está a doutrina segundo a qual existe também, subjacente ao todo integral, uma “razão” sobre-humana, da qual o homem, e só ele entre todos os seres, participa. O terceiro círculo intelectual, também ele já há muito transformado em tradição, é o da ciência moderna da natureza e da psicologia genética: o homem seria o resultado final, muito tardio, da evolução do planeta Terra – um ser que se distingue das formas animais que o precederam só pelo grau de complicação das combinações de energias e de capacidades que, em si, se encontram já na natureza infra-humana. Estes três círculos de ideias não têm entre si unidade alguma. Possuímos assim uma antropologia científico-natural, uma antropologia filosófica e uma antropologia teológica, que mutuamente se ignoram – *do homem, porém, não possuímos nenhuma ideia unitária*. (SCHELER, 2008, p. 15)

Segundo ele, o que distinguiria o *homem* dos demais seres vivos não diria respeito ao estágio de evolução biológica desse em relação aos demais seres vivos, mas pelo fato do *homem* ser portador de uma espiritualidade, um ímpeto original que se apresenta como abertura para a compreensão do mundo.

Para Scheler, a tarefa de definir o *homem* se apresenta mais como um desafio à própria capacidade humana de obter conhecimento sobre si mesmo e o mundo, do que como um projeto de investigação com um objetivo possível de ser alcançado. Por ser uma área de interesse de vários campos de conhecimento, o que esperaríamos seria um cenário altamente favorável para o seu desenvolvimento, porém não é isso o que vemos, pelo contrário, “[...] por valiosa que possa ser a multiplicidade sempre crescente das ciências especiais que tratam do homem, ela, em vez de elucidar, oculta a sua essência” (Scheler, 2008, p. 15).

O *Dicionário de Filosofia* (2012) de Nicola Abbagnano (1901-1990) expõe um cenário muito semelhante ao apresentado por Scheler a respeito das tentativas de definição do *homem*. De acordo com Abbagnano podemos classificar em três diferentes tipos as tentativas até aqui feitas com o objetivo de definir o *homem*.

A primeira se dá a partir do estabelecimento de uma relação de comparação entre os conceitos de *Homem* e *Deus*, ora estabelecendo uma semelhança, ora uma diferença, ora afirmado ser o *homem* a imagem de Deus, ora negando a subordinação do *homem* à Deus quando é afirmado que aquilo que o *homem* pensa de Deus serve, também, para definir a si próprio. A segunda, pela compreensão de que o homem possui uma capacidade que lhe é própria e exclusiva, como a de ter uma alma, ter consciência, alcançar conhecimento, produzir tecnologia, dentre outras. A terceira, pela compreensão de que o homem é um ser indeterminado que pode escolher ser o que quiser, um ser que se *autoprojeta*.

Nesse contexto, podemos perceber que são variadas as definições de *homem*, a tal ponto que somos induzidos a concluir que elas parecem se referir a coisas totalmente diferentes. Assim como são variadas as definições, variados são os modos como elas são produzidas. Esse esforço de definição do *homem*, que se apresenta como um lugar de poucos acordos acerca dos muitos e diferentes resultados produzidos com vistas à realização de tal propósito, o vemos como exemplo daquilo que poderia ser a própria imagem do *homem*, que o definiria enquanto tal, ou seja, algo que se vê como distinto e tão singular que não possui um modo necessário de ser definido, nem muito menos uma definição suficiente do que seja.

Nesse cenário, para o qual convergem diferentes atores com papéis que se antagonizam na encenação de uma peça, cujo roteiro representa não haver, os esforços de definição do *homem* parecem enfrentar um estado de crise permanente que os impede de irem além do ponto de partida.

Ernst Cassirer (*Antropologia Filosófica*, s/d) considera que a crise na busca pela definição do *homem* tem origem no antagonismo entre concepções filosóficas que se contrapunham quanto ao papel do homem no tocante ao conhecimento. O antagonismo entre a tese do autoconhecimento e a da necessidade de conhecimento do mundo exterior é considerado por Cassirer a razão da instauração desse estado de crise permanente envolvendo a tentativa de definição do *homem*. A impossibilidade de se decidir sobre qual tipo de conhecimento é o mais importante na tarefa de se produzir uma definição do *homem* criou um impasse, um obstáculo com a aparência de ser intransponível,

contribuindo para a situação de crise na sua realização. Além disso, há a necessidade de que todos os conhecimentos alcançados sobre o homem sejam articulados, o que representaria a superação do impasse ora reinante.

De acordo com Cassirer, a situação de impasse só aumentou desde que ela teve início com os posicionamentos distintos de Platão e Aristóteles. Enquanto o primeiro julgava o conhecimento como algo para o qual as percepções sensíveis não concorriam em nada no seu desenvolvimento; o segundo, embora não considerasse desprezível a função do intelecto na obtenção de conhecimento, julgava necessário o testemunho da percepção. Há um ambiente em que se verifica a existência de uma tensão permanente entre a necessidade de se conhecer o mundo com a de conhecimento de nós mesmos. Mas, mesmo quando o âmbito é restrito ao conhecimento sobre o próprio homem, o estado de tensão não deixa de vigorar.

Tal é a estranha situação em que se encontra a filosofia moderna. Nenhuma outra idade se viu em posição tão favorável no que concerne às fontes do conhecimento da natureza humana. A psicologia, a etnografia, a antropologia e a história reuniram um cabedal de fatos surpreendentemente rico e de constante crescimento. Nossos instrumentos técnicos de observação e experimentação foram imensamente aperfeiçoados e nossas análises se tornaram mais apuradas e mais penetrantes. Apesar disto, não parece que tenhamos encontrado ainda um método para o domínio e a organização deste material. Cotejado com nossa própria abundância, o passado pode parecer paupérrimo. Entretanto, nossa riqueza de fatos não é necessariamente uma riqueza de pensamentos. A não ser que consigamos encontrar o fio de Ariadne que nos tire deste labirinto, não poderemos ter uma visão do caráter geral da cultura humana, e continuaremos perdidos no meio de um conjunto de dados desconexos e desintegrados, carente, ao que parece, de toda unidade conceitual. (CASSIRER, s/d, p. 45)

A multiplicidade de esforços não representa a garantia de um desenvolvimento rápido e cooperativo na busca pela definição do *homem*. Porém, isto não é o mais desconcertante, a situação de impasse que impede o ir além do primeiro passo é que parece ser a principal razão da crise na definição do homem. Não é difícil de se compreender tal situação de impasse quando a vemos como a representação do entrelaçamento das buscas pelas origens do mundo e dos homens. Essas questões estão

profundamente relacionadas entre si, o que torna indecidível a questão de saber qual tipo de conhecimento é mais importante. No entanto, Cassirer observa que, mesmo sendo envolvida numa situação de forte antagonismo, acontece com tentativa de definição do *homem* algo que não esperaríamos, qual seja, o fato de uma explicação absorver elementos de outra.

Nas primeiras explicações mitológicas do universo encontramos sempre uma *antropologia* primitiva ao lado de uma cosmologia primitiva. O problema da origem do mundo está inextricavelmente entrelaçado com a da origem do homem. A religião não destrói estas primeiras explicações mitológicas. Pelo contrário, preserva a cosmologia e a antropologia mitológicas dando-lhes nova forma e nova profundidade. A partir desse momento, já não se concebe o conhecimento de si mesmo como um interesse meramente teórico. Não é simplesmente um tema de curiosidade ou especulação; passa a ser proclamado a obrigação fundamental do homem. Os grandes pensadores religiosos foram os primeiros a inculcar essa exigência moral. Em todas as formas superiores da vida religiosa, a máxima “Conhece-te a ti mesmo” é considerada como um imperativo categórico, lei religiosa e moral básica. Neste imperativo sentimos, por assim dizer, uma súbita inversão do primeiro instinto natural de conhecer – percebemos uma transposição de todos os valores. Na história de todas as religiões do mundo – judaísmo, budismo, confucionismo e cristianismo – podemos observar os passos individuais deste movimento.

O mesmo princípio vale para a evolução geral do pensamento filosófico. Em suas primeiras fases, a filosofia grega parece exclusivamente interessada pelo universo físico. A cosmologia predomina claramente sobre todos os outros ramos da investigação filosófica. Não obstante, o que caracteriza a profundidade e a amplitude do espírito grego é o fato de quase todo pensador grego representar, ao mesmo tempo, um novo tipo geral de pensamento. Além da filosofia física da Escola de Mileto, os pitagóricos descobriram uma filosofia matemática, enquanto os pensadores eleáticos são os primeiros a conceber o ideal de uma filosofia lógica. (CASSIRER, s/d, p. 18)

Mesmo com críticas como a de Cassirer de que a proliferação de tentativas de definir em que consiste a natureza humana antes de favorecer o alcance de tal objetivo, representa, na verdade, uma dificuldade a mais, o surgimento de novas perspectivas de tratamento da questão não enfraqueceu. Atualmente a questão é vista de uma perspectiva em que a tecnologia ocupa um lugar de suma relevância, pois ela é apresentada como a

possibilidade do homem alcançar a tão desejada condição de controle dos acontecimentos.

Em *El hombre postorgánico* (SIBILIA, 2005) nos é apresentada o que passou a ser considerada como a quarta ferida narcísica na imagem do *homem*, pois, mais uma vez, o homem deixa de ser o centro dos esforços e preocupações no desenvolvimento do conhecimento. Nesta obra a autora aborda a distinção entre a tecnociência *prometeica* e a tecnociência *fáustica*.

O primeiro tipo de tecnociência tinha um objetivo amplo, a emancipação do homem, a superação das dificuldades que impedem uma existência plena ao suplantar as limitações impostas à espécie pela natureza e pela própria sociedade humana.

Se a tradição prometeica pretendia dobrar tecnicamente a natureza, o faz mirando ao “bem comum” da humanidade e a emancipação da espécie, sobretudo das “classes oprimidas”. Acreditando no papel libertador do conhecimento científico, este tipo de saber almeja melhorar as condições de vida através da tecnologia. O desenvolvimento gradual desse tipo de saber levaria à construção de uma sociedade racional, fundamentada numa sólida base científico-industrial capaz de erradicar a miséria humana. (SIBILIA, 2005, p. 45, tradução nossa)

O segundo tipo de tecnociência rompe com os ideais de emancipação humana e passa a mirar o aumento da capacidade humana em prever e controlar, tanto os acontecimentos quanto os próprios indivíduos.

De acordo com a perspectiva fáustica, então, os procedimentos científicos não teriam mais como meta a verdade e o conhecimento da natureza íntima das coisas, senão uma compreensão restrita dos fenômenos para exercer a previsão e o controle; ambos propósitos estritamente técnicos. É inevitável associar os critérios fáusticos à tecnociência contemporânea. Até poderíamos insinuar que existe uma certa afinidade entre a técnica fáustica – com seu impulso para a apropriação ilimitada da natureza (humana e não-humana) – e o capitalismo, com seu impulso para a acumulação ilimitada de capital. Esse projeto parece estar alcançando seu ápice hoje em dia, como se observa na vertiginosa corrida tecnológica e sua inextricável relação com os mercados globalizados. (SIBILIA, 2005, p. 50, tradução nossa)

A tecnociência fáustica é o tipo de mudança que indica que a percepção de Cassirer, de que a tarefa de definição do *homem* se revela como uma grande confusão de tendências e perspectivas com pouca possibilidade de unificação, estava correta.

Há um elemento que deve ser destacado nessa situação problemática. Ao mesmo tempo em que é o objeto da definição, o *homem* é o próprio agente dela, o que torna a situação praticamente sem solução possível. Nesta situação peculiar o objeto de observação é alterado a cada observação que faz de si próprio, o que faz com que a observação seja, também, constantemente alterada. Ao tentarmos nos definir, a própria busca faz com que alteremos constantemente nossa perspectiva daquilo que deve ser buscado. Tal situação pode nos levar a dois tipos de atitudes, resignarmo-nos de que o conhecimento do homem envolve uma submissão àquilo que intuímos ou de que o mesmo é impossível de ser alcançado.

Talvez o maior desafio da capacidade humana de conhecer esteja em superar o paradoxo que parece ser o esforço do homem em se autodefinir.

Algumas definições de ‘homem’

As tentativas de definir o *homem* são empreendidas em praticamente todas as esferas em que a expressividade humana pode ser exercida. Todavia, como vimos acima, o fato de haver uma multiplicidade de esforços, não faz com que tal tarefa se torne mais fácil de ser executada, pelo contrário, ela se torna, com isso, ainda mais complexa e difícil de ser levada à cabo.

Ao refletirmos sobre o que somos, nos colocamos em uma determinada posição hierárquica com relação a todas as demais coisas. À princípio parecemos nos submeter a uma posição secundária, mas verdadeiramente fazemos com que tudo acabe por indicar que o topo da pirâmide é merecidamente ocupado por nós. Contudo, existem abordagens que negam essa preponderância do *homem* sobre as demais coisas. Gregory Bateson (*Mind and Nature*, 1979), antropólogo norte-americano, aborda o *homem* a partir da ideia de sistema. Para ele, a totalidade de tudo que há constitui um sistema composto de partes

que competem entre si ao mesmo tempo, porém dependem umas das outras. De acordo com essa concepção, nada existe de maneira isolada, para que algo seja é necessário o suporte de um sistema que lhe garanta condições para a sua sobrevivência e desenvolvimento. Ele critica a separação entre objetos mentais e objetos não-mentais, para ele mente e natureza fazem parte de uma mesma coisa, não possuem propósitos distintos e separados. O *homem* é considerado como sendo mais um dos elementos que constituem a composição de um determinado ambiente.

Assim, num contexto em que são muitas e variadas as formas de tratamento da questão, como é o da definição de *homem*, optamos por destacar três aspectos enfatizados em distintas definições deste por entendermos que eles possuem profunda relevância para as questões relativas a hipótese de trabalho dessa tese. O primeiro define o homem como um “animal simbólico”; o segundo, como um “ser que produz tecnologia” e; o terceiro, como um “ser propenso à autodestruição”.

O ‘animal simbólico’

Ernst Cassirer em *Antropologia Filosófica* (s/d) apresenta um aspecto específico do homem que o distingue de todas as demais coisas vivas. Para ele o homem encontrou uma maneira própria de se adaptar ao meio, não se limitando a utilizar os mesmos recursos empregados pelos outros seres vivos.

De acordo com a investigação biológica, os seres vivos se distinguem das demais coisas por possuírem dois sistemas: um que recebe os estímulos do meio e, outro que responde a esses estímulos. É da interação entre esses dois sistemas que a vida de um organismo é mantida e desenvolvida, o que foi chamado por Johanes von Uexküll como *círculo funcional do animal*.

Conforme von Uexküll, a investigação biológica, embora seja considerada uma ciência da natureza, não procede exatamente como a Física e a Química, pois a Biologia trata com um fenômeno que possui uma realidade própria, a vida. Esta não pode ser descrita levando-se em consideração meramente as características físico-químicas dos

seres vivos. Enquanto os seres inanimados estão submetidos a uma realidade que pode ser considerada a mesma para todos, os seres vivos possuem, cada um, sua própria realidade em função dos estímulos e respostas que dão ao longo de suas vidas, o que constitui suas experiências particulares.

Cassirer considera Uexküll um behaviorista porque este pretende produzir uma interpretação da vida animal baseado não em elementos inerentes ao observador, mas às características próprias do objeto em observação para tentar reconstruir a experiência vivida pelo objeto em questão.

Porém, há uma distinção marcante entre os seres humanos e os outros seres vivos. Enquanto os demais seres vivos possuem um duplo sistema de manutenção e desenvolvimento da vida, os seres humanos possuem um sistema triplo, que nos dá a capacidade de criar um mundo próprio, um mundo em que a natureza humana encontra todas as condições para se desenvolver de modo pleno em confronto com as suas próprias condições e limitações.

Uma das evidências a favor da tese da existência de um terceiro sistema no círculo funcional do homem são os diferentes tipos de respostas que são dadas por este conforme é estimulado pelo meio. Uma é orgânica e rápida; enquanto a outra é mediada e lenta, revelando possuir uma origem e desenvolvimento distintos.

Existe uma diferença inequívoca entre as reações orgânicas e as respostas humanas. No primeiro caso, a resposta dada a um estímulo exterior é direta e imediata; no segundo, a resposta é diferida. É interrompida e retardada por um lento e complicado processo de pensamento. (CASSIRER, s/d, p. 49)

É nesse pequeno intervalo entre uma resposta instantânea e uma elaboração complexa que se situa o mundo dos homens. Um mundo com características e leis que lhes são próprias, um mundo que nos distancia daquele marcado pelo duplo sistema de manutenção e desenvolvimento da vida. Os seres humanos estão submetidos de forma incondicional ao sistema tripartite de estímulo e resposta ao meio, e isso não representa necessariamente uma vantagem evolutiva nossa em relação aos demais seres vivos, pelo

contrário, essa inovação se apresenta, às vezes, mais como um elemento complicador envolvido em nossas tomadas de decisão do que propriamente um facilitador.

Contudo, o mais instigante nessa concepção é a consideração de que os seres humanos vivem praticamente num mundo à parte, num mundo em que as leis naturais parecem se encontrar em estado de suspensão em detrimento de um conjunto de novas regras e leis elaboradas de acordo com o funcionamento do nosso próprio pensamento. O conjunto de experiências que constituem a realidade humana não só se equipara aos fatos que determinam a vida dos outros seres vivos, como se sobrepõe a eles, de tal modo que a vida humana perde o sentido se eles forem simplesmente omitidos.

Em lugar de lidar com as próprias coisas, o homem, em certo sentido, está constantemente conversando consigo mesmo. Envolveu-se por tal maneira em forma linguísticas, em imagens artísticas, em símbolos míticos ou em ritos religiosos, que não pode ver nem conhecer alguma senão pela interposição desse meio artificial. (CASSIRER, s/d, p.50)

Essa peculiaridade do mundo humano é a razão dos diferentes tipos de respostas que damos aos estímulos do meio ambiente. Todo esse incremento constitui o *sistema simbólico do homem*, o elemento a mais na sua constituição cósmica e, que acaba por determinar nossa singularidade.

A linguagem é um fenômeno constituído de diferentes níveis, sendo o mais básico aquele que diz respeito às emoções, elas se encontram em praticamente todos os níveis da expressividade, “com exceção talvez das sentenças formais puras da matemática – que não tenha certo matiz afetivo ou emocional” (CASSIRER, s/d, p. 56). Essa capacidade de expressão não é exclusiva do homem, outros seres vivos também são capazes de se comportar simbolicamente, isto é, são capazes de estabelecer uma relação entre os elementos que constituem o mundo. No entanto, os seres humanos interagem de uma forma bem mais intensa com os símbolos do que outros animais.

O domínio do simbólico implica a possibilidade de construção de uma nova realidade que não se fundamenta diretamente num substrato material, mas sobre o *significado*. É essa capacidade que torna o mundo humano tão peculiar e próprio. Isso ocorre devido ao salto que o homem faz do sinal para o símbolo. Diferente dos demais

animais, o conteúdo expressivo da linguagem humana transpõe os limites da experiência subjetiva, ela se universaliza porque nos tornamos capazes de realizar a operação de substituir um correlato do mundo por um nome. O simbólico é instaurado ao compreendermos o processo de nomeação de um objeto.

Tocamos aqui o ponto crucial de todo o nosso problema. A diferença entre a *linguagem proposicional* e a *linguagem emocional*, verdadeiro marco divisório entre o mundo humano e o mundo animal. Todas as teorias relativas à linguagem animal carecem de significação se deixarem de reconhecer esta diferença fundamental. Em toda a literatura sobre o assunto parece não haver uma única prova concludente de que algum animal tenha dado o passo decisivo da linguagem *subjetiva* para a *objetiva*, da linguagem afetiva para a proposicional. (CASSIRER, s/d, p. 57)

O salto para a dimensão simbólica, que nos conferiu uma nova condição de existência, depende de vários fatores, um deles, necessariamente, é o fato de nos tornarmos capazes de dar nomes às coisas. Sem esse passo a universalização do simbólico permaneceria em estado de latência, não haveria possibilidade de compartilhamento dos significados. A instauração do simbólico no *círculo funcional do homem* depende, necessariamente, da concorrência de dois fatores. De um lado estão as emoções, os afetos, os sonhos e os desejos humanos; de outro, um fenômeno de linguagem, qual seja, nossa capacidade de dar nomes às coisas. O ser humano “[...] precisou compreender que tudo tem um nome – que a função simbólica não se restringe a casos particulares, mas é um princípio da aplicabilidade universal, que abarca todo o campo do pensamento humano”. (CASSIRER, s/d, p. 64)

O salto para o simbólico se realiza plenamente quando o símbolo se torna autônomo, não sendo mais submetido às mesmas determinações que os objetos, “[...] na mentalidade primitiva muito raramente se atinge esta consciência: o símbolo ainda é considerado propriedade da coisa, à semelhança de outras propriedades físicas” (CASSIRER, s/d, p. 67).

No mundo simbólico temos uma autodeterminação que não é possível de ser alcançada no mundo físico. Podemos realizar praticamente o que quisermos com os símbolos, eles nos permitem cria-los e moldá-los *significativamente* conforme a nossa

imaginação. Os símbolos, ao se tornarem independentes das coisas que nomeiam, abriram ao *homem* um novo mundo que permite plena expressividade dos seus sentimentos e emoções.

A característica mais acentuada dessa virada objetiva, no tocante à expressividade humana, está profundamente relacionada com as infinitas possibilidades de uso que os símbolos nos permitem fazer deles. Este foi o grande ganho obtido pelo *homem* quando saltou para a dimensão simbólica, encontrou algo que é capaz de se ajustar à sua capacidade criadora. Com isso, as possibilidades de arranjos significativos se tornam praticamente infinitas.

Pelo fato de tudo ter um nome, a aplicabilidade universal é uma das maiores prerrogativas do simbolismo humano. Mas não é a única. Existe outra característica dos símbolos, que acompanha e completa esta última, e forma seu correlativo necessário. Um símbolo não é apenas universal, porém extremamente variável. (CASSIRER, s/d, p. 66)

Um símbolo humano genuíno não se caracteriza pela uniformidade, mas pela versatilidade. Não é rígido nem inflexível, é móvel. (CASSIRER, s/d, p. 67)

Os seres humanos possuem uma atividade interior que irrompe na forma da expressividade simbólica, um âmbito da vida que não se restringe às leis que governam os corpos, ela surge como um desejo de não somente constatar algo, mas de imprimir algo de si sobre o mundo.

Um ser que produz tecnologia

Uma das características consideradas mais marcadamente humana é o fato deste produzir tecnologia. Sua relação com o que é essencial no homem torna a sua compreensão a condição *sine qua non* para o entendimento deste.

Os questionamentos e problematizações em torno da *tecnologia* adquiriram bastante importância e desenvolvimento ao longo do século XX, ao ponto de passarem a ser considerados um âmbito próprio de reflexão. A tecnologia é compreendida, dentre

outras coisas, como uma espécie de compensação para com a situação de incompletude do homem. Quando assim compreendida, ela vem acompanhada de expectativas e angústias, esperanças e apreensões.

Há uma gama bastante variada de abordagens da *tecnologia* que procuram entender a história da reflexão sobre essa atividade do homem dando relevância a seus diferentes aspectos.

A tecnologia adquiriu, nos últimos dois séculos, maior relevância na compreensão da relação do homem com o mundo. Don Ihde (*Technics and práxis*, 1979) elaborou uma classificação de quatro tipos de diferentes relações estabelecidas entre o *homem* e o *mundo*, que ele chamou de *embodiment*, *hermeneutic*, *alterity* e *background*, para explicar como as tecnologias modelam (*shaping*) tanto um quanto o outro. De acordo com Ihde, as modelagens são provocadas pela acentuação (*magnification*) ou diminuição (*reduction*) de certas características do homem e do mundo.

A visão das montanhas da lua, através de todo o poder de transformação do telescópio, remove a lua da sua configuração na extensão dos céus. Mas se nossas tecnologias fossem apenas para replicar nossa experiência imediata e corporal, elas seriam de pouco uso e, finalmente, de pouco interesse. (IHDE, 1979, p. 76, tradução nossa)

Essa dupla face do processo de modelagem da tecnologia sobre o mundo e o homem além de expressar como se dá esse processo, também serve para analisar a própria tecnologia.

Com bastante frequência não consideramos que a questão seja se uma tecnologia emergente é boa ou ruim, ou se devemos escolher que ela entre nas nossas vidas ou não, mas como podemos modelar nossas vidas em face dessa tecnologia. Expor a dupla face, portanto, contribui para uma importante ampliação da base em que avaliamos as tecnologias, tanto no nível social quanto no nível pessoal. (KIRAN, In: ROSENBERGER & VERBEEK, 2015, p. 124, tradução nossa)

Nos enganamos quando pensamos que a reflexão sobre a *tecnologia* devesse começar com a elaboração da sua definição, mesmo porque este é considerado *um termo escorregadio* (BIJKER, HUGHES & PINCH, *The Social construction of technological*

systems, 2012). Essa tarefa não é colocada como um objetivo central daqueles que se dedicam a refletir sobre as questões suscitadas por esse tipo de atividade.

Thomas J. Misa (BERG OLSEN *et alli*, *An Companion to the Philosophy of Technology*, 2009) no capítulo 1, afirma que assim como os historiadores da *arte* não sentem necessidade de definir o que seja a *arte* a fim de dar continuidade às suas atividades, também os historiadores da tecnologia não veem que seja necessário definir o que seja a *tecnologia* para se certificarem de que este é um âmbito de reflexão que possui um espaço próprio e relevância assegurados. Tais atividades envolvem um elevado nível de complexidade que se torna quase impossível o entendimento do que exatamente elas são.

Talvez essa seja a razão principal para que o objetivo de definir em que consiste a *tecnologia* não seja considerado altamente relevante para a reflexão em torno dela, sendo, assim, relegado a um plano secundário. De acordo com Misa (2009), duas abordagens da tecnologia se destacam ao longo da história. Uma que visa ao tratamento de questões que digam respeito ao desenvolvimento estrito dos sistemas e dispositivos tecnológicos e a outra que trata de questões relacionadas aos impactos produzidos pela tecnologia no quotidiano dos seus usuários e neles próprios.

A maioria dos historiadores que escrevem sobre tecnologia tem definido o termo mais pela apresentação e discussão de exemplos pertinentes. Muitos dos historiadores que estudam o século XX concentram seus estudos nos grandes sistemas tecnológicos, tais como eletricidade, produção industrial e transporte, surgidas nas primeiras décadas e que se tornaram mais ou menos difundidas no ocidente durante a segunda metade desse século. No entanto, outros historiadores desse mesmo século, preferiram firmemente examinar as tecnologias da perspectiva da *vida cotidiana*, da perspectiva do *usuário*. (BERG OLSEN ET ALLI, *An Companion to the Philosophy of Technology*, 2009, p. 7, tradução e itálicos nossos)

Embora autores que se dedicam a refletir sobre a tecnologia critiquem algumas das abordagens que buscam nas raízes filosóficas do problema os elementos que tornem compreensível tal atividade, como é o caso de Joseph Pitt em *In search of a new Prometheus* (PAUL T. DURBIN, *Broad and narrow interpretations of philosophy of technology*, 2012) que critica o trabalho de Frederick Ferré por considerar este um esforço

para que permaneçam obscuras as questões de fundo que se fazem presentes na reflexão sobre a tecnologia, é inegável que os conceitos de *homem* e *tecnologia* estejam profundamente relacionados um com o outro.

A ideia de que o ser humano não é o ser mais forte e completo dentre os seres vivos já é mencionada no início da filosofia quando o personagem de Platão afirma que a condição humana é a de estar despreparado, desprotegido e desamparado quando comparado aos demais seres vivos (*Protágoras*, 321 c). Esta condição inicial impõe a necessidade de elaboração e construção de um suporte que garanta as condições de sobrevivência necessárias à manutenção e ao desenvolvimento da forma peculiar de existência que é a do *homem*, a quem caberia a realização de tais propósitos. É como se a natureza desse ao *homem* uma chance para este se tornar viável, contudo, depende totalmente de seus esforços o alcance dessa condição. Desse ponto de vista, não há como ser dissociada do *homem* a sua relação especial com a tecnologia, já que ela é o suporte que garante a sua sobrevivência.

Houve uma época em que a ciência e a tecnologia eram vistas como algo que representava a possibilidade de uma vida para o homem em que este teria como preocupação o ocupar-se em bem viver. As invenções técnicas humanas resolveriam, através da substituição, muitas das atividades que desviam os homens do bem viver. O conhecimento científico-tecnológico era enaltecido como uma fonte de soluções para muitos dos problemas que afligem aos homens.

Francis Bacon (1561-1626) afirma no *Novum Organum* que o esforço de conhecimento do homem pode reparar, em parte, a perda produzida a este pelo pecado original.

Pelo pecado o homem perdeu a inocência e o domínio das criaturas. Ambas as perdas podem ser reparadas, mesmo que em parte, ainda nesta vida; a primeira com a religião e com a fé, a segunda com as artes e com as ciências. Pois a maldição divina não tornou a criatura irreparavelmente rebelde; mas, em virtude daquele diploma: *Comerás do pão com o suor de tua fronte*,²⁶¹ por meio de diversos trabalhos (certamente não pelas disputas ou pelas ociosas cerimônias mágicas), chega, enfim, ao homem, de alguma parte, o pão que é destinado aos usos da vida humana. (BACON, <http://br.egroups.com/group/acropolis/>, p. 232)

Bacon é considerado a pai da ciência moderna porque, dentre outras coisas, atribuiu a ela uma tarefa que transcende aos seus propósitos de natureza científica, qual seja, o nobre objetivo de tornar melhor o viver humano. Fundamentando-se no texto bíblico, ele classifica o esforço humano de alcançar conhecimento rigoroso como revelador da capacidade deste de conhecer, em parte, como foram feitas as obras de Deus.

Deus moldou a mente do homem como um espelho ou vidro capaz de refletir a imagem do universo, e feliz de receber a impressão dele, como o olho é feliz de receber luz; e que não apenas se deleita com a contemplação da variedade das coisas e as vicissitudes dos tempos, como também se eleva para descobrir e discernir as ordenanças e decretos que através de todas essas mudanças são infalivelmente observados. E embora ele insinue que a lei ou suma suprema da natureza, que ele denomina *a obra que Deus faz do princípio ao fim*, é impossível de ser descoberta pelo homem, ele, contudo, não menoscaba a capacidade da mente, mas o põe como algo que pode ser atribuído a impedimentos tais como a brevidade da vida, a má conjunção de esforços, a transmissão defeituosa do conhecimento de mão em mão, e muitas outras inconveniências a que a condição do homem está sujeita. Pois que nenhuma parcela do mundo está vedada à inquisição e invenção do homem, ele decreta em outra passagem, quando diz: *O espírito do homem é como a lâmpada de Deus, com a qual ele esquadriinha a interioridade de todos os segredos.* (BACON, 2007, p. 21)

A tecnologia e o conhecimento dele derivados eram instrumentos considerados indispensáveis para que o ser humano conseguisse atingir um viver pleno, um viver voltado para a contemplação da obra divina, uma espécie de redenção parcial do homem frente ao desastre do pecado original.

A reflexão de Bacon já nos apresenta alguns questionamentos que darão origem ao que é chamado hoje de *filosofia da tecnologia*. Embora a reflexão de Bacon sobre o homem e o conhecimento já contenha elementos que se farão presentes na reflexão posterior sobre a tecnologia como, por exemplo, a necessidade de incorporação de instrumentos que ampliem nossa capacidade de observação das coisas. Ela não chega a ser ainda uma filosofia da tecnologia, não era considerada um âmbito próprio de reflexão.

Isso só viria a ocorrer no final do séc. XIX, com a publicação da obra de Ernst Christian Kapp (1808-1896) intitulada *Grundlinien Einer Philosophie der Technik* no ano de 1877.

A *filosofia da tecnologia* possui duas tradições (MITCHAM, 1989), uma derivada do interesse dos engenheiros, cientistas e técnicos envolvidos na pesquisa e aplicação de conhecimentos por problemas de natureza filosófica relacionados à tecnologia (*engineering philosophy of technology*), e outra originada pelo interesse de filósofos, sociólogos, antropólogos na compreensão e problematização da tecnologia (*humanities philosophy of technology*). Conforme este pensador, a tecnologia se distingue da técnica na medida em que a primeira tem origem no conhecimento científico, sem deixar de empregar as técnicas próprias do agir humano.

Como explicita no título, Kapp trata do que devemos levar em consideração se pretendemos tornar as questões em torno da tecnologia um campo próprio de reflexão. É considerado o primeiro a pensar de forma sistemática e aprofundada as questões pertinentes à filosofia da tecnologia.

[...] a relação intrínseca que se estabelece entre os instrumentos e os órgãos, relação que deve ser esclarecida e destacada – embora a mesma seja mais um descobrimento inconsciente do que uma invenção consciente-, é que nos instrumentos o humano se reproduz continuamente a si mesmo. Como o fator de controle é o órgão cuja utilidade e poder devem ser aumentados, a forma apropriada de um instrumento só pode ser derivada desse órgão. A riqueza das criações espirituais brota, pois, da mão, do braço e dos dentes. Um dedo dobrado se converte em um gancho, a palma da mão em um prato; na espada, na lança, no remo, na pá, no rastilho, no arado e na enxada, se observam diversas posições do braço, da mão e dos dedos, cuja adaptação para a caça, a pesca, a jardinagem e as ferramentas agrícolas é facilmente visível. (KAPP, *Grundlinien einer philosophie der technik*, 1877, p. 44, citado por Carl Mitcham em *Que es la filosofia de la tecnologia?*, 1989, p. 30)

Conforme Carl Mitcham (1941) (*Que es la filosofia de la tecnologia?*, 1989), são três os momentos inaugurais da *engineering philosophy of technology*. O primeiro com a publicação do livro de Ernst Kapp em que a tecnologia é vista como projeção dos órgãos humanos, o segundo com Peter Klimentievich Engelmeier e o emprego da tecnologia na

criação da tecnocracia e o terceiro com a visão mística de Friedrich Dessauer onde a tecnologia se torna uma experiência religiosa.

De acordo com Mitcham (*Que es la filosofía de la tecnología?*, 1989), Kapp via a tecnologia como uma potencialização dos órgãos humanos através da ampliação das suas capacidades e possibilidades.

[...] por meio de uma reflexão a partir da sua própria experiência, empreendeu a tarefa de formular uma filosofia da tecnologia, na qual os instrumentos e armas eram compreendidos como diferentes tipos de “projeções dos órgãos”. (MITCHAM, 1989, p. 29, tradução nossa)

Este foi fortemente influenciado por George Wilhelm Friedrich Hegel (1770-1831), chegando a escrever uma obra com o propósito de tornar aplicável a teoria do idealismo de Hegel.

Como Marx, Kapp foi um hegeliano de esquerda. Sua melhor obra, editada em dois volumes sob o título de *Vergleichende allgemeine Erdkunde* (1845) revela, como os Manuscritos Econômico-Filosóficos de 1844 de Marx, a intenção em traduzir o idealismo dinâmico de Hegel para sólidos termos materialistas. (MITCHAM, 1989, p. 25, tradução nossa)

Contudo, Kapp possuía uma concepção de história diferente da hegeliana, ela não se trataria da implantação necessária da *Ideia Absoluta*, mas do registro das tentativas humanas de enfrentamento dos desafios impostos pelos diferentes ambientes nos quais ele habita. Para ele, o homem almeja a condição de se tornar independente da natureza.

Kapp também foi influenciado pelo pensamento do geógrafo alemão Karl Ritter (1779-1859), que propôs que as características geográficas de um determinado ambiente produzem influência no ordenamento social e cultural.

Porém, enquanto o materialismo de Marx aspirava sintetizar a teoria da história de Hegel com a nova ciência da economia, o de Kapp buscava vincular a história com a nova ciência da geografia de Ritter. A “geografia universal comparada” de Kapp se antecipou ao que hoje poderia ser denominada de *filosofia do meio ambiente*. Por um lado, este trabalho enfatizava, como o de Ritter, as influências formativas da geografia, especialmente os corpos de água, na ordem sociocultural. Os rios, os mares e os oceanos, afetam não só as economias e culturas

gerais, como também as estruturas políticas e as organizações militares. (MITCHAM, 1989, p. 25, tradução e itálicos nossos)

O homem não é um ser isolado que vive sem a influência das coisas que existem a sua volta. Pelo contrário, elas influenciam tanto a existência humana que se torna necessário ao homem a elaboração de uma estratégia que garanta a sua sobrevivência. A execução dessa estratégia de sobrevivência constitui, para Kapp, o que ele chamou de *processo de colonização* do ambiente. Tal processo teria por objetivo tornar o ambiente natural menos hostil e perigoso ao homem, favorecendo, assim, o seu desenvolvimento. Por outro lado, a realização plena de tal objetivo demandaria ao homem a necessidade de ajustar-se às mudanças feitas com vistas à criação das condições para o seu desenvolvimento. Isto é, o *processo de colonização* tem dois âmbitos: no primeiro, as mudanças são relativas ao ambiente natural; no segundo, referem-se ao próprio homem, cujo desenvolvimento pleno só será alcançado se este for capaz de se adequar às mudanças no meio ambiente.

Isto requer o cultivo do espaço (por meio da agricultura, da mineração, arquitetura, engenharia civil, etc.) e do tempo (principalmente mediante os sistemas de comunicação, desde as línguas até os telégrafos). Estes últimos, em sua forma aperfeiçoada, viriam a constituir uma “telegrafia universal”, relacionando as línguas mundiais, a semiótica e as invenções em uma transformação global da terra e do verdadeiro *habitat* humano. Porém, isso só será possível quando a colonização externa do ambiente natural for complementada por uma colonização interna do ambiente humano. (MITCHAM, 1989, p. 26, tradução nossa)

Para Mitcham, Ernst Kapp não se limitou a traçar uma comparação entre os órgãos humanos e os instrumentos.

Seu livro inclui (capítulo 10) a primeira reflexão filosófica sobre a nova ciência da engenharia mecânica (da mesma maneira que a análise de Franz Reuleaux em sua clássica obra *Theoretische Kinematik: Grundzuge einer Theorie des Maschinellwesens* [1875]), na qual encontramos semelhanças entre a descrição de Reuleaux da máquina como algo dotado de limitações metodológicas e o caráter da ética, que também propunha impor limites de princípio à ação humana. (MITCHAM, 1989, p. 32, tradução nossa)

Nessa passagem podemos ver que a relação entre a tecnologia e a ética é bastante forte. A reflexão em torno do funcionamento das máquinas e de tudo o que é produto da tecnologia levanta questões de natureza ética, principalmente no que diz respeito ao controle do desenvolvimento tecnológico.

Conforme Mitcham (*Philosophy and Technology II*, 1986), a tecnologia não teria muito a oferecer em termos de controle do desenvolvimento tecnológico, sendo esse um dos fatores que levam à desconfiança da tecnologia.

Esse tema é abordado por Mitcham por meio da análise do conceito de *incontinência*. Num breve artigo ele trata da questão de por que optamos por algo, mesmo sabendo que ele não é o mais recomendável racionalmente? A *incontinência* é essa forma de agir em que o sujeito se mostra incapaz de proceder de acordo com um conhecimento. Isso seria consequência do livre arbítrio. As tecnologias da informação pouco teriam a contribuir com relação ao controle racional da tecnologia, pois os seres humanos quando agem, não o fazem somente empregando racionalidade.

A ação humana, em última instância, não é determinada por uma razão. Há algo de mais fundamental, básico e real, nomeadamente a vontade. E isso, é testemunhado pelo fato da incontinência; sabendo o que é bom num determinado nível de racionalidade, os seres humanos, no entanto, frequentemente fazem outra coisa. (MITCHAM, 1986, p. 255, tradução nossa)

Contemporâneo de Ernst Kapp, embora 47 anos mais jovem, Peter K. Engelmeier (1855-1941) teve grande importância no desenvolvimento da reflexão da *filosofia da tecnologia* como um campo próprio de investigação.

Engelmeier empregou o termo *filosofia da tecnologia* em artigos que publicou em revistas alemãs. Para ele, a tecnologia não se resume na elaboração de produtos bons e baratos. Hoje a tecnologia está presente em praticamente todas as atividades humanas, da construção de estradas à administração pública.

O emprego da tecnologia nos mais diversos setores da vida seria uma das consequências do crescimento econômico. Essa realidade coloca um desafio à tecnologia e ao tecnólogo, já que as expectativas em torno das possibilidades da tecnologia são altamente positivas, qual seja, estariam os tecnólogos preparados para dar conta das novas

demandas? Para que essa questão possa ser respondida, é necessário, na opinião de Engelmeier, que seja realizada uma investigação mais aprofundada das influências da tecnologia sobre a sociedade e vice-versa.

Esta pergunta dificilmente pode ser respondida afirmativamente, porque ele não somente inclui o manejo de nossa especialização no sentido de prática tecnológica, senão que, também, aponta para uma visão de grande alcance, as interações entre a tecnologia e a sociedade. Temos de investigar o que representa a tecnologia, os principais objetivos que perseguem seus distintos ramos, que tipo de método usa, onde termina seu âmbito de aplicações, que áreas da atividade humana a cercam, sua relação com a ciência, a arte, a ética, etc. Devemos desenvolver um quadro completo da tecnologia no qual analisemos tantas manifestações técnicas quanto seja possível [...] porque a tecnologia é a primavera no grande relógio mundial do desenvolvimento humano. (ENGELMEIER, In: *Que es la filosofía de la tecnología?*, MITCHAM, 1989, p. 33, tradução nossa)

Posteriormente, numa comunicação intitulada *Philosophie der Technik*, feita no IV Congresso Mundial de Filosofia realizado em Bolonha no ano de 1911, Engelmeier apresentou uma reformulação da compreensão de como deveria se dar essa investigação ao procurar compreender a razão que leva os homens a produzirem tecnologia, evidenciando, desse modo, a relevância da compreensão sobre o próprio ser humano.

Ao iniciar com uma descrição do “império da tecnologia”, [Engelmeier] considera a relação entre ciência e tecnologia e faz uma análise filosófica da essência da tecnologia para dar conta do alcance deste domínio. Ele o encontra, finalmente, no desejo humano e na vontade inerente voltada à tecnologia. (MITCHAM, 1989, p. 34, tradução nossa)

Defendeu a racionalidade tecnológica de tal modo que foi um dos responsáveis pela implantação da tecnocracia na administração do estado soviético, vindo a ser morto pela burocracia comunista ao defender tal concepção de organização estatal. Em sua concepção, os tecnólogos e engenheiros ocupariam um lugar de maior relevância que os próprios políticos, as mudanças, correções de rumo e aperfeiçoamento da política, de acordo com o seu pensamento, deveriam seguir princípios tecnológicos.

Sua reflexão em torno dos problemas relacionados à tecnologia o levou a conceber uma série de problemas que deveriam ser abordados, tendo em vista elevar a *filosofia da tecnologia* à condição de um campo próprio de investigação. Ao fazê-lo, verificamos o quanto a tecnologia está presente nos mais diversos aspectos da vida humana. Por isso é compreensível seu engajamento político, ele não havia abandonado a nobre tarefa atribuída à tecnologia de tornar melhor a vida dos homens.

O Círculo sobre Problemas Gerais da Tecnologia [...] se abstém de todo gênero de propaganda. Para o curto prazo, foram traçadas as seguintes tarefas: desenvolver um programa para a filosofia da tecnologia que inclua as tentativas de definir o conceito de tecnologia, os princípios da tecnologia contemporânea, a tecnologia como fenômeno biológico, a tecnologia como fenômeno antropológico, o papel da tecnologia na história da cultura, a tecnologia e a economia, a tecnologia e a arte, a tecnologia e a ética e outros fatores sociais (MITCHAM, 1989, p. 34, tradução nossa)

Segundo Mitcham (*Que es la filosofia de la tecnologia?*, 1989), o quarto engenheiro e pesquisador a empregar o termo *filosofia da tecnologia* dentro da tradição da *engineering philosophy of technology* foi o alemão Friedrich Dessauer (1881-1963). Ele foi um dos primeiros a produzir máquinas de raios X após fundar a empresa VIEFA Werke, vindo a morrer por causas relacionadas à exposição aos raios X.

Segundo Dessauer, a filosofia da tecnologia deveria ter por objetivo “[...] compreender o poder do conhecimento científico-tecnológico, que se converteu, por meio da engenharia moderna, num novo modo de existência dos seres humanos no mundo” (MITCHAM, 1989, p. 46, tradução e itálico nossos). Ele vê na filosofia da tecnologia uma reflexão que é capaz de superar as dificuldades da filosofia da ciência em compreender o impacto gerado pela tecnologia no quotidiano. Para o empreendimento de tal tarefa, ele toma como referência o pensamento kantiano presente nas três críticas. Em certa medida, Kant sempre tem como ponto último de chegada, seja quando está refletindo sobre o conhecimento, sobre a moral ou sobre a estética, o mundo *noumenico*, o mundo das coisas-em-si, o mundo da necessidade, o mundo onde os princípios que determinam o modo de ser das coisas são seguidos absolutamente.

A proposta de Dessauer seria acrescentar uma quarta crítica à obra de Kant, a *crítica da tecnologia*, a qual consistiria em mostrar que os sistemas e dispositivos tecnológicos seguem determinações que estão no nível daquilo que é *noumenico*. Dessauer considera as invenções o resultado da relação harmônica entre as soluções de natureza técnica e as leis naturais. Nesse sentido, a criação tecnológica desencadeia um processo que vai além da expectativa de tornar o viver do homem melhor.

Dessauer argumenta que o fazer, particularmente na forma de invenção, estabelece contato preciso com as coisas-em-si. “A essência da tecnologia não se encontra nem na manufatura industrial (que é meramente uma invenção para a produção em massa), nem nos produtos (que são meramente utilizados pelos consumidores), senão o ato da criação técnica. Uma análise do ato da criação técnica revela que esta tem uma relação harmônica com as leis da natureza tendo em vista propósitos humanos, porém, que os propósitos humanos e naturais são só condição necessária, mas não suficiente para a sua existência. Todavia, há algo mais, o que Dessauer chama “elaboração”, que liga a mente do inventor com um “quarto reino” das soluções preestabelecidas para resolver problemas técnicos. (MITCHAM, 1989, p. 47, tradução nossa)

Tal visão é considerada uma espécie de atitude mística diante da tecnologia, pois vê nesta o cumprimento de desígnios que estariam para além da nossa capacidade de compreensão.

Pelo lado da tradição da *humanities philosophy of technology*, na qual está incluído Heidegger, destacam-se, como pensamentos inaugurais que determinaram a configuração do debate em torno da tecnologia, os de Lewis Mumford e Ortega y Gasset.

Mitcham considera Lewis Mumford um dos fundadores da reflexão em torno da tecnologia por ter mostrado que é necessário dar atenção às mudanças que se dão nas formas de vida e da cultura humanas provocadas pela participação das máquinas no nosso quotidiano. Mas essa não é considerada a sua contribuição mais valiosa para o desenvolvimento da *filosofia da tecnologia*. Mumford escreveu a obra *Technics and Civilization* (1934) considerada uma obra clássica na história da reflexão sobre a tecnologia, nela

[...] reflete possuir um profundo conhecimento das obras de Ure, Zschimmer, Veblen e Dessauer. Aqui, pela primeira vez, Mumford emprega sua concepção de natureza humana para proporcionar uma extensa análise do amplo domínio da civilização mecânica [...]. (MITCHAM, 1989, p. 53, tradução nossa)

Mumford descreve a civilização mecânica como possuindo uma história constituída de três momentos marcantes que ele chamou de *ondas*. Ao observar que as máquinas produzem impacto na vida e cultura humanas, ele se propôs a entender o modo de surgimento e desenvolvimento da *civilização mecânica*.

O que se chama ordinariamente de revolução industrial, a série de mudanças industriais que começaram no séc. XVIII, foi uma transformação que teve lugar no curso de percurso muito mais longo. A máquina revolveu nossa civilização em três ondas sucessivas. A primeira onda, que se deu no séc. X, acumulou força e velocidade enquanto outras forças da civilização se enfraqueciam e dissipavam: este triunfo inicial da máquina foi um esforço para alcançar ordem e poder por meios puramente externos, seu êxito se deveu, em parte, ao fato de que contornava muitos dos problemas autênticos da vida e se afastava das graves dificuldades sociais e morais que não haviam nem sido enfrentadas, nem resolvidas. A segunda onda veio no séc. XVIII, depois de uma longa estagnação na Idade Média, com os aperfeiçoamentos na mineração e a siderurgia; aceitando todas as premissas ideológicas do primeiro esforço para criar a máquina, os discípulos de Watt e Arkwright pretendiam universalizá-las e tirar proveito das consequências práticas. Durante esse esforço, vários problemas morais, sociais e políticos que haviam sido deixados de lado em detrimento do exclusivo desenvolvimento da máquina, se apresentaram, então, com redobrada urgência: a mesma eficiência da máquina foi radicalmente diminuída pelo fracasso em alcançar na sociedade um conjunto de fins harmoniosos e integrados. A regulamentação externa e a resistência e desintegração internas iam de mãos dadas: aqueles afortunados membros da sociedade que estavam em completa harmonia com a máquina alcançaram dito estado somente ao fechar vários caminhos importantes da vida. Finalmente, começamos em nossos próprios dias a observar as crescentes energias da terceira onda: por trás desta, tanto na tecnologia quanto na civilização, há forças que se manifestam agora em todos os setores da atividade e que tendem para uma nova síntese do pensamento e a uma renovada sinergia na ação. Como resultado deste terceiro movimento, a máquina deixa de ser um substituto de Deus ou de uma sociedade ordenada; em vez de seu êxito seja medido pela mecanização da vida, seu valor se torna cada vez mais mensurado em termos de sua própria aproximação ao orgânico e ao vivo. As ondas de retrocesso das primeiras duas fases diminuem um pouco a força da terceira: porém, a imagem segue sendo exata ao sugerir

que a onda que agora está nos transportando se move numa direção oposta à do passado. (MUMFORD, 1934, p. 4)

Ele faz uma importante observação, de que as grandes invenções tecnológicas são precedidas por mudanças de mentalidade. As mudanças tecnológicas não ocorrem por razões estritamente dessa natureza, para que elas se tornem efetivas é necessária uma preparação dos seus novos usuários, uma preparação de naturezas “ideológica e social”, “[...] antes que os novos processos industriais sejam implantados em larga escala, se faz necessária a reorientação de desejos, costumes, ideias e objetivos”. (MUMFORD, 1934, p. 3, tradução nossa). Ao observar que a tecnologia é um fenômeno que está profundamente relacionado com outros aspectos da vida do homem, propõe-se a compreender de que maneira elementos de naturezas psicológica e social, juntamente com as condições materiais influenciam no surgimento e desenvolvimento da tecnologia.

A tecnologia se faz presente no meio da cultura humana porque o homem é, no seu entendimento, um animal pensante. A fabricação e o uso de instrumentos não são considerados por ele os principais vetores de desenvolvimento do homem, esses são dados pela atividade de pensamento do homem que procura, ao desenvolver a tecnologia, mais do que resolver o problema da produção de alimento ou do domínio da natureza, visa dar vazão à capacidade criadora do homem.

A esse respeito, Mitcham (1989) afirma que

Mumford sustenta que embora o ser humano esteja logicamente concentrado em atividades terrenas, ele não deve ser entendido como *homo faber*, mas como *homo sapiens*. Não é o fazer, senão o pensar; não é o instrumento, senão a mente, o que constitui a base da humanidade. (MITCHAM, 1989, p. 54, tradução nossa)

Mumford deu atenção à questão dos reflexos e impactos do desenvolvimento da tecnologia provocados sobre a cultura humana e desta sobre a tecnologia. Ele critica a compreensão de que a tecnologia é um ramo de atividade que não depende de fatores externos para que seu desenvolvimento ocorra. Para ele a tecnologia se dá num contexto que é determinado por razões de várias naturezas, razões estas que determinam o próprio desenvolvimento tecnológico.

Steve L. Goldman (*Philosophy, engineering and western culture*, 1990) salienta que a crítica de Mumford tem semelhança com o pensamento de Vico ao este reconhecer que a tecnologia influencia e é influenciada pelos diferentes aspectos da vida humana.

A filosofia da ciência foi por muito tempo empreendida como se o logos da ciência fosse separável da história da ciência e da prática quotidiana da ciência, como se a essência da ciência fosse um modo abstrato de raciocínio e, assim, acessível ao pensamento puro. (PITT & DURBIN, 1990, p. 147)

Em outra obra (*El mito de la máquina*, 2010), livro em que investiga as influências e transformações produzidas pelas ciências na tecnologia, Mumford distingue dois tipos de tecnologias, a saber: a *politécnica* e a *monotécnica*. A primeira é um tipo emancipatório de tecnologia, que busca desenvolver a potencialidade humana em toda a sua plenitude; a segunda, não visa a emancipação do homem, é um tipo de tecnologia que coloca seus próprios objetivos à frente de qualquer outro, ela tem em vista a domesticação do homem a fim de torna-lo um usuário seu. Enquanto a primeira tem um modo de compatibilidade com os princípios da natureza, o que a torna uma espécie de encaminhadora da felicidade humana; a segunda, antepõe-se ao que é natural e se impõe ao *homem* mesmo em detrimento das necessidades desse. Encontramos duas fontes antagônicas gerando tecnologia, duas forças cujo portador é o homem.

Quase desde o início da civilização, existiram de forma paralela duas tecnologias díspares: uma democrática e descentralizada, a outra, totalitária e centralizada; a “democrática” baseada em operações de artesanato em pequena escala, se mantinha viva em inúmeras aldeias junto com a granja e a criação de gado, estendeu-se às pequenas cidades rurais, chegando, finalmente, nos grandes centros urbanos. (MUMFORD, 2010, p. 387)

O homem revela, assim, mais um paradoxo inerente àquilo que é. Ao comentar o pensamento do antropólogo americano Alfred Louis Kroeber (1876-1960) sobre a propensão do homem à destruição, Mumford expõe a situação paradoxal a que o mesmo se impõe ao se mostrar tanto capaz de elaborar e construir algo quanto de destruí-lo completamente.

Esta propensão construtiva é evidente desde as primeiras etapas de nosso crescimento; se deixarmos uma criança que, todavia, não saiba falar, com uns blocos de papelão ou madeira, espontaneamente construirá algo, colocando uns blocos sobre os outros (como demonstrou Arnold Gesell com os mais variados experimentos); assim como, seguramente, em qualquer momento, o derrubará completamente, esparramando pelo chão esses mesmos blocos com gestos de poderio selvagem. Temos, portanto, razão ao imputar a nossos primitivos antepassados as mesmas qualidades que Erich Fromm atribui aos sonhos de hoje: “a expressão simultânea daquilo que é mais profundo e irracional, junto com as mais altas e valiosas funções das nossas mentes”. (MUMFORD, 2010, p. 94)

Mumford acredita que uma reflexão filosófica sobre a tecnologia que leve em consideração a liberdade e emancipação humanas deverá observar esses diferentes aspectos inerentes à relação do homem com os seus constructos tecnológicos.

Carl Mitcham (1989) considera que a reflexão de Jose Ortega y Gasset sobre a tecnologia possui aspectos semelhantes à de Lewis Mumford. Ortega y Gasset também fundamenta sua reflexão numa teoria da natureza humana. Para ele, a vida do homem é compreendida como um enfrentamento duplo. Na obra *Meditación de la técnica* (1964), ele conclui que “[...]a vida humana não é só luta com a matéria, senão, também, luta do homem com sua alma”. (ORTEGA Y GASSET, 1964, p. 375, tradução nossa)

Ao investigar o modo como a técnica ingressou na cultura humana, passando a fazer do modo de vida ocidental, ele verificou que o homem desenvolve a técnica como um meio de sobrevivência. Seu uso visa a mudança da natureza, visa remover do horizonte da existência humana as necessidades impostas pela natureza. Isso torna o homem um ser especial, pois ao não se ajustar de forma harmônica às vicissitudes da natureza, fazendo com que sua existência seja algo penoso e difícil.

Daí que nossa vida seja pura tarefa e inexorável fazer. A vida de cada um de nós é algo que não nos é dado pronto e acabado, senão algo que está por ser feito. A vida dá muito o que fazer; porém, ademais não é senão esse fazer que dá a cada qual um fazer, repito, não é uma coisa, senão algo ativo, num sentido que transcende a todos os demais. Porque no caso dos demais seres os supomos como alguém ou algo que já é, age; porém aqui se trata precisamente de que para ser é necessário agir, que nada somos sem esse agir. O homem, queira ou não, tem que fazer-se a si mesmo, autofabricar-se. (ORTEGA Y GASSET, 1964, Tomo V, p. 341, tradução nossa)

Essa concepção da natureza humana contrasta com a noção heideggeriana de *homem*. Para Ortega Y Gasset o homem é um ser que se comporta como algo completamente estranho ao ambiente natural, sofre ao ser submetido às necessidades que a natureza lhe impõe. Por sua vez, para Heidegger o homem existe no mundo como um ser natural. A explicação de Ortega y Gasset para a inserção da técnica no quotidiano humano se deve justamente à desarmonia entre este a natureza.

Essa diferença de opinião entre Ortega y Gasset e Heidegger foi originada pela discordância do primeiro da interpretação dos termos da língua alemã, *wohnen* e *bauen*, feita por Heidegger que considerou ambas portadoras do significado de *ser* no sentido de *estar vivo*. Essa equivalência faz Heidegger considerar o homem como mais uma coisa natural que habita o mundo. Ortega y Gasset discorda dessa interpretação porque Heidegger, ao reconstruir etimologicamente aqueles termos, não levou em consideração o que ele chamou de *campo pragmático*, isto é, um conjunto de palavras que constituem o todo significativo de um determinado contexto da vida. Em suma, Ortega y Gasset entende que o homem não habita a Terra como os outros seres, ele não o faz de um modo natural como acreditava Heidegger.

A vida do homem é entendida por Ortega y Gasset como um permanente enfrentamento com a natureza, a técnica entra na vida e cultura humanas porque ela torna possível a criação de um ambiente diferente do natural. Essa técnica leva à criação de um ambiente completamente humano e adaptável às suas demandas.

A fim de explicar de que maneira a técnica passa a influenciar o viver humano, Ortega y Gasset usou dois procedimentos: 1) descrevendo a situação de penúria a que o homem fica submetido quando lhe são impostas necessidades pela natureza e, 2) recorrendo a uma narrativa mítica da origem do homem.

O primeiro procedimento empregado por Ortega y Gasset para explicar a inserção da tecnologia no meio da cultura humana encontramos em *Meditación de la técnica* (1964). Essa elucidação adota um procedimento bastante distinto do procedimento anterior ao fazer um relato dos efeitos provocados pelo clima no comportamento humano. Ao fazê-lo, Ortega Y Gasset sintetiza seu pensamento e determina a amplitude da sua reflexão, sua preocupação em compreender o fenômeno da tecnologia é porque ele, em

última instância, quer saber o que faz o homem não desistir de viver se o ambiente que ele habita não se mostra em quase nada favorável à manutenção da sua existência.

Ocorre que, quando chega o inverno, o homem sente frio. Este sentir frio do homem é um fenômeno em que aparecem unidas duas coisas muito distintas. Uma, é o fato de que o homem encontra em torno de si essa realidade chamada frio. Outra, que essa realidade o agride, que se apresenta diante dele com um caráter negativo. O que quer dizer, aqui, negativo? Algo bastante claro. Tomemos o caso extremo. O frio é tal que o homem se sente morrendo, isto é, sente que o frio o mata, o aniquila, nega-o. Acontece que o homem não quer morrer, ao contrário, normalmente deseja sobreviver. Estamos tão habituados a ver nos outros e em nós mesmos este desejo de viver, de afirmarmos diante de toda circunstância negativa, que nos custa um pouco nos darmos conta do estranho que é, nos parecendo absurda ou talvez ingênua a pergunta: Por que o homem prefere viver a deixar de ser? Sem dúvida, se trata de uma das perguntas mais justas e sérias que podemos nos fazer. Uma tentativa de responder à questão fala do instinto de conservação. Acontece, que a ideia de instinto é, em si mesma, muito obscura e nada esclarecedora e, ainda que fossa uma ideia clara, é notório que no homem os instintos estão quase apagados, porque o homem não vive, definitivamente, dependente de seus instintos, senão que se governa mediante outras faculdades como a reflexão e a vontade, que agem sobre os instintos. A prova disso é que alguns homens preferem morrer a viver e, pelo motivo que for, anulam em si esse suposto instinto de conservação. (ORTEGA Y GASSET, 1964, p. 319, tradução nossa)

A explicação da vontade de viver do homem, que o leva a produzir tecnologia, não pode se dar apelando ao instinto de conservação deste. Essa explanação é mais complexa, pois envolve uma contradição. Ao lutar contra o mundo, o homem percebeu que poderia construí-lo totalmente de acordo com as suas necessidades. Talvez este seja o motivo principal da vontade de viver do homem, a possibilidade de construir um mundo próprio, totalmente concebido para si.

O segundo procedimento ele apresenta em *El mito del hombre allende la técnica* (ORTEGA Y GASSET, 1965) em que ele narra uma história, com características de um relato da Biologia, sobre o processo evolutivo do homem. Nessa narrativa Ortega Y Gasset descreve a origem humana como a biologia clássica. Os ancestrais do homem viviam nas árvores, estando adaptados a esta forma de vida. Uma modificação, provocada ao acaso pela natureza, na fisiologia desses primatas, talvez uma febre, fez com que os

seus cérebros se tornassem hipertrofiados, aumentando, assim, a sua capacidade de funcionamento. A inserção da tecnologia na vida humana se deve a um fato fortuito da natureza, por algo inesperado.

Como vocês sabem, os animais superiores que precederam ao homem, e isso já foi demonstrado, possuem entendimento, porém não possuem, ou apenas tem memória; o que é o mesmo, não tem fantasia, a qual, igual a memória, é, às vezes, produtiva e outras, improdutiva. Os pequenos chimpanzés, por exemplo, esquecem logo o que lhes aconteceu, mesmo sendo bastante inteligentes; mais ou menos o que acontece com muitos homens quando não dispõem de nenhum conteúdo para o seu entendimento e por isso não podem seguir discorrendo sobre um assunto. Porém, esse animal que se converteu no primeiro homem, encontrou em si mesmo uma enorme riqueza de figuras imaginárias. Estava, naturalmente louco, cheio de fantasia, como não havia acontecido a nenhum animal antes dele, isto significa dizer que diante do mundo que o circundava foi o único que encontrou, em si, um mundo interior. Possui um interior, um dentro, o que os outros animais não podem ter em absoluto. (ORTEGA Y GASSET, 1965, p. 621, tradução nossa)

Essa explicação de Ortega y Gasset se assemelha muito com a explicação, do ponto de vista biológico, que James Trefil (*Somos diferentes? Um cientista explora a inteligência ímpar da mente humana*, 1999) dá para o surgimento da inteligência humana, o que ele chamou de *barganha evolutiva*, isto é, um ganho inesperado e fortuito no desenvolvimento do processo de adaptação biológica.

O fato que, de acordo com a narrativa mítica de Ortega y Gasset, determinou o surgimento da humanidade está vinculado a um acontecimento determinado pelo acaso e que provocou o aumento da capacidade do cérebro de indivíduos da espécie que deu origem ao homem. Um eco da filosofia clássica que reconhece na atividade de pensamento o fator de distinção da nossa espécie em relação a todas as demais.

Da atividade de pensar emerge um novo mundo, um mundo em que a capacidade criadora e imaginativa humana determina pelo seu exercício a alteração do mundo natural. Essa singularidade de possuir uma interioridade, fez com que o homem se encontrasse, “[...] pela primeira vez, ante estes dois projetos totalmente diferentes: ante os instintivos, que ainda estavam presentes nele e, os fantásticos, por isso tinha que escolher e selecionar” (ORTEGA Y GASSET, 1965, p. 622, tradução nossa).

Em ambas explicações, o homem se encontra em confronto com as circunstâncias determinadas pelo ambiente no qual habita. Além disso, ambas possuem como elemento central o fato de a vida estar no entendimento do que ela seja a compreensão das várias questões relativas ao existir do homem, o que fez com que seu pensamento a esse respeito passasse a ser denominado de *racionvitalismo*.

Um ser propenso à autodestruição e o medo da tecnologia

Em uma das muitas cenas antológicas de *2001: Uma odisseia no espaço*, (STANLEY KUBRICK, 1968) um humanoide transforma a queixada de um animal morto, após uma observação empírica perspicaz, numa ferramenta, que o cineasta sugere ser uma arma. Não foi um objeto que não representasse um perigo para outrem. O fogo não deu nenhuma contribuição para esse invento, a não ser a radiação solar que limpou aquela ossada, foi o pensamento o grande vetor dessa transformação ao aplicar um movimento específico a um objeto determinado.

Se nos nossos primórdios e sem a sofisticação da ciência a imagem da destruição é a que se desenvolve na cena humana em que este ainda é uma criança no domínio da sua mais nova capacidade; na era moderna essa imagem adquiriu cores mais intensas.

A tecnologia está de tal forma presente na cultura contemporânea que é praticamente inimaginável um mundo futuro em que nos imaginemos sem ela, pelo contrário, tendemos a pensar que ela sempre se tornará mais sofisticada, abrangente e capaz. Essa presença é verificada nos mais diversos setores da vida humana, até naquilo que julgamos que só os humanos são capazes de realizar, como a atividade política.

No artigo *Global Technology and the Promise of Control*, em que analisa, dentre outras coisas, o processo de disseminação da tecnologia em outras culturas que não a ocidental, Trish Glazebrook (TABACHNICK & KOIVUKOSKI, *Globalization, Technology and Philosophy*, 2004) apresenta uma reflexão crítica em torno do processo de propagação da tecnologia ao considerá-la como uma nova forma de dominação do homem sobre ele mesmo. De acordo com Glazebrook, as tecnologias não estão

desprovidas de um embasamento ideológico e não dar atenção a esta influência faz da tecnologia uma potencial fonte de opressão, “pois sem uma reflexão cuidadosa sobre a ideologia subjacente à tecnologia moderna, o sonho tecnológico de controlar a natureza se derrama em práticas políticas e sociais de exploração” (TABACHNICK & KOIVUKOSKI, 2004, p.143).

Ainda conforme Glazebrook, ao longo da história da reflexão em torno da tecnologia, houve uma modificação profunda da compreensão do que ela é, bem como dos seus objetivos transformando a promessa de liberdade e autonomia do homem em ameaça de perda da própria humanidade. Aquilo que no período clássico era compreendido como a circunscrição do âmbito da técnica é de forma considerável, ao ponto “[...] da distinção entre ciência e tecnologia, tão clara para Aristóteles, seja turvada na modernidade, [sendo] que o projeto tecnocientífico moderno é subscrito por uma lógica de dominação e controle”. (TABACHNICK & KOIVUKOSKI, 2004, p.144)

A tecnologia possui um grande poder de transformação, não somente modificando e aperfeiçoando os métodos e procedimento estritamente tecnológicos. Ela também tem capacidade para alterar as relações entre os homens, bem como o de fazer com que estes revisem seus valores éticos e morais, por exemplo. Mas, por outro lado, ela pode ser veículo de reafirmação daquilo que já está instaurado.

Albert Borgman em *Philosophical Reflections on the Microelectronic Revolution* (MITCHAM, 1986), critica a noção de que a *revolução* provocada pelos computadores e dispositivos microeletrônicos, apesar de ser uma espécie de segunda revolução industrial, apesar de mostrar capacidade de resolução de problemas de difícil tratamento, prometendo disponibilizar, num futuro breve, energia, alimento e saúde em abundância, ela não alterou de maneira significativa as tendências de comportamento dos últimos dois séculos.

Ao mesmo tempo em que a tecnologia microeletrônica se desenvolve para tirar o homem de trabalhos degradantes, ela desenvolve dispositivos que servem meramente para o lazer, a diversão e o passatempo. De acordo com Borgman, esta segunda linha de desenvolvimento será, com o tempo, intensificada, produzindo uma espécie de alienação, um estado de degradação do homem. O resgate da tecnologia dessa condição se daria por

meio de práticas comprometidas com a liberdade do homem e pela compreensão dos princípios que subjazem à ciência e a tecnologia. Somente assim seria possível o exercício da cidadania plena na sociedade tecnológica.

No horizonte do paraíso da tecnologia sempre há uma barra escura que indica uma promessa de perigo. Uma das razões para a apreensão causada pelo desenvolvimento tecnológico é o fato de esta potencializar no homem uma tendência que este possui, a de autodestruir-se.

As repercussões e consequências das aplicações tecnológicas na vida humana passaram a ser vistas com mais atenção após as duas grandes guerras mundiais, porque o crescimento irrefreável da tecnologia deu ao homem um poder de destruição sobre o planeta nunca antes possível a ele.

Lewis Mumford (*El mito de la máquina*, 2010) afirma que

A história toda está cheia de evidências de que muitas e frequentes vezes o curso destrutivo foi seguido, às vezes, no momento preciso em que as energias coletivas do grupo foram aumentadas pelo domínio superior de seus recursos físicos, AL Kroeber apontou que certos parentes remotos do homem, os chimpanzés, quando vistos em liberdade, são deliberadamente destrutivos: "eles gostam de demolir; como as crianças que cresceram sem controle ou educação, tem satisfação em espionar, quebrar, esmagar e, deliberadamente, triturar. Assim que eles têm algo em suas mãos, é raro que eles desistem até diminuí-lo aos seus componentes elementares ».

Kroeber acreditava que tal propensão poderia explicar um dos fenômenos da cultura humana, ou seja, o muito antigo precedente histórico da pedra talhada sobre a polida. (MUMFORD, 2010 p. 94).

Essa tendência do homem de empregar para um fim negativo o desenvolvimento de sua capacidade criadora gera sobre a tecnologia um sentimento de apreensão devido à sua capacidade em aumentar e fortalecer tal tendência. Segundo Arthur Lovejoy (*Reflections on human nature*, 1961), essa tendência é acentuada ou diminuída pela atitude de exaltação ou autocritica, respectivamente.

No livro intitulado *O fantasma da máquina* (1969), Arthur Koestler expõe que, na evolução do homem os poderes de criação e de destruição são constantes. Sua suspeita é de que pode haver na própria constituição do homem algo que o impele a se autodestruir.

Koestler entende que a espécie humana assimilou um tipo de erro evolutivo que se manifesta como uma patologia, a propensão à autodestruição. As hipóteses para a explicação desse comportamento do homem vão desde a doutrina do pecado original até a hipótese do instinto de morte.

A expressão *fantasma da máquina* foi empregada por Gilbert Ryle em *O Conceito de Mente* para criticar o dualismo cartesiano. Nesta obra Ryle defende a posição de que os termos como *crença, desejo, imagem* e outros não possuem sentido por terem origem em equívocos provocados pelo uso da linguagem, o que ele chamou de *erro categorial*. Ryle não concorda com o modelo cartesiano de matriz dualista, porque este postula a existência de duas substâncias que funcionam harmonicamente em paralelo sem conseguir explicar como se dá tal harmonia.

A preocupação com os desdobramentos que podem se dar por meio do desenvolvimento da tecnologia passou a ocupar a agenda não só de cientistas, mas principalmente de políticos e ambientalistas.

Nick Bostrom em *Superintelligence, paths, dangers, strategies* (BOSTROM, 2014) pergunta se o desenvolvimento da tecnologia da inteligência artificial não produzirá uma inteligência superior à do homem, o que poderá colocar a liberdade e autodeterminação humanas em xeque. Por outro lado, Raymond Kurzweil em *A era das máquinas espirituais* (KURZWEIL, 2007) argumenta que a biologia desenvolveu o cérebro humano para produzir um resultado mais rápido do que ela própria seria capaz de conseguir com o seu método de desenvolvimento aleatório. Kurzweil argumenta que o cérebro humano tem a propriedade de acelerar o desenvolvimento do conhecimento e, com isso, acelerar os processos da própria natureza, o que ele chamou de *lei dos retornos acelerados*.

Andy Stirling em *The Precautionary Principles* (BERG OLSEN *et alli*, *A Companion to the Philosophy of Technology*, 2009) aborda um tema que adquiriu maior relevância desde a década de 1970, a relação entre o desenvolvimento científico-tecnológico e a preocupação com a preservação ambiental. Essa área de abrangência do princípio da precaução foi bastante aumentada nas décadas seguintes, estendendo-se em praticamente todas as direções.

De acordo com o Princípio 15 da Declaração do Rio de 1992: “...Onde houver ameaça de dano grave ou irreversível, a falta de certeza científica absoluta não será utilizada como razão para o adiamento de medidas economicamente viáveis para prevenir a degradação ambiental. (BERG OLSEN *et alli*, 2009, p. 248)

A invocação de um princípio como o da precaução mostra que a apreensão do homem com a tecnologia passou a fazer parte da agenda do quotidiano moderno e que a tomada de consciência dos perigos potenciais advindos do desenvolvimento científico e tecnológico são reais e preocupantes.

O Conceito de “inteligência”

Ortega y Gasset em *El mito del hombre allende la técnica* (1965) faz uma reconstrução etimológica do conceito de *inteligência* ao relacioná-lo com o processo de evolução biológica do homem.

O homem terá que ser, desde o princípio, um animal essencialmente *elector*. Os latinos chamavam ao fato de eleger, escolher, selecionar, *eligere*; e ao que o fazia, o chamavam *eligens*, *elegens* ou *elegans*. O *elegans* ou elegante nada mais é do que aquele que escolhe e escolhe bem. Assim, pois, o homem possui de antemão uma determinação elegante, tem que ser elegante. Mas há mais. O latim alertou – como é comum em quase todas as línguas – que depois de um certo tempo a palavra *elegans* e o fato de ser elegante – a elegância - ser elegante – a elegância – haviam perdido algo, por isso era necessário fortalecer a questão e começou-se a dizer *intellegans*, *intellegentia*: inteligente. (ORTEGA Y GASSET, 1965, p. 622, tradução nossa)

Ao comentar esta passagem da obra de Ortega y Gasset, Carl Mitcham (*Que es la filosofía de la tecnología?*, 1989, p. 62), afirma que “esta inteligência dá lugar à insatisfação, a um descontentamento com o mundo; ao desejo de criar um novo mundo e, assim, à técnica”.

O homem é um ser que escolhe, não age de modo meramente instintivo, avalia, analisa, interpreta e depois toma uma decisão sobre como deve proceder. O homem parece não se submeter totalmente à natureza, parece possuir um domínio próprio de existência. A passagem para a condição de animal possuidor de uma inteligência superior é o momento de ruptura do homem com o meio que lhe deu origem.

Quando trata da questão do homem como um animal simbólico, Ernst Cassirer (*Antropologia Filosófica*, s/d) menciona o surgimento da inteligência nos animais e faz uma constatação com referência a esse conceito, afirmando que

[...] a ambiguidade e vagueza do próprio termo “*inteligência*” tem sempre impedido uma solução clara. Como poderemos responder a uma pergunta cujo significado não compreendemos? Metafísicos e cientistas, naturalistas e teólogos têm empregado a palavra *inteligência* com sentidos variados e contraditórios. (CASSIRER, s/d. p. 61, tradução e itálico nossos)

O esforço de definição do que seja a *inteligência* parece semelhante a metafísica do ponto de vista kantiano, não vai além da intenção inicial. Sempre envolvido em posições antagônicas e diferentes pontos de vista, seja sobre como defini-la, seja como reconhecê-la. Contudo, alguns pensadores dedicados a refletir sobre o conceito de *inteligência* não são pessimistas quanto à essa situação de indefinição. É o caso de Jeff Hawkins e Sandra Blackeslee (*Intelligence*, 2005) que acreditam, de acordo com a tradição da crítica da linguagem, que a situação atual tem todas as características que o identificam como um contexto envolto em muita confusão.

Muitos cientistas dizem que é porque o cérebro é bastante complicado, que levará muito tempo até que possamos compreendê-lo. Eu discordo. Complexidade é um sintoma de confusão, não uma causa. Ao invés disso, argumento que temos algumas intuições, mas suposições incorretas que nos enganam. Sendo um dos maiores erros acreditar que a inteligência possa ser definida pelo comportamento inteligente. (HAWKINS & BLAKESLEE, 2005, p. 7)

Essa situação de desacordo é uma constante nos esforços de definição do que seja a *inteligência*. Talvez por estar profundamente relacionado ao conceito de *homem*, a *inteligência* seja um conceito que envolve o mesmo tipo de dificuldade que a definição daquele conceito enfrenta.

O debate em torno da inteligência humana é realizado de forma intensa e polêmica, sendo bastante estimulado pelo posicionamento antagônico das partes nele envolvidas.

Um desses debates, tendo por tema a localização da *inteligência* e envolvendo duas das mais importantes universidades do Reino Unido, Oxford e Cambridge, é descrito por Miguel Nicolelis no livro *Muito além do nosso eu* (2011). De acordo com o relato, essa é só mais uma das grandes discussões envolvendo essas instituições em que uma toma um posicionamento antagônico ao da outra.

No debate realizado em Oxford no ano de 1946 o professor de Cambridge lorde Edgar Douglas Adrian sustenta a posição tomada por seus colegas de instituição no debate com os professores de Oxford, no final do século XVII, de que a inteligência se localiza num órgão do corpo, o cérebro, e não espalhada pelo corpo, como acreditavam os docentes de Oxford.

O argumento apresentado por Edgar Douglas Adrian afirma que o cérebro

[...] é uma estrutura particular de células e fibras nervosas, encontrado em alguns, mas não em todos, os animais. Aqueles seres que não possuem um cérebro no mais estrito senso anatômico, apesar disso são capazes de realizar tarefas complexas, bem adaptadas às circunstâncias, definindo um comportamento que pode ser chamado de inteligência. E dentro de nosso corpo existem muitas células nadando livremente na corrente sanguínea e se comportando como seres vivos mais ou menos independentes, evitando o que lhes é maléfico e se aproveitando do que é bom. Teriam essas células algum direito a reivindicar a posse de uma mente própria? (EDGAR DOUGLAS ADRIAN *apud* NICOLELIS, 2011, p. 60)

O professor de Cambridge está convencido de que o fator primordial que deve ser levado em consideração para a determinação da presença de inteligência num determinado ser é o fato deste possuir um cérebro.

O contexto do debate em torno da definição do conceito de *inteligência* é de bastante controvérsia, ora ela sendo definida como algo oposto à natureza e uma característica peculiar do homem, ora como algo que cumpre inexoravelmente um desígnio natural. Por essas razões esse conceito é cercado de precauções quanto às suas distintas definições devido à falta de clareza quanto ao seu significado.

Destacaremos também, como fizemos com o conceito de homem, dois aspectos desse conceito que são relevantes nessa abordagem que fazemos acerca dos objetivos do projeto de IA. O primeiro é o que estabelece a identificação da *inteligência* com a capacidade de calcular; o segundo, com a capacidade de elaboração de uma obra artística.

A relação entre o cérebro humano e a inteligência

A relação entre o cérebro humano e a inteligência é algo natural no nosso tempo. Uma das teorias que dá suporte à compreensão de que os seres humanos ocupam uma posição de dominância no planeta afirma que isto ocorre devido à capacidade do cérebro humano, muito superior aos das demais espécies. De acordo com esta linha de compreensão, o desenvolvimento da linguagem, da tecnologia e do viver numa sociedade complexa se devem ao cérebro humano. O cérebro é a sede do pensamento e é nele que a faculdade responsável por esta atividade reside.

Em *The mind is not in the head* (<http://cepa.info/575>), Humberto Maturana problematiza o estabelecimento da relação entre o nosso sistema nervoso e o mundo externo por entender que um possui um sistema estrutural determinado distinto do outro. De acordo com ele, isso ocorre, dentre outras coisas, devido a erros conceituais cometidos pelos neurocientistas ao confundirem um campo de domínio de fenômenos com outro.

Através desse erro, confundimos domínios de fenômenos sem intersecção que são vinculados por uma relação generativa e tratados como se um deles estivesse incluído no outro. (MATORANA, <http://cepa.info/575>, tradução nossa)

As pesquisas em torno do funcionamento do cérebro tiveram um grande impulso quando os pesquisadores passaram a empregar os aparelhos de tomografia por emissão de pósitrons (PET) e de formação de imagens por ressonância magnética funcional (fMRI) a partir da metade da década de 80 do século passado. A grande vantagem no estudo do cérebro proporcionada por estes equipamentos foi tornar possível o seu estudo *in vivo*. Nessas condições a observação das funções desempenhadas pelos neurônios e os demais componentes cerebrais se tornou robusta o suficiente para cogitar-se a descrição do seu funcionamento neurônio a neurônio.

Ambas as técnicas dependem da maneira pela qual os neurônios, como todas as outras células do corpo, obtêm suas energias a partir das moléculas transportadas pelo sangue. Quando a célula está desempenhando sua função, seja ela uma célula muscular se contraindo ou um neurônio disparando, ela exige mais energia do que quando está em repouso. O corpo satisfaz essa necessidade pelo aumento do fluxo de sangue e nutrientes para essas células – esta é a razão pela qual seus batimentos cardíacos aumentam quando você faz exercícios.

Da mesma maneira, quando os neurônios em seu cérebro disparam, eles “bombeiam”. O fluxo sanguíneo para a região ativa aumenta, e, embora esse aumento seja pequeno quando comparado àquele observado nos músculos, ele é real e mensurável. A PET e a fMRI são duas técnicas diferentes para medir o aumento do fluxo sanguíneo para regiões do cérebro que estejam sendo usadas. (TREFIL, 1999, p. 85)

Com isso, foram possíveis de ser visualizadas as partes do cérebro em funcionamento ao termos os pensamentos. Essa condição de investigação permitiu averiguar que nem todo o cérebro é empregado quando pensamos algo, dependendo do que pensamos diferentes áreas são ativadas, é isso que as imagens produzidas pela tomografia mostram.

Quando ela se tornou disponível pela primeira vez nos anos 80, a literatura científica subitamente foi tomada por fotografias coloridas de seções do cérebro com diferentes partes realçadas para ilustrar os efeitos de diferentes atividades mentais. As pessoas rapidamente viram que diferentes partes do cérebro estavam envolvidas, por exemplo, na ação de pensar em uma palavra, pensar em dizer uma palavra e efetivamente dizer a palavra. Foi como se uma barreira fundamental ao entendimento do cérebro humano tivesse sido removida. (TREFIL, 1999, p. 86)

No entanto, nem sempre pensamos que os nossos pensamentos são realizados em nossas cabeças. Na Grécia Antiga do período de Homero ainda não havia sido formada a ideia de *corpo* como a que possuímos hoje

[...] o corpo era visto como um aglomerado de membros, representados no ritmo dos seus movimentos, o corpo não é representado através de um termo unitário; pelo contrário, faz-se uso de uma multiplicidade de termos e representações para falar do corpo (Reale, 2002). Assim, o homem homérico empregava os termos *guya* ou *gyia* (se referindo aos membros do corpo em movimento) e *melea* (se referindo aos membros dotados de força muscular). Era empregada a palavras *demas* para designar a estrutura, o tamanho e a semelhança (Snell, 2001). (CASTRO & LANDEIRA-FERNANDEZ, 2011, p. 799).

Também não havia sido formada a ideia de que as atividades de pensamento possuíam uma sede que as reunisse. Parece ser uma característica deste período (séc. VII e VIII a.C.) a falta de unidade das coisas do mundo, os conceitos unificadores dos diversos fenômenos ainda estavam em gestação. Assim como o corpo era visto como uma mera junção de membros que possuíam movimentos independentes, não havendo uma unidade, da mesma forma eram vistas as atividades de pensamento.

Segundo Snell (2001), os gregos homéricos não tinham ainda desenvolvido uma concepção unitária da vida psíquica. Para esse autor, a ausência de uma palavra referente a esta unidade psíquica pode ser considerada como uma evidência da inexistência desse conceito. Assim, não há termos em Homero para estruturas da atividade mental compatíveis com as palavras “alma” ou “mente”, por exemplo. (CASTRO & LANDEIRA-FERNANDEZ, 2011, p. 799)

Entretanto, já aparece no *Corpus Hippocraticum* (séc. V a.C.) a ideia de que algumas doenças mentais e transtornos de comportamento estariam relacionados com o mal funcionamento do cérebro.

No *Corpus Hippocraticum*, o cérebro é apontado como a sede do julgamento, das emoções e de todas as atividades do intelecto, assim como a causa dos transtornos neurológicos, tais como espasmos, convulsões e desordem da inteligência (Hipócrates, trad. 2002; Panourias, Skiadas, Sakas & Marketos, 2005). No Tratado **Dos Ferimentos da Cabeça**, numerosas referências a distúrbios de

movimento são feitas, incluindo vários tipos de paralisia (Hipócrates, trad. 2002, Panourias et al., 2005). Ainda nesse tratado, há referências a distúrbios na fala, além da recomendação do uso de trepanação no tratamento de lesões cranianas (Finger, 2000; Hipócrates, trad. 2002; Panourias et al., 2005). (CASTRO & LANDEIRA-FERNANDEZ, 2011, p. 802)

As sequelas dos campos de batalhas nos tempos de Homero parecem ter dado um grande auxílio na compreensão posterior da relação do cérebro com o resto do corpo. Na obra de Homero são feitas referências aos diferentes tipos de ferimentos de guerra e as suas consequências, mesmo sem haver um vocabulário que auxiliasse na elaboração de uma explicação relacionando as funções cerebrais e sua relação com o funcionamento do corpo.

Na **Ilíada**, observam-se diversas alusões a inúmeros ferimentos (cerca de 141 a 147), localizados nas regiões da cabeça, pescoço, peito, abdômen e membros; ferimentos estes por vezes acompanhados de fraturas ou mesmo esmagamento de ossos, causados por lanças, punhais, flechas ou pedras. Além disso, variam desde aqueles capazes de gerar morte instantânea (como uma flecha que penetra por trás do pescoço) até aqueles ocorridos devido a uma comoção cerebral, como uma pancada à altura do capacete (Crivellato & Ribatti, 2007; Demetrios, 2001; Porto, Moreira & Simão, 2001). (CASTRO & LANDEIRA-FERNANDEZ, 2011, p. 802).

Não há um consenso, entre os filósofos do período clássico, a respeito da localização da atividade de pensar. Para Platão (*Timeu*), a atividade de pensar se localizava na cabeça, que abrigava a alma. A argumentação de Platão para sustentar esta hipótese é a de que a configuração do nosso corpo possui uma hierarquia, os membros localizados na parte superior do corpo são os que realizam as atividades mais elevadas. Assim, para Platão, fazia totalmente sentido crer que a atividade mais elevada que realizamos está no lugar mais elevado dos nossos corpos, a cabeça. A alma (*psique*), entidade que nos dá a capacidade de produzir pensamentos, teria necessariamente de ser abrigada no cérebro.

Já Aristóteles acreditava que a sede do pensamento era o coração. Ele apontava várias razões para isso, o coração produzia calor, era o órgão responsável pela produção

e circulação do sangue, era o dínamo que colocava em movimento todas as demais partes do corpo.

Contudo, já somos advertidos por ele no princípio do *De Anima* de que “[...] em todo caso e de todo modo, é dificílimo obter alguma convicção a respeito da alma”. (ARISTÓTELES, 2006, p. 45). Uma grande incerteza, para não dizer incontornável, sempre rondou os esforços de compreensão das atividades da parte imaterial do homem. A razão que Aristóteles aponta para isso é o fato de não haver um método único de investigação daquilo que é, pois é necessário que um método adequado de estudo seja concebido para o estudo de cada coisa, pois para coisas distintas há princípios distintos, como, por exemplo, para os números e as superfícies.

Hoje o enfoque no estudo da atividade do pensamento está bastante voltado para o desenvolvimento e funcionamento do cérebro humano, desde as suas mais diferentes linhas e áreas de investigação.

James Trefil em *Somos diferentes? Um cientista explora a inteligência ímpar da mente humana* (1999) apresenta o cérebro como um órgão que não foi desenvolvido primordialmente pela natureza para o fim de produzir pensamentos. Para ele, o processo evolutivo da biologia é um processo caótico de alteração dos seres vivos em decorrência de alterações do meio ambiente que, uma vez alterado, produzirá como efeito a suspensão daquilo que não for capaz de suportar as novas condições determinadas por tais alterações.

O mais relevante na exposição de Trefil não está em ele aceitar a tese darwiniana da evolução, mas na ênfase que dá a constatação de que alguns ganhos evolutivos não são advindos diretamente das alterações acontecidas num determinado meio ambiente. Algumas mudanças nos organismos são derivadas do próprio processo de adaptação ao meio ambiente alterado.

Um exemplo de mudanças derivadas das próprias adaptações ocorre com o surgimento das asas dos insetos. De acordo com esta teoria, as asas dos insetos teriam derivado de protuberâncias nos seus dorsos, elas serviriam como aletas de refrigeração. Posteriormente, os indivíduos que possuíam aletas maiores puderam experimentar o voo

planado. As vantagens dos indivíduos com esta capacidade fizeram com que aqueles que possuíssem protuberâncias maiores fossem selecionados ao longo de eras.

Finalmente, é claro, a movimentação desajeitada das aletas provavelmente teria invalidado qualquer vantagem em tê-las de tamanho maior. Todavia, mais ou menos neste ponto, as aletas já são grandes o bastante para permitir que o inseto plane. De repente, há um nicho ecológico inteiramente novo disponível ao inseto. Em vez de rastejar sobre uma única árvore, ele agora pode planar de árvore em árvore em busca de alimento e escapar de predadores. Como consequência, o que originalmente era uma aleta de refrigeração desempenha agora uma função inteiramente diferente – que permite a um inseto realizar voos rudimentares. Uma vez ultrapassado esse limiar, o desenvolvimento de asas completas não é difícil de imaginar. Esse processo, no qual um determinado órgão primeiro é útil para um determinado propósito e depois serve para outro, ocorre repetidamente na história da evolução. Eu o denomino “barganha evolutiva”. (TREFIL, 1999, p. 99)

Com o desenvolvimento do nosso cérebro ocorreu algo semelhante. Ele não foi adaptado para ser gerador de conhecimento, ele na verdade sofreu mudanças porque outros órgãos e membros dos *Australopithecus* sofreram alterações que, por sua vez, produziram mudanças na fisiologia e morfologia dos seus cérebros.

Mas, como ocorreu na conversão das aletas de refrigeração em asas, uma vez que essa mudança tenha tido lugar, novas possibilidades se abriram para a seleção natural. O palco estava armado para outra “barganha” evolutiva. Os cientistas afirmam que a evolução da inteligência humana, como a evolução do voo dos insetos, pode muito bem ser exemplo de um benefício aparentemente não-intencional em uma área em decorrência de desenvolvimento em outra. (TREFIL, 1999, p. 104)

No processo de adaptação evolutiva do cérebro humano houve um aumento na sua capacidade de entender e compreender o funcionamento das coisas. Contudo, isto não o livrou de enfrentar dificuldades e obstáculos impostos pelo seu próprio funcionamento quanto a esses objetivos. As críticas de Oliver Sacks e Vilayaun Ramachandran apontam para um limite do cérebro humano ao estudar a si próprio, o que passou a ser chamado de *chauvinismo hemisférico*.

Uma razão importante da desatenção para com o hemisfério direito, ou “não dominante”, como sempre foi chamado, é que, enquanto é fácil demonstrar os efeitos de várias lesões localizadas no hemisfério esquerdo, as síndromes correspondentes do hemisfério direito são muito menos distintas. Presumia-se, em geral, depreciativamente, que ele fosse mais “primitivo” que o esquerdo, sendo este último visto como a flor única da evolução humana. E, em certo sentido, isso é correto: o hemisfério esquerdo é mais complexo e especializado, um produto muito recente do cérebro primata e especialmente hominídeo. Por outro lado, é o hemisfério direito quem controla as faculdades cruciais de reconhecimento da realidade que todo ser vivo precisa possuir para sobreviver. O hemisfério esquerdo, como um computador anexado ao cérebro básico da criatura, é projetado para programas e esquematizações; e a neurologia clássica se interessa mais por esquematizações do que pela realidade, de modo que, quando finalmente emergiram algumas das síndromes do hemisfério direito, foram consideradas bizarras. (OLIVER SACKS *apud* CHRISTIAN, 2013, p. 79)

[...]

O hemisfério esquerdo é especializado não só na produção efetiva dos sons da fala, mas também na imposição da estrutura sintática na fala e em boa parte do que se denomina semântica, a compreensão do sentido. O hemisfério direito, por sua vez, não controla as palavras faladas, mas parece ocupar-se de aspectos mais sutis da linguagem, como as nuances da metáfora, alegoria e ambigüidade – habilidades que não são enfatizadas de modo adequada em nossas escolas elementares, mas são vitais para o avanço das civilizações através da poesia, mito e drama. Tendemos a classificar o hemisfério esquerdo como o hemisfério principal ou “dominante” porque ele, como um chauvinista, se encarrega totalmente da fala (talvez também de boa parte do pensamento interno), reivindicando a posição de repositório do mais elevado atributo da humanidade, a linguagem. (VILAYAUN RAMACHANDRAN *apud* CHRISTIAN, 2013, p. 80)

Portanto, há um cenário bastante problemático no tocante ao estudo do cérebro humano, pois o seu funcionamento produz influência sobre o seu próprio estudo. Isso produz um desconforto epistêmico, já que temos razões para duvidar daquilo que empregamos como elemento primordial na busca pelo conhecimento. O próprio cérebro pode nos induzir a engano do mesmo modo que a experiência dos sentidos.

As esquematizações como forma de funcionamento predominante no hemisfério esquerdo acabam por impor um modelo de funcionamento que se aplica ao cérebro, ao mundo, às coisas. É necessário que se investigue qual o *status* desta relação, pois pode ser que estejamos diante de uma determinação incontornável.

Algumas definições de ‘inteligência’

O ‘saber calcular’

Apontada como uma das principais atividades humanas que atestam o caráter especial do homem é a nossa capacidade de saber calcular.

Roger Penrose (*O grande, o pequeno e a mente humana*, 1998) procura explicar o surgimento da mente como o resultado da relação de três diferentes mundos, o *físico*, o *cultural* e o *mental*. Segundo o seu pensamento, o mundo mental não pode ter surgido do mundo físico pura e simplesmente, pois

[...] os conceitos que tratamos na física são matéria, objetos massivos, partículas, espaço, tempo, energia, coisas físicas e demais. Como pode ser que nossas sensações, nossa percepção do vermelho ou da felicidade tenham algo a ver com a física? Creio que isso é um mistério. (PENROSE, 1999, p. 580, tradução nossa)

Ao examinar a relação entre a matemática e a física, Penrose faz uma observação, a de que o mundo físico parece seguir as leis da matemática de forma bastante precisa. Ele considera este mais um mistério. O mundo físico é descrito com conceitos da matemática, conceitos que são considerados construções mentais. Mas, segundo Penrose, essa não é uma forma adequada de se abordar a questão, pois, embora os mundos físico, cultural e mental estejam relacionados entre si, isto não significa dizer que um surja do outro.

Apesar de adotar um ponto de vista da física, já que para ele a *consciência* é um fenômeno que merece uma tentativa de explicação por parte da ciência, isso não significa dizer que é totalmente possível de ser colocada num computador qualquer descrição matemática do mundo feita pela ciência.

Penrose distingue duas perspectivas antagônicas de tratamento dessa questão: de um lado a suposição de que as leis da física possuem um caráter determinista e não podem ser consideradas os elementos responsáveis pela emergência do mental; de outro, a suposição de que

[...] somos simples computadores; pode ser que ainda não saibamos como descrever tais coisas mas, se de algum modo soubéssemos o tipo certo de cálculos que haveríamos de realizar, seríamos capazes de descrever todas as coisas mentais enumeradas na Figura 3.4. A palavra emergência é muitas vezes empregada para descrever esse processo. Essas qualidades *emergem*, segundo essas pessoas, como o resultado do tipo certo de atividade computacional. (PENROSE, 1999, p. 60, tradução nossa)

Embora não se julgue em condições de definir o que seja a *consciência*, ele destaca dois aspectos inerentes a esta, o que ele denominou de suas manifestações, passivas e manifestações ativas. A primeira diz respeito ao aspecto de receptividade da consciência e a segunda evidencia o caráter de liberdade e autodeterminação desta. Essas manifestações revelam um caráter ambivalente da *consciência*. Ela se relaciona, ao mesmo tempo, tanto àquilo que é determinação e necessidade, quanto com o que é livre e indeterminado. Penrose introduz essa distinção para apresentar um terceiro tipo de manifestação da consciência que não possui nem o caráter ativo, nem o caráter passivo. É esse terceiro elemento que lhe permitirá dizer alguma coisa sobre a *consciência*, o que ele chama, sem saber exatamente o que significa, de *entendimento* ou *intuição*.

Me concentrarei, principalmente, naquilo que implica a consciência de um modo essencial. É diferente dos aspectos passivo e ativo da consciência e, talvez, ocupe um lugar entre ambos. Me refiro ao uso do termo *compreensão*, ou talvez *intuição direta*, uma palavra muito mais adequada. Não definirei nenhum destes termos; não sei o que significam. Existem outros dois termos que, também, não entendo: *conhecimento* e *inteligência*. Então, por que estou falando sobre coisas que não sei realmente o que elas significam? Provavelmente se deva ao fato de que sou um matemático e, os matemáticos, não pensam muito sobre os conceitos. Não necessitamos de definições precisas dos elementos com os quais estamos trabalhando, contanto que possamos dizer alguma coisa sobre as relações entre eles. O primeiro ponto chave é que, para mim, a inteligência é algo que requer compreensão. Empregar o termo inteligência num contexto em que negamos que esteja presente qualquer compreensão me parece pouco razoável. De modo análogo, compreender sem nenhum conhecimento é, também, algo absurdo: compreender requer algum tipo de conhecimento. Este é o segundo ponto chave. Disso se deduz que a inteligência requer conhecimento. Mesmo que não esteja definindo nenhum desses termos,

me parece razoável insistir sobre essas relações entre eles. (PENROSE, 1999, p. 60, tradução nossa)

Sua abordagem da questão não tem como ponto de partida o significado dos conceitos que investiga. Seu objetivo é de natureza matemática, ele pretende investigar as relações que existem entre aqueles conceitos. Nesse sentido, se quisermos compreender o que é a *consciência*, é pertinente que seja analisada a relação entre o *entendimento* e a *inteligência*, dado que ambos demonstram possuir em comum o conhecimento. Além disso, para ele não faz sentido pensar a *inteligência* sem o *entendimento*, nem o *entendimento* sem o *conhecimento*.

Assim, existem, na sua opinião, várias abordagens da *receptividade*:

O ponto de vista que chamo de **A**, às vezes chamado de *inteligência artificial forte* (IA forte) ou *funcionalismo* (computacional), afirma que todo pensamento é simplesmente a realização de certos cálculos e, em consequência, se forem realizados os cálculos adequados, o resultado será o conhecimento.

Rotulei como **B** o segundo ponto de vista, segundo este princípio, seria possível simular a atividade de um cérebro quando o seu dono tivesse conhecimento de algo. A diferença entre **A** e **B** é que, ainda que dita atividade pudesse ser simulada, essa mera simulação não poderia em si mesma, segundo **B**, ter nenhum sentimento ou conhecimento – há algo mais, que, talvez, tenha a ver com a construção física do objeto. Por isso, um cérebro composto de neurônios e equivalentes pode ter conhecimento, enquanto uma simulação da atividade do dito cérebro não teria conhecimento.

Eu sou um defensor do ponto de vista de **C**. Existem, no entanto, várias variedades de **C**. É o que poderíamos chamar de *C fraco* e *C forte*. **C** propõe que, de alguma forma, com física a física atual bastaria olharmos com um pouco mais de cuidado para encontrarmos certos tipos de atividade que estão para além da computação. Quando digo *para além da computação*, serei mais explícito logo a seguir. De acordo com o *C fraco*, não há nada que necessitemos procurar fora da física atual para encontrar a atividade não computacional devida. O *C forte*, por outro lado, exige que algo exista fora da física conhecida; nossa compreensão física não é adequada para a descrição do conhecimento. É incompleta e, como você deve ter deduzido, acredito que nossa imagem física é incompleta, como indiquei na figura 2.17. Do ponto de vista da *C forte*, talvez a ciência, no futuro, seja capaz de explicar a natureza da consciência, coisa que a ciência atual não faz.

Finalmente, existe ainda o ponto de vista **D**, segundo o qual é um erro considerar essas questões nos termos de ciência. Talvez o conhecimento não possa ser explicado em termos científicos. (PENROSE, 1999, p. 61, tradução nossa)

As possibilidades de tratar da questão do *conhecimento* vão desde a que acredita que o pensamento pode ser totalmente executado num computador até a que afirma o diametralmente oposto, de que esse objetivo está fora dos domínios da ciência, passando pelas perspectivas intermediárias que aceitam uma ou outra em parte. Para justificar o seu ponto de vista, Penrose responderá à questão, se existe algo *para além* da computação presente em nossa compreensão. Penrose menciona o exemplo do computador que adotou uma estratégia no jogo de xadrez totalmente equivocada, porque a sua forma de proceder, empregando a realização de cálculos computacionais, não lhe possibilitou vislumbrar as jogadas para além daquela em jogo, pelo menos não um grande número de jogadas seguintes. Isso demonstraria que a execução de um procedimento computacional não pode ser entendida como uma ação acompanhada de compreensão.

A questão é que o xadrez é um jogo computacional. Neste caso, o jogador humano vê a barreira de peões e comprehende que ela é impenetrável. O computador não possuía essa compreensão, simplesmente computava um movimento após o outro. Assim, este exemplo ilustra a diferença entre a mera computação e a qualidade da compreensão. (PENROSE, 1999, p. 62, tradução nossa)

O fato de ser capaz de realizar operações que lhe permitam efetivar uma escolha ou deliberação não significa dizer que o computador possua a mesma capacidade de escolha que um ser humano, pois este faz escolhas que parecem não estar baseadas em operações do tipo computacional.

A fim de dar mais robustez a seu argumento, ele imagina uma máquina de Turing com uma capacidade de memória e de processamento ilimitados. Seu objetivo é mostrar que um cálculo computacional não consiste meramente na realização de operações matemáticas, mas também de operações lógicas.

Para demonstrar o que seja uma computação, ele apresenta quatro tipos de tarefas dadas a um computador para realizar. A primeira tarefa o computador tem como chegar a uma conclusão ou término das operações, nas demais não. As tarefas que consistiriam na realização de operações computacionais que não tem como chegar a um término, ele classificou em três diferentes níveis: fácil, difícil, dificílimo.

Esse tipo de operação computacional é empregado por Penrose como um indício para sustentar que algumas operações envolvendo a realização de cálculos computacionais não podem ser levadas à cabo.

Eis um exemplo mais complicado: encontrar um número par maior que 2 que não seja a soma de dois números primos. Essa computação será concluída em algum momento? A crença geral é que não o conseguirá, porém isto é uma mera conjectura, conhecida como a *conjectura de Goldbach* e é tão difícil que ninguém sabe com certeza se é verdadeira ou não. Assim, pois, há (provavelmente) três cálculos que não terminam: um fácil, um difícil e um terceiro que é tão difícil que ninguém sabe, todavia, se realmente se ele chega a um término ou não.

Coloquemos agora a seguinte pergunta: os matemáticos estão empregando algum algoritmo computacional (vamos chama-lo de A) para convencer-se de que certas computações não terminam? (PENROSE, 1999, p. 64, tradução nossa)

A forma especulativa de funcionamento do pensamento humano leva Penrose a suspeitar de que esta forma de procedimento não pode se fundamentar na realização de operações de natureza estritamente *computacional*, pois esse tipo de operação não é capaz de produzir como resultado uma *compreensão*. Os computadores, apesar de realizarem operações complexas, não podem ser consideradas máquinas inteligentes assim como uma roda d'água que coloca em funcionamento uma serraria não é inteligente, embora seja capaz de transformar troncos de árvores em tábuas.

Sua argumentação nos leva a crer que a atividade humana de pensamento possui um componente que não se faz presente nos computadores quando estes executam seus programas. De acordo com Penrose, os seres humanos possuem a capacidade de *convencerem-se a si mesmos* de alguma coisa, e isso não se deve ao exercício da

capacidade de execução de um algoritmo. Um exemplo disso seriam as demonstrações de um tipo especial de proposição da matemática chamada π_1 .

Turing e Gödel possuem opiniões distintas a respeito do que torna possível aos matemáticos terminarem um cálculo quando temos razões para acreditar que tal cálculo é interminável? Para Turing, a parada seria nada mais do que a execução de um algoritmo errôneo, o seguimento de uma regra que não deveria ser seguida; enquanto Gödel não acreditava que a intuição matemática fosse redutível à computação, pelo menos não por um algoritmo acerca do qual pudéssemos ter certeza da sua validade. Se para Turing, o procedimento de parada é devido ao exercício de um algoritmo errôneo, a execução de uma *operação computacional frouxa*; para Gödel, ela é produzida por um algoritmo cuja validade não temos como aferir. Penrose não concorda com nenhuma das duas posições, já que para ele a inteligência não é o exercício da mera capacidade de se seguir regras. Além do ser humano ser capaz de seguir regras, ele também *toma conhecimento* das coisas.

Se, na elaboração da demonstração de uma proposição matemática não estão envolvidas somente atividades computacionais, como então, matemáticos como Gödel foram capazes de concluir coisas acerca das quais não temos como sustentar, por meio de uma demonstração matemática, a sua validade?

Os seres humanos teriam uma capacidade de preencher uma lacuna que não é possível de ser preenchida pelo exercício simples de se seguir regras.

Apesar do fato de que não há nenhuma forma computacional de caracterizar os números naturais, qualquer criança sabe o que são. Tudo que fazemos é mostrar-lhe diferentes quantidades de objetos, como ilustra a Figura 3.9, e depois de algum tempo, poderão abstrair a noção de número natural a partir desses exemplos concretos. Não damos à criança um conjunto de regras computacionais, o que estamos fazendo é capacitar à criança a compreender o que são os números naturais. Eu diria que a criança é capaz de estabelecer algum tipo de contato com o mundo platônico das matemáticas. (PENROSE, 1999, p. 68, tradução nossa)

Ele recorre a um argumento epistêmico-metafísico para sustentar essa última afirmação dizendo que os números naturais, de certa maneira, estão aí desde sempre e que temos acesso a eles porque possuímos a capacidade de conhecer as coisas.

Se fôssemos simples computadores eletrônicos, não teríamos dita capacidade. As regras não são fatores que nos capacitam a compreender a natureza dos números naturais, como mostra o teorema de Gödel. Compreender o que são os números naturais é um bom exemplo do contato com o mundo platônico. (PENROSE, 1999, p. 68, tradução nossa)

Ele sugere que a explicação para essa capacidade do homem esteja na estrutura do cérebro que é constituído por células que realizam atividades bastante complexas. Mais especificamente, são determinadas microestruturas que estão presentes nos neurônios (*microtúbulos*) os prováveis responsáveis pela capacidade humana de terminar um cálculo interminável. Essas microestruturas realizam, “provavelmente, muitas coisas, incluindo o transporte de moléculas neurotransmissoras dentro da célula, porém uma atividade na qual parecem estar envolvidos é na determinação das intensidades das sinapses”. (PENROSE, 1999, p.75, tradução nossa)

O neurônio é uma estrutura bastante complexa e cujo fluxo de entrada e saída de informação envolve um processo que necessita, na opinião de Penrose, do desenvolvimento de um novo modelo de conhecimento.

Como conclusão, Penrose acredita que somente através do desenvolvimento de atividades de natureza não-computável seremos capazes de estabelecer um nexo entre as teorias da física clássica e da física quântica. E, talvez com isso, sejamos capazes de explicar fenômenos que parecem não fazer sentido.

O ‘saber elaborar uma obra artística’

No segundo capítulo dessa tese, quando tratamos do procedimento de Turing para reconhecer inteligência nas máquinas, mencionamos a objeção feita por Geoffrey Jefferson a essa estratégia de reconhecimento.

O argumento de Jefferson está na conferência que fez quando recebeu o prêmio *Lister Oration*, premiação concedida pelo *Royal College of Surgeons of England* de 1949, intitulada *The mind of mechanical man*.

Ele inicia sua conferência afirmando que um dos objetivos de conhecimento mais almejados atualmente pelo homem é o de compreender a relação entre o cérebro e a mente: *um, finito; a outra, inexata e furtiva* (JEFFERSON, 1949, p. 1105). Pressionado pela matemática e pela física, o debate em torno da relação *cérebro/mente* foi bastante influenciado pela comparação entre os computadores e o sistema nervoso humano. Embora visse como positiva a aproximação dos físicos e matemáticos desse debate, Jefferson não deixou de advertir que é sempre necessária uma certa precaução, a de não ir além daquilo que estão em acordo, para evitar cair nos erros induzidos pelo entusiasmo científico. Ele conclui que a relação *cérebro/mente* sempre envolverá uma discussão arriscada.

Após ilustrar fartamente que o sonho humano de construir autômatos é muito antigo e presente em diferentes lugares e épocas, perguntar se, assim como artesãos, relojoeiros e outros, construíram, no passado, máquinas admiráveis que conseguiam enganar uma plateia esclarecida, os conhecimentos científico e tecnológico atuais não seriam capazes de construir algo que não fosse vivo, mas que parecesse ser, um simulacro que fosse bastante convincente? Como fazer, então, para distinguir um simulacro convincente de algo genuíno?

Jefferson se apoia no pensamento de Descartes para sustentar que um autômato pode até fazer algumas coisas melhores do que um homem, mas não poderia ser programado para responder a todos os estímulos que um homem é capaz de responder, porque não possui uma mente, sua ação seria um mero encadeamento de movimentos das partes que o constituem.

Muitos são os exemplos que mostram a obra extraordinária que é um organismo vivo, sua eficiência energética, sua capacidade de se manter em funcionamento. Mesmo que o cérebro venha a ser copiado e imitado pelo conhecimento humano, ele permanecerá único, pois realiza atividades que não podem ser explicadas levando-se em consideração somente os fenômenos físicos.

Descartes resolveu a dificuldade tornando a mente sobrenatural, propondo uma mente imaterial independente do organismo na pineal. Este era o antigo refúgio daqueles que enfrentaram o inexplicável na natureza, como ainda vemos nos povos primitivos e supersticiosos. Hoje podemos seguramente duvidar que um agente sobrenatural na base dos processos mentais. (JEFFERSON, 1949, p. 1106)

A grande diversidade de comportamento que um ser vivo é capaz de desempenhar torna difícil a reprodução artificial do mesmo. Não é porque há um elemento de comparação entre o computador e o sistema nervoso, ambos empregam eletricidade nos seus funcionamentos, que pode ser afirmado que são semelhantes e possuem as mesmas capacidades. Para ele, o hiato que existe entre as ações desempenhadas mecanicamente e as ações que demandam a necessidade do emprego do pensamento só seriam respondidas se as máquinas conseguissem compor um soneto ou um concerto musical. De acordo com esse pensamento, a produção de uma obra estética só é possível porque o homem possui uma vida interior que necessita de expressividade, não são os processos eletroquímicos que produzem um soneto ou uma ópera.

Para Jefferson, as máquinas estão destituídas da capacidade criadora, da capacidade de agir e refletir sobre este agir. Da capacidade de elaborar algo cuja necessidade de expressão se torne uma manifestação de liberdade. Esta capacidade do homem está envolta em mistério, pois ela parece não estar submetida às leis da natureza e as suas criações não encontram, no mundo natural, nenhum elemento que lhes seja correspondente. Ele via no fazer artístico algo que torna o homem um ser especial e que uma das grandes dificuldades a ser superada é a atitude dos físicos que não dão a devida relevância à complexidade e riqueza do sistema nervoso humano.

Por outro lado, ao abordar a relação entre a arte e neurologia, Ramachandran e Hirstein (*The Science of art*, 1999) desenvolveram uma teoria em que os componentes

neurais cumprem papel relevante na experiência estética. Ainda que muito do que é artístico seja considerado, por eles, como uma manifestação idiossincrática de artistas, tornando quase impossível uma explicação dessa atividade humana. Eles julgam que a maior parte do que é considerado arte se deve a aplicação de princípios estéticos subjacentes a ela. Assim, a tarefa do artista consistiria em dois momentos, a captura daquilo que é essencial em algo e a amplificação dessa essência por meio da obra de arte, cuja função seria a de ativar de forma vigorosa os mecanismos neurais implicados numa experiência estética. Assim, a experiência estética seria um tipo de atividade que visaria, em última instância, a ativação de determinada área do cérebro.

O conceito de ‘artifício’

A princípio, a distinção entre o que é considerado *natural* e o que é *artificial* é bastante óbvia, simples, fácil de verificar e definir. *Natural* é definido como sendo tudo aquilo que existe independente da vontade e ação humanas, enquanto *artificial* seria exatamente o oposto, isto é, tudo o que existe em razão da capacidade e atividade humanas.

Em *As ciências do Artificial* (1981), Herbert Simon afirma que uma das razões que nos leva a acreditar que haja uma diferença entre o natural e o artificial é que “[...] pela sua subserviência às leis da natureza os fenômenos naturais se rodeiam de uma aparência de “necessidade”, os fenômenos artificiais aparecem “contingentes”, moldáveis pelo ambiente”. (SIMON, 1981, p. 13)

A preocupação de Simon está relacionada à tarefa de elaboração de uma teoria que fosse capaz de dar conta das alterações e correções de rumo necessários ao seu desenvolvimento, como poderá ocorrer com uma *ciência do artificial*, uma vez que tal ciência não se debruçaria sobre o necessário, mas sobre o contingente “[...] não [trata] de como as coisas são, mas de como podem ser – em resumo, da projeção”. (SIMON, 1981, p. 15)

Ele busca responder à questão, se é possível propor a criação de uma ciência artificial, uma ciência que seja capaz de tratar com o problema da contingência. Para ele é necessário, antes, que seja superado um preconceito que recai sobre a noção de *artificial*.

O meu dicionário define *artificial* como: “produzido mais pela arte que pela natureza; não genuíno ou natural; afetado; não pertence à essência do objeto”. Propõe como sinônimos: afetado, factício, manufaturado, fingido, simulado, falsificado, forjado, não-natural; como antônimos: real, genuíno, honesto, natural, verdadeiro, sincero. A nossa linguagem parece refletir a profunda desconfiança do homem pelos seus próprios produtos. Não tentarei verificar a validade de tais interpretações ou explorar as suas possíveis raízes psicológicas: o leitor compreenderá que uso o termo “*artificial*” de forma tão neutra quanto possível, significando “fabricado pelo homem por oposição a natural. (SIMON, 1981, p. 25)

A neutralidade de Simon no uso de *artificial* é criticada por SYTSE STRIJBOS (*Computer and world picture. In: Broad and Narrow interpretations of philosophy of technology*. PITT, J.C. & DURBIN, P.T., 1990) por considerá-la uma expressão do comprometimento da abordagem de Simon com uma imagem técnica do mundo. Em outras palavras, Simon acredita que a fronteira entre o natural e o artificial não é tão bem demarcada como pressupõe a definição do dicionário. Ele defende que é possível através do emprego da simulação alcançarmos compreensão, ou seja, para ele o computador é uma ferramenta útil para a compreensão da própria atividade humana de pensamento, um computador poderia realizar praticamente a mesma coisa que um ser humano faz ao pensar, só que em outra base.

A imitação é possível por poderem ser organizados sistemas físicos distintos que exibem comportamento quase idênticos. A mola amortecida e o circuito amortecido obedecem à mesma equação diferencial linear de segunda ordem; logo, podemos usar qualquer desses sistemas para imitar o outro. (SIMON, 1981, p. 40)

Simon defende o emprego de um procedimento experimental na investigação da fronteira entre o natural e o artificial, entre o funcionamento da mente humana e o de um computador. Essa forma de tratamento da questão é duramente criticada por Strijbos por

não considerar esta metodologia um modo adequado de demarcação da fronteira entre o humano e a máquina, entre o *natural* e o *artificial*.

Simon é, como ele diz, um radical tecnológico e um filósofo pragmático. Aqui devemos perguntar de forma crítica se a prática pode, como ele acredita, servir de árbitro entre as posições extremas do conservadorismo filosófico e o radicalismo. Em outras palavras, pode o desenvolvimento atual em curso da tecnologia da informática decidir se existe ou não uma fronteira estrita entre as possibilidades humanas e as do computador? Simon continua na suposição incorreta de que a prática é uma caso independente. Prática, neste caso o desenvolvimento da tecnologia do computador, sempre é avaliada a partir de uma perspectiva particular e desenvolvida à luz dessa mesma perspectiva. Se o radical acredita que os seres humanos são únicos e que os computadores podem, em princípio, fazer tudo o que os seres humanos fazem, então a prática não será capaz de refutá-lo. Quando o computador não puder fazer certas coisas, a defesa será sempre a de que a evolução futura irá resolver o problema.(PITT & DURBIN, 1990, p.83, tradução nossa)

A crítica de Strijbos afirma que o comprometimento da posição de Simon não é somente com uma determinada visão filosófica de mundo. Aquilo que ele considera uma visão defeituosa de Simon produz implicações de natureza social e política. O pensamento de Simon, que estabelece uma identidade entre o homem e a máquina, serve para justificar uma determinada compreensão de liberdade, organização e autoridade e, consequentemente de mundo e sociedade. Ela traz consigo consequências concepção filosófica do mundo, seu comprometimento é de natureza política da concepção de Simon

A visão defeituosa que Simon possui do homem traz consequências para a prática, ela se adapta à imagem técnica do mundo. A visão antropomórfica do computador e a visão computadoromórfica da humanidade não deixam de se refletir no processo de tecnicização. Considere novamente a visão de Simon de liberdade e responsabilidade numa organização. Simon defende uma visão funcionalista da autoridade. Autoridade é uma função necessária da organização, de modo que alguns tenham o poder de tomar decisões enquanto outros devem obedecer. De acordo com essa concepção, certa desigualdade de poder é essencial para uma relação de autoridade. Autoridade implica uma relação hierárquica entre duas pessoas, uma relação de chefe e subordinado. (PITT & DURBIN, 1990, p.84, tradução nossa)

De acordo com Strijbos, Simon não era muito simpático à crítica de que a autoridade é algo que desumaniza o ser humano por não se preocupar com os objetivos de autorrealização dos indivíduos, porque para ele a prática mostra o contrário. Strijbos considera que a visão do mundo artificial expressada por Simon acredita que o pensamento humano é um sistema artificial semelhante ao computador.

A distinção entre *natural* e *artificial* envolve um grande desafio para aqueles que, no exercício da atividade de historiadores das ideias, buscam compreender suas razões e significados, o que levou Pierre Hadot a afirmar que *escrever a história do pensamento é escrever a história de mal-entendidos* (*The veil of Isis*, 2006, p. 17). Para mostrar que a distinção entre *natural* e *artificial* também é uma história de mal-entendidos, ele faz referência às diferentes interpretações dadas ao aforismo de Heráclito aos diferentes significados atribuídos

De acordo com Hadot, a origem da distinção entre *natural* e *artificial* já encontramos na Grécia antiga. Sua abordagem do aforismo de Heráclito demarca o início da sua reflexão sobre o conceito de *natureza* e afirma que este possui duas interpretações: uma original e outra escolar, quando passou a ser ensinado nas escolas gregas. Três termos são essenciais na compreensão desses significados: *philein*, *physis* e *kryptein*.

Quanto ao emprego do termo *philein*, ele afirma possuir dois significados, expressando tanto um sentimento quanto uma tendência natural das coisas.

Em Heródoto, por exemplo, “O vento ama (está acostumado) a soprar”, ou em Demócrito, “Graças a esses exercícios (os esforços para aprender a ler e música), o respeito “ama” desenvolver-se (é habitualmente desenvolvido). Aqui a palavra “ama” (*philei*) denota não um sentimento, mas uma tendência natural ou habitual, um processo que ocorre necessariamente ou com frequência. (HADOT, 2006, p. 7, tradução nossa)

O termo *physis*, por sua vez, também possui um duplo significado, pode expressar tanto a ideia de nascimento e crescimento quanto a de um tipo de ideal personificado, uma aspiração.

Não é possível escrever detalhadamente a história de sua interpretação, contudo gostaria de me concentrar principalmente em mostrar como a

palavra *physis* primeiro designou um processo de crescimento e finalmente, veio a significar uma espécie de ideal personificado. (HADOT, 2006, p. 17, tradução nossa)

Por sua vez, o termo *kryptein* também possui um duplo significado, o primeiro expressa a ideia de que algo é ocultado, por ser difícil de compreender ou porque deve ser escondido dos outros; o segundo expressa desaparecimento e morte.

No entanto, de uma maneira muito interessante para a nossa reflexão, os versos de Sófocles colocam em oposição os verbos *philein* e *kryptein*, nos encorajando a manter as duas interpretações que propus, isto é, no sentido ativo: “O que faz com que as coisas apareçam, tende a fazê-las desaparecer”, “O que causa o nascimento tende a causar a morte” ou, “O que causa o nascimento tende a causar a morte”, ou “O que revela é também o que esconde”, e no sentido passivo, “o que aparece tende a desaparecer” ou “O que nasce necessita morrer”. (HADOT, 2006, p. 11, tradução nossa)

O sentido de *natureza* está relacionado tanto a algo que pode ser a expressão de um desejo, vontade, de um agir tendo em vista um fim específico, quanto de uma lei natural no sentido estrito, isto é, um acontecer que não é determinado por nenhuma finalidade previamente deliberada. A interpretação de Hadot nos permite cogitar que natural é um predicado que também pode ser aplicado à ação humana.

Lynne Rudder Baker (1944) em *Ontology of Artifacts* trata da distinção entre os objetos naturais e os artefatos empregando a noção de *constituição*. Sua abordagem do problema da distinção entre os objetos naturais e os artefatos é crítica em relação ao modo como tradicionalmente esse tema foi tratado pela filosofia, [...] *quero defender o status ontológico de artefatos contra filósofos – de Aristóteles a Leibniz até Peter van Inwagen – que consideram os artefatos deficientes ontologicamente.* (BAKER, 2004, p. 1, tradução nossa)

A chamada *visão constitucional* adotada por Baker para tratar deste problema, não acredita que a identidade de alguém possa ser estabelecida pelo seu corpo ou pelo conteúdo da sua memória, pois tanto um quanto o outro são alterados constantemente. A *constituição* de algo não é o mesmo que a sua identidade, enquanto a identidade se remete à noção de necessidade, a constituição está sob o escopo da contingência.

Constituição, eu acredito que seja a cola do mundo. *Constituição* é uma relação muito geral que todos nós conhecemos (embora provavelmente não sob esse rótulo). Um rio, em qualquer momento, é constituído por um agregado de moléculas de água. Mas o rio não é idêntico ao agregado de moléculas de água que o constitui naquele momento. Uma vez que o mesmo rio - chamado "R" - é constituído por diferentes agregados de moléculas em diferentes momentos, ele não é idêntico a nenhum dos agregados de moléculas de água que o compõem. Então, a constituição não é identidade. (BAKER, 2004, p. 1, tradução e itálicos nossos)

Na visão de Baker a *constituição* relaciona dois tipos de propriedades constitutivas de um objeto ou artefato, ao que ela chamou de propriedade *não-derivada* e propriedade *derivada*. Dito de outro modo, um objeto ou artefato possui elementos essenciais e elementos contingentes, elementos que fazem com que ele seja identificado como tal e não como uma coisa qualquer e elementos que tornam possível a sua materialização, seja com que material for. O que faz com que um determinado objeto seja um objeto de determinado tipo não é a matéria com a qual ele é feito, mas uma propriedade distinta dessa.

Na visão constitucional, a realidade se apresenta em tipos fundamentalmente diferentes. Cada coisa existente é de um tipo primário. O tipo primário de uma entidade é dado pela resposta à pergunta aristotélica: o que é x mais fundamentalmente? Não existe uma "mera coisa" por trás ou subjacente à instância de um tipo primário. Todo objeto tem essencialmente seu tipo primário, e entidades de diferentes tipos primários têm diferentes condições de persistência. A constituição é uma relação entre coisas de diferentes tipos primários. A constituição traz à existência nova objetos de tipos primários de nível superior ao que estava lá anteriormente. (BAKER, 2004, p. 1, tradução nossa)

De acordo com ela, não são somente incluídos nos tipos primários aquelas entidades determinadas por sua estrutura, pelo material com que são constituídas ou por uma "essência subjacente", mas também pela função que desempenham.

Subjacente à visão constitucional está a idéia de que o mais fundamental é muitas vezes determinado pela capacidade de fazer de alguma coisa pode fazer - suas habilidades e capacidades - em vez de pelo que é feito. Isso é óbvio no caso de artefatos: o que faz de algo um relógio é a função

de dizer tempo, não importa do que seja feito. (BAKER, 2004, p. 3, tradução nossa)

A *constituição* dos objetos se dá a partir do estabelecimento de relações entre tipos primários de entidades que ocasionam o surgimento de novas entidades. O que faz com que uma combinação de tipos primários de entidades seja mais do que um mero agregado é o fato de podermos atribuir ao que resulta dessa combinação uma capacidade de ação e de produção de efeitos no mundo.

Em cada caso, novas coisas de novos tipos, com novos tipos de poderes causais, surgem. Um organismo - mas não o agregado de células que o constituem - pode comer sua presa. Uma bandeira - mas não o agregado de pedaços de pano – pode fazer com que um veterano chore. (BAKER, 2004, p. 3, tradução nossa)

Os artefatos humanos, na sua opinião, não podem ser considerados ontologicamente deficientes porque eles vão além da formação de um mero agregado de coisas, eles possuem uma estrutura, uma função que lhes capacita a produzir mudanças onde estão presentes, e isto é suficiente para lhes garantir o mesmo *status* ontológico atribuído aos objetos naturais. Sua suspeita é de que a gênese desse pensamento se deve à suposição de que os seres humanos não possuem a capacidade de estabelecer as condições de ligação entre os elementos que constituem uma nova entidade, isso só seria realizado pela natureza.

A perspectiva de Baker estabelece uma paridade entre os artefatos naturais e os produzidos pelo homem. Os artefatos produzidos pelo homem não teriam uma deficiência ontológica em relação aos naturais por serem derivados desses, pelo contrário, eles compartilham com os artefatos da natureza uma mesma capacidade, a de produzir mudanças no mundo. A função mais relevante dos artefatos produzidos pelo homem seria a sua capacidade em permitir o reconhecimento da ação do próprio homem.

Quando Guttenberg inventou a imprensa, surgiu um novo tipo de coisa: e mudou o mundo. Seria estranho supor que instrumentos que produziram alterações gigantescas não sejam substâncias genuínas, ou não possuam *status* ontológico. Considerando as mudanças no mundo a partir dos efeitos produzidos pela imprensa e o telefone (e inúmeros

outros tipos de artefatos), os artefatos requerem um *status* ontológico tão forte quanto os objetos naturais. (BAKER, 2004, p. 17, tradução nossa)

A equiparação dos objetos naturais com os produzidos pelo homem está presente no pensamento de Spinoza. Alan Gabbey em *Spinoza on the natural and the artificial* (BENSAUDE-VINCENT & NEWMAN, *The artificial and the natural*, 2007) afirma que na obra de Spinoza não encontramos a diferença presente na obra de Descartes onde a ação humana vai de encontro aos objetos e processos naturais. Spinoza faz uma distinção entre o natural e o artificial, “[...] mas a distinção é entre duas perspectivas com que pode ser vista a mesma coisa, não entre um estado de coisas que expressa processos naturais e outro originado da vontade dos seres humanos que interveem na ordem natural das coisas” (GABBEY In: BENSAUDE-VINCENT & NEWMAN, 2007, posição 2992 de 4285 70%).

Uma volta estranha

Os conceitos de *Homem*, *Inteligência* e *Artifício* são conceitos que estão envolvidos numa série de problemas que vão desde a falta de concordância quanto às suas definições até a forma de como eles devem ser estudados.

Essa situação pode ser vista com certo desconforto de pensamento, porque a falta de um acordo geral sobre o significado de conceitos causa a sensação de que podemos estar enganados ao levarmos em consideração aquilo que não deveríamos. No entanto, para nós, é exatamente esse tipo de situação que nos permite formular e apresentar a resposta que apresentamos à nossa hipótese de trabalho. Nos servimos daquilo que é considerado uma *anomalia* para mostrar que as possibilidades de reflexão em torno de um problema podem ser ampliadas se não os rejeitarmos como tal.

Ao longo da história da filosofia e do próprio desenvolvimento da cultura ocidental fomos orientados a pensar que existe um determinado rigor e necessidade no pensamento. Assim como podemos compreender o pensamento como uma forma de exercício de plena liberdade, podemos pensá-lo como um exercício pleno de necessidade e rigor. Contudo,

suspeitamos que uma das formas de alcançarmos rigor de pensamento tenha se sobrepujado a outras, que por falta de incentivo precisamos ainda descobrir.

Não vemos a diversidade semântica desses conceitos como algo problemático, mas uma evidência daquilo que nos parece bastante razoável pensar, que nossos instrumentos e ferramentas de pensamento (conceitos e teorias) não são capazes de resolver o problema da coexistência de coisas que se negam absolutamente e cuja negação não posso aceitar se não estiver comprometido com a preservação de uma determinada.

Ao propormos uma aproximação da reflexão filosófica da forma de pensar nas artes, pretendemos indicar que existe um continente a ser explorado pelo pensamento filosófico no tocante ao modo como construímos nossas explicações para os problemas que se apresentam a nós. Esperamos que o ruído que aqui produzimos seja transformado, pela generosidade da audição da plateia, numa bela melodia.

Como o próprio nome expressa, uma *volta estranha* é algo insólito e incomum, pois ela não segue aquilo que, à princípio, pensávamos que deveria ser seguido. Nela podemos perceber um movimento que assume um novo sentido sem que, aparentemente, haja uma razão para a mudança. Ela resulta no que Hofstadter chamou de *hierarquia entrelaçada*, ou seja, um tipo de hierarquia em que não há níveis onde uns elementos se sobreponham a outros. Há um dinamismo nessa forma de arranjo hierárquico, pois ele se constitui pela alternância dos elementos que o constituem, o que não dá a possibilidade de manutenção de uma estrutura rígida. Ela pode ser considerada uma espécie de ilusão de ótica, mas isto não retira dela a relevância como um elemento provocador de inquietação, sua elaboração provoca uma sensação de que algo extraordinário está acontecendo.

Como vimos no capítulo anterior, a *volta estranha* foi empregada pela primeira vez na música por Johann Sebastian Bach como o elemento principal da concepção harmônica das suas *fugas*. Contudo, é a representação plástica de Escher que nos permite compreender mais claramente esse movimento classificado como *estranho*.

Existem muitos tipos de *voltas estranhas*, diferenciando-se entre si pelo número de etapas que as constituem. Escher foi fortemente influenciado pela representação geométrica do espaço, a influência da matemática na sua obra é manifesta. Não saberia dizer se ele é um matemático expressando através da arte ideias matemáticas ou se é um

artista se utilizando da matemática para expressar uma criação artística. A técnica da *tesselação*, bastante empregada por ele nas suas obras, que consiste na divisão precisa do espaço em figuras que o preenchem completamente sem que uma se sobreponha a outra, é uma demonstração da influência da matemática na composição de suas obras.



Figura 5: *Systematic Study* de M.C. Escher (1936)
<https://www.ime.usp.br/~thaicia/quasicristais/Tess.html>

No entanto, foi o uso da geometria hiperbólica que lhe deu a possibilidade de construir suas *voltas estranhas*. Esse ramo da geometria surgiu da preocupação dos matemáticos em provar se o *Postulado das Paralelas*, de Euclides, era um postulado que poderia ser provado a partir do emprego de outros postulados da geometria euclidiana ou se era um postulado independente. No início do séc. XX, após séculos de esforços dos geômetras e com a colaboração de figuras do quilate de Carl Friedrich Gauss, Nikolai Ivanovich Lobatchevsky e Henri Poincaré, foi provado que o *Postulado das Paralelas* se trata, verdadeiramente, de um postulado independente. Esse esforço resultou no desenvolvimento de um novo tipo de geometria tão consistente quanto a euclidiana, pois

a validade de suas proposições pode ser determinada pelo processo de axiomatização, procedimento de prova que emprega um conjunto de proposições cuja validade aceitamos sem comprovação ou demonstração.

A hipérbole é a linha que contorna o corte longitudinal que se faz num cone, por essa razão são chamadas de *cônicas*. Alguns estudiosos de Escher caracterizam seu trabalho como a expressão de uma *perspectiva cônica*,

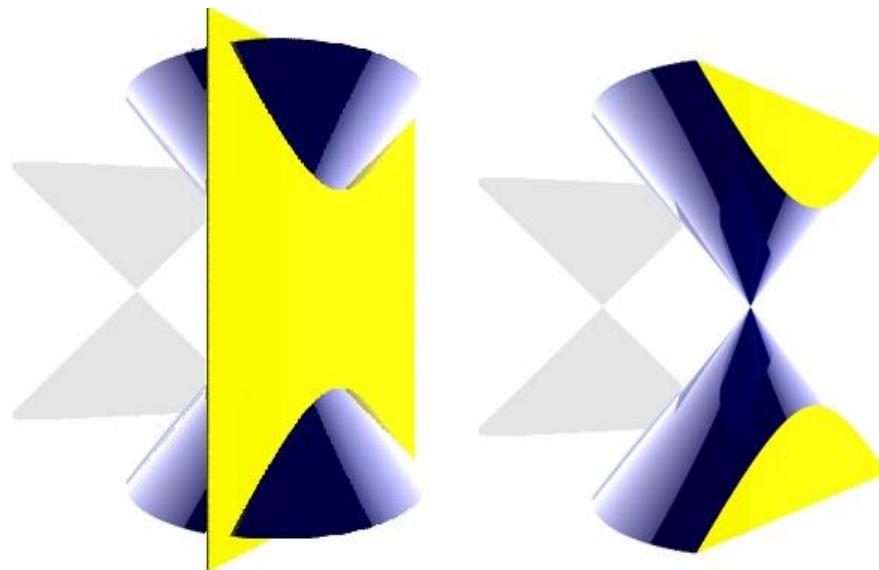


Figura 6: Representação extraída da internet

O emprego dessa geometria possibilitou a Escher a construção de imagens que não temos como transferi-las do meio bidimensional para o tridimensional. Talvez devido a isso, sejamos fortemente induzidos a acreditar que tais imagens nada mais sejam do que uma simples ilusão de ótica. Essa comparação não deixa de ser correta, mas isso não lhe retira importância na discussão que aqui fazemos. A atitude artística de Escher rompe com a concepção de que a representação do mundo tem de produzir uma imagem espelhada deste, as figuras de Escher não podem ser consideradas representações espelhadas do mundo, pelo contrário, elas subvertem toda a ordem que consideramos natural.

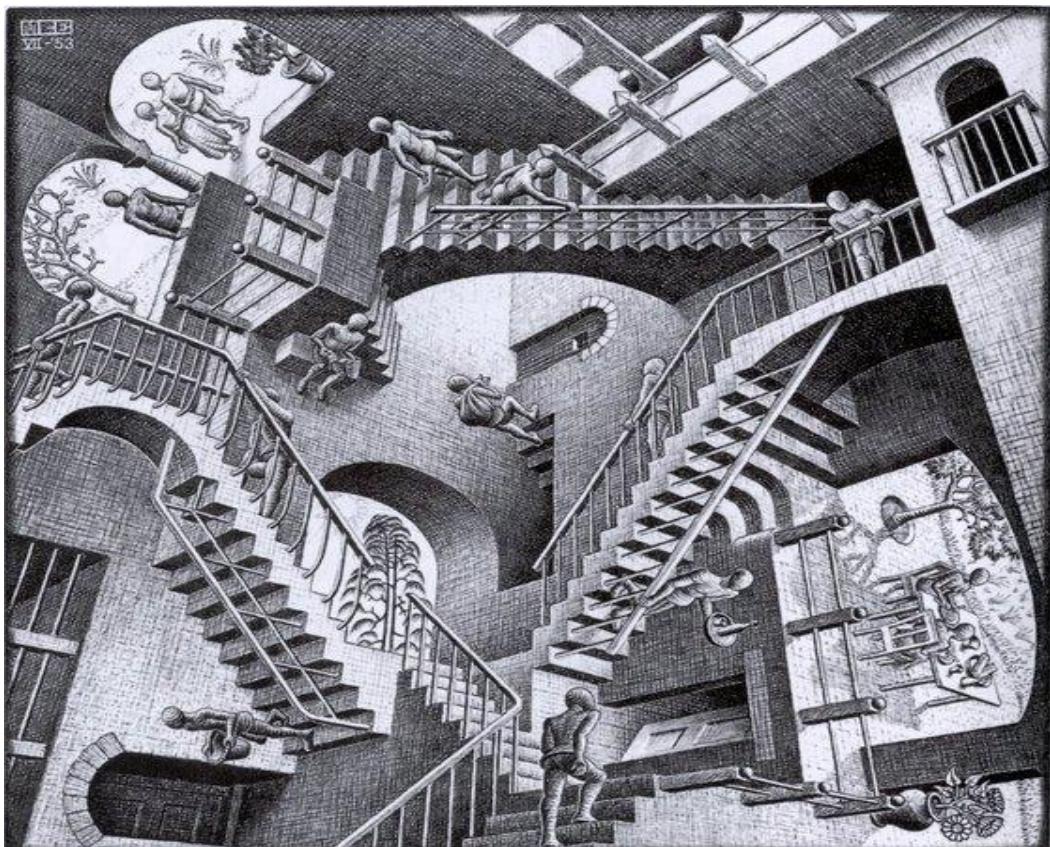


Figura 7: *Relativity* de M.C. Escher (1953)

Na figura acima, podemos ver que não existe somente uma fonte de gravidade, não temos como adotar um ponto de vista privilegiado. As figuras humanas representadas parecem estar em mundos distintos. Se compararmos algumas delas com as outras, por exemplo, a condição de uma nega a física a que a outra está submetida. Dependendo do ponto de vista que adotarmos, a gravidade produzirá ou uma força para baixo, ou uma força para o lado ou, por último, uma força para cima. Os diferentes pontos de vista constituem um conjunto de realidades paralelas, subsistindo ao mesmo tempo e num mesmo lugar. A composição de Escher possui características marcantes, como

[...] uma perspectiva quase sempre representada com três pontos de fuga, sem horizonte específico e levando o observador a mergulhar no universo da imagem que está sendo representada. Escher usou esse tipo de recurso tanto para a representação realística de cenários como para a criação de universos

imaginários com importante presença de ilusão de ótica.
(FURTADO, s/d, p. 16)

Como tratamos na presente tese de analisar a relação entre três conceitos importantes para a reflexão sobre os objetivos do projeto de IA, nos sentimos naturalmente inclinados a adotar o tipo de *volta estranha* constituída de três etapas, formando a figura que Penrose chamou de *triângulo de impossibilidade*.

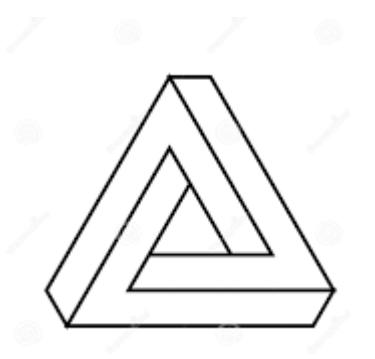


Figura 8: *Triângulo de Impossibilidade*

Esse esquema de composição de imagens foi bastante empregado por Escher na elaboração de muitas das suas obras, como podemos ver na litografia *Relativity* de 1953.

Uma das principais razões para empregarmos essa forma de arranjo de ideias se deve a ele nos produzir uma dupla estranheza. A primeira é a de que, quando vemos uma figura que compõe uma *volta estranha*, imediatamente percebemos que há algo de “errado” nesta. Ela parece envolver erros lógicos que se dão por meio de saltos inferenciais. No mínimo poderíamos classificá-la como um tipo de *petição de princípio*, pois acaba concluindo aquilo que adota como fundamento inicial. Numa volta de três etapas teríamos três saltos inferenciais sem razão alguma os sustentando. A segunda é que, como observou Penrose, quando separamos os vértices de uma *volta estranha*, não temos como indicar onde se encontra a causa da impossibilidade que aparenta possuir quando vemos a figura completa.

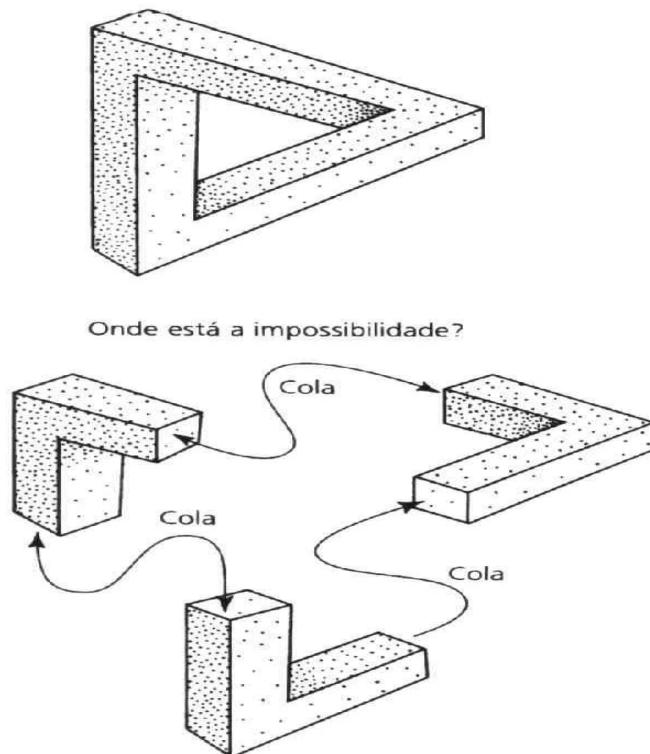


Figura 9: Penrose (1999, p. 80)

Nosso objetivo, ao apresentarmos a *volta estranha* envolvendo os conceitos de *homem*, *inteligência* e *artifício* é responder à hipótese de trabalho desta tese, qual seja, se podemos ou não considerar a IA um âmbito de conhecimento cujos objetivos são, ao menos, admissíveis. A IA é um ramo do conhecimento que cogita a possibilidade de se construir um novo homem, dado que o homem é o único ser que possui inteligência a criação de um novo ser dotado de inteligência também teria *status* de ser humano. Nesse sentido, perguntamo-nos se a IA não promete mais do que é capaz de realizar.

Uma das vantagens do modelo *triângulo de impossibilidade* de uma *volta estranha* é que ele é capaz de representar as forças e movimentos que geram a estranheza produzida por esse tipo de arranjo.

Quando vimos os significados de *homem*, *inteligência* e *artifício* destacamos as dificuldades encontradas ao longo da história do pensamento para definir esses termos e que os mesmos receberam diferentes sentidos nas tentativas de defini-los. Nenhum deles possui uma definição geral que seja capaz de abranger todos os diferentes significados a

eles atribuídos. Quando um termo passa a ser empregado, parece não haver algo que o relate a um e somente um significado. Aquilo que ele significa pode ser alterado de tal maneira que não precisa sequer estar em acordo lógico com os significados anteriores. Ocorre algo semelhante a uma gota d'água quando cai e se choca com o chão, ela é espargida em todas as direções. Não há compromisso dos termos em estabelecer relação com um e somente um significado, pelo contrário, parece ser inerente a eles a condição de que podemos lhes atribuir mais de um significado.

Essa possibilidade de riqueza semântica é propiciada por um tipo de força locucionaria, como diria Austin, que chamamos de *dispersão semântica*. É essa força a responsável pelos saltos que propiciam a construção de uma *volta estranha*, sem ela ficaríamos presos a uma órbita semântica em torno de um determinado termo. Acreditamos que ela seja a principal responsável pela quebra de uma hierarquia rígida, condição necessária para a construção de uma *volta estranha*.

A *volta estranha* aqui proposta representa a interposição dos pensamentos de John Langshaw Austin, Reinhart Koselleck e Douglas Hofstadter. Uma condição a ser observada a fim de que a volta estranha não soe como algo sem sentido é a de que devemos abstrair de toda e qualquer relação de natureza, puramente, dicotômica entre os significados dos conceitos que formam a referida volta. A sensação provocada por uma volta estranha é de que *alguma coisa que parece impossível aconteceu*, algo que não se coloca em concordância com nossas crenças e convicções.

Austin se dedicou a refletir sobre o conteúdo de nossas percepções e concluiu que o impasse gerado entre as duas doutrinas filosóficas que tentam justificar qual seja esse conteúdo se trata de uma grande confusão. Essa discussão está repleta de questões e conceitos bastante indefinidos, controversos e problemáticos. No entanto, a crítica de Austin não chega a apresentar uma forma de superação desse impasse. Como romper com a dicotomia? Se a representação da situação epistêmica não pode ser mais expressa por um par dicotômico, existe alguma forma de arranjo de ideias que possa superar a dicotomia? Austin não tinha uma resposta pronta para essas indagações.

A dicotomia pode ser considerada um modo de pensar pouco convincente, além de expressar uma atitude irracional ao afirmar que uma determinada concepção está

completamente correta, enquanto a outra está completamente errada. A ideia que cada uma das posições antagônicas forma sobre o seu par não é razoável, pois entendemos que tanto uma quanto a outra trazem contribuições para o desenvolvimento do conhecimento. Num certo sentido, seu antagonismo não deixa de trazer algo de positivo ao esforço de conhecimento, dado que a adoção da atitude de perspectiva crítica máxima de um para com o outro influencia a ambos na busca por maior rigor nas suas respostas.

A discordância com a exacerbação dicotômica se deve ao fato de que ela nos obriga a optar por algo que sabemos não haver razão para concordar, qual seja, a de que a atividade de um dos membros que forma o par dicotômico não acrescenta nada de positivo ao esforço de obtenção de conhecimento.

Uma alternativa de arranjo de ideias que apresenta características que superam as dificuldades impostas pela forma dicotômica de organização de ideias está presente na obra de Reinhart Koselleck (*Estratos do Tempo*, 2014) com a sua concepção estratificada do tempo. No nosso entendimento essa concepção rompe com o modelo dicotômico, porque ela pressupõe a relação de mais do que dois elementos antagônicos necessários à produção de uma compreensão do mundo. No entanto, essa forma de organização do pensamento possui uma limitação, ela não permite um dinamismo nas relações entre os elementos que a constituem.

O *triângulo de impossibilidade* empregado por Bach, Gödel e Escher e abordado por Douglas Hofstadter (*Gödel, Escher e Bach: um entrelaçamento de gênios brihantes*, 2001) é, na nossa opinião, a forma de arranjo que se ajusta ao propósito de superar a dicotomia, pois ela consegue absorver as mudanças de significado que estão presentes e dão origem a uma *volta estranha* ao quebrar a hierarquia rígida pressuposta por um par dicotômico.

Outro elemento indispensável à compreensão adequada de uma *volta estranha* é a concepção de Arthur Lovejoy (*The Great Chain of Being*, 2001) de que os diferentes significados de um determinado conceito formam o que ele chama de *grande cadeia*. Essa noção é fundamental aqui, porque ela apresenta uma inovação quanto ao modo como podemos compreender a relação entre os diferentes significados que um termo possui ao nos permitir entender esses diferentes significados como elementos que constituem uma

determinada totalidade semântica. Assim, esses diferentes significados podem passar a não serem compreendidos negativamente como uma ruptura do tecido semântico das palavras.

Numa cadeia que constitui uma *volta estranha* o sentido dos termos está submetido a um outro tipo de encadeamento, distinto do empregado nas construções lógicas tradicionais. No uso tradicional, que pressupõe uma hierarquia rígida de conceitos, ao ser enunciado um termo, devemos seguir um eixo de raciocínio que exclui todas as demais possibilidades de significação se quisermos construir algo considerado logicamente válido ou, pelo menos, com sentido. Na construção de uma *volta estranha* não há um eixo a ser seguido. A estrutura de propagação semântica é ramificada, ou seja, quando um termo é enunciado, abre-se um leque de opções de significação; ao contrário do modo tradicional em que devemos restringir as possibilidades de significação. A principal distinção de funcionamento da linguagem que encontramos entre esses diferentes modos de pensamento é que no dito tradicional as possibilidades de significação formam uma convergência; enquanto numa *volta estranha* a dispersão de significados é a característica. Por isso somos levados a acreditar que não há relação de compromisso exclusivo dos termos com um e somente um significado dos muitos que lhe podem ser atribuídos.

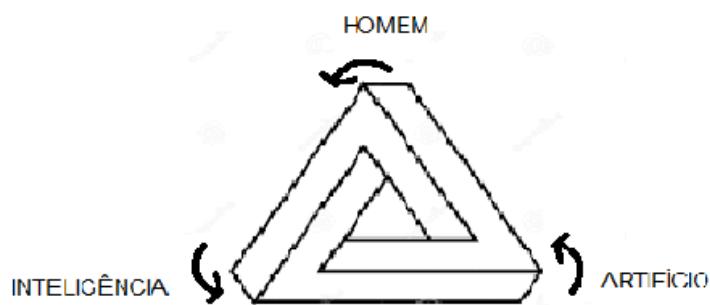


Figura 10: *Triângulo de Impossibilidade Homem, Inteligência e Artifício*

Com esse diagrama pretendemos representar o seguinte raciocínio: *com a sua inteligência, o homem produz artifícios, como o próprio homem*. A primeira consideração

que devemos fazer, a fim de construirmos a *volta estranha* envolvendo os três conceitos acima referidos, é evitarmos cair na armadilha das relações de natureza dicotômica entre os diferentes significados desses conceitos.

Acreditamos que a forma de organização dicotômica dos conceitos e ideias é uma das razões que faz com que seja adotada uma perspectiva negativa sobre o alcance do objetivo último do projeto de IA, ao ponto de não o considerar, nem ao menos, como admissível. É possível superarmos os obstáculos originados pela forma dicotômica de organização das ideias se entendermos que as diferenças semânticas dos significados dos conceitos constituem uma totalidade, uma *grande cadeia*, em que os conteúdos semânticos se relacionam livremente, além daquelas consideradas válidas pela lógica clássica, por exemplo.

O argumento que constitui a *volta estranha* é realizado em três etapas. Ele tem a forma de um triângulo, em que o primeiro vértice é constituído pelos conceitos de homem e artifício, o segundo pelos conceitos de inteligência e homem, e o terceiro pelos conceitos de artifício e inteligência. Nesse triângulo, talvez a compreensão mais difícil de ser superada seja a de que os artifícios que nós humanos construímos também podem ser considerados como produtos da natureza. A ideia de que não há uma continuidade ontológica entre nós e nossas criações é uma das principais razões que impedem, na nossa opinião, a aceitação da admissibilidade do objetivo último do projeto de IA.

Pensamos que a resistência à tese de que vale a pena empregar esforços e recursos para na realização de um objetivo tão ambicioso quanto o de construção de inteligência artificial se deve a problemas e questões que não são propriamente de naturezas científica, prática e tecnológica.

Nosso propósito ao construirmos uma *volta estranha* envolvendo o emprego dos conceitos acima mencionados visa mostrar que é possível, ao tomarmos como ponto de partida o conceito de *homem*, retornarmos a um novo homem por meio dos conceitos de *inteligência* e *artifício*. Se entendermos o resultado do trabalho humano como algo tão natural quanto uma maçã, se aceitarmos que a inteligência não é algo misterioso, embora difícil e, sobretudo, que nossa humanidade não é determinada pelo material com o qual somos construídos, é totalmente aceitável cogitar a possibilidade de se construir uma

máquina que seja dotada do mesmo atributo ontológico que nós seres humanos. Com isso, procuramos mostrar como é possível, por meio do emprego de uma *volta estranha*, tomarmos como ponto de partida o homem biológico e chegarmos ao homem artifício. (Novo homem).

Para finalizar, um esclarecimento, não entendemos que o *novo* homem seja *novo* no sentido de que algo completamente *inusitado* passe a existir, consideramos que o *novo* se restringe ao modo como o ser humano poderá ser engendrado. Se suspendermos algumas considerações que julgamos válidas na realização de nossos julgamentos seremos surpreendidos com as possibilidades de conclusões que tal estratégia nos propicia.

CONSIDERAÇÕES FINAIS

A reflexão sobre a possibilidade de obtenção de inteligência por meio do conhecimento humano é um campo de investigação para o qual convergem interesses de diferentes naturezas. Muitos veem no desenvolvimento dessa tecnologia a possibilidade de realização de negócios, outros do estabelecimento de novos vínculos com o mundo, terceiros de um sonho de poder, outros ainda de alcançarmos uma nova compreensão do que seja o conhecimento. Por estar relacionado a vários aspectos da vida humana, o problema da IA adquire uma relevância especial devido ao seu potencial de mudança da vida, do ambiente e da própria sociedade.

Ao investigar esse tema, pude observar que há uma grande efervescência de ideias e concepções sobre o homem, o conhecimento, a filosofia, a ciência, a arte que foram revigoradas pelo debate em torno da IA. Nesses anos de pesquisa a literatura em torno deste assunto mostrou-se bastante rica, tanto no âmbito filosófico quanto fora dele. Nessa teia bibliográfica se relacionam temáticas que tornam a discussão cada vez mais abrangente e ampla, não permitindo que uma determinada área de conhecimento se aproprie do debate. À obtenção de inteligência artificial não é suficiente o desenvolvimento científico-tecnológico de dispositivos ditos inteligentes, pois a definição do que seja a inteligência não é uma tarefa que compete somente a essa área.

A tarefa de discutir o significado das expressões que empregamos para explicar o que acontece ou o que é, é tão relevante quanto o próprio objetivo prático de construção de um dispositivo inteligente. A necessidade de esclarecimento dos significados dos conceitos é uma condição necessária para que sejamos capazes de justificar o reconhecimento de IA.

No entanto, assim como não julgamos que o reconhecimento de IA possa se dar pela mera apresentação de um dispositivo sobre o qual é dito que é inteligente, não entendemos que a tarefa de esclarecimento dos conceitos será capaz de produzir o reconhecimento de IA. Talvez o reconhecimento de inteligência artificial no futuro venha a se dever mais aos resultados da interação entre máquinas inteligentes e os seres humanos do que por uma justificativa prévia em favor do reconhecimento de semelhança entre ambos. Assim como numa relação que envolve diferenças de qualquer ordem, o procedimento de discriminação/segregação/separação ocorre quase que de forma

inxorável, fazendo com que aquilo que é aparentemente distinto seja, de fato, considerado distinto. A perspectiva de inclusão e de reconhecimento do que aparentemente é distinto, como a cor da pele, das crenças religiosas, se de carne ou de lata, dissolve o *status* da relação de impertinência entre um e outro. Por isso, estamos mais inclinados a pensar que o reconhecimento de IA será feito pelo convívio com máquinas que nos convencerão da sua inteligência e capacidade. Assim como há e houve discriminação de humanos com humanos e que só foram superadas pelo convívio, acreditamos que o mesmo ocorrerá com as máquinas, se for concretizado o objetivo de construção de IA. Após isso, faremos os ajustes necessários para que nossa compreensão de mundo entre em conformidade com essa nova experiência.

O que apresentamos aqui como resposta ao problema do projeto de IA, mais do que uma tentativa de solução definitiva para a questão, é um convite para o debate acerca do próprio homem, do conhecimento e do mundo. O problema da IA nos remete a problemas que são tipicamente humanos, como o reconhecimento de direitos, por exemplo. A característica de ser um problema que tem o poder de arrebatar o interesse de tantos e de tantas diferentes áreas e âmbitos do viver humanos faz com que sua consideração seja inevitável.

A *volta estranha* aqui apresentada nos mostra que o objetivo principal do projeto de IA é admissível, pois as barreiras apresentadas a aceitação dessa ideia são originadas pelo modo dicotômico de organização das ideias. Os conceitos que formam o triângulo de impossibilidade são demasiado ambíguos para nos garantirem alguma certeza acerca de qual dos membros que formam a relação antitética entre a admissibilidade e a não aceitação é o correto e aceitável. Enquanto a vertente céтика reflete a questão levando em consideração as fragilidades das tentativas de provar inteligência artificial, a vertente que acredita na viabilidade do projeto de IA dá relevância aos avanços práticos e teóricos alcançados pela busca dos objetivos desse projeto. É importante salientar que, tanto quando é vista negativamente, como forma possível de controle sobre o homem, ou positivamente, como busca por um melhor viver do homem, a IA é considerada como algo possível de ser alcançado.

Mesmo que o ser humano não consiga alcançar tal propósito, já que não está fora do terreno das possibilidades a de que a humanidade seja destruída antes disso, o projeto de IA não perde em importância, pois nos faz refletir sobre o conhecimento, a tecnologia, o próprio homem, sua relação consigo mesmo e com o mundo. Isso por si só justificaria tal importância, mas há, ainda, um elemento a mais na sustentação de sua pertinência, o propósito de produzir inteligência artificialmente nos incita a buscar novas formas de compreensão do mundo e de nós mesmos, nos impelindo a não nos contentarmos com aquilo que temos de conhecimento.

Acreditamos que não está fora do terreno do possível a pretensão de construção do *novo homem*, substrato oriundo do sucesso do projeto de IA. Não se trata de um *novo homem* melhor ou pior do que esse que já existe, mas *novo* no sentido de ser engendrado de uma maneira diferente do que a que vimos até aqui ser empregada.

Quem sabe, o objetivo de construção de uma máquina inteligente por parte do homem seja uma tentativa deste de fugir da solidão. Numa vastidão cósmica sem ter com quem dialogar, já que a ideia de Deus é frustrante, pois é difícil dialogar com Deus, a humanidade traça o seu caminho ao buscar companhia na sua viagem pelo tempo.

BIBLIOGRAFIA

- ADAMSON, D. Blaise Pascal. *Mathematician, physycist and thinker about God*. St. Martin Press, Inc. New York, 1995.
- ADLEMAN, L. *Molecular computation of solutions to combinatorial problems*, Science, 266, 1021-1024 (Nov, 11, 1994)
- ALCHEMY Website. Archive August 2002. Online. Available at: http://www.alchemywebsite.com/a-archive_aug02.html. Accessed on 23 Jan 2009.
- ALFONSO-GOLDFARB, A.M. *Da Alquimia à Química*. São Paulo-EDUSP, 1988.
- ALFONSO-GOLDFARB. A.M. *A passagem da alquimia à química: uma história lenta e sem rufar de tambores*. Revista ComCiência, nº 130. Campinas, 2011.
- ARISTÓTELES. *De Anima*. Apresentação, tradução e notas de Maria Cecília Gomes dos Reis. São Paulo. Ed. 34, 2006.
- ARISTÓTELES. *The Complete Works of Aristotle* (editado por Jonathan Barnes). Princeton; Princeton University Press, 1995.
- AUSTIN, J.L. *Sentido e Percepção*. Tradução: Armando Manuel Mora de Oliveira. São Paulo: Martins Fontes, 1993. Coleção Tópicos.
- AUSTIN, J.L. *Philosophical Papers*. Oxford: Clarendon Press. 1961.
- BACON, Francis. *Novum Organum ou Verdadeiras Indicações Acerca da Interpretação da Natureza*. Tradução de José Aluysio Reis de Andrade. Homepage do grupo: <http://br.egroups.com/group/acropolis/>
- BACON, Francis. *O progresso do conhecimento* / Francis Bacon; tradução, apresentação e notas Raul Fiker. São Paulo: Editora UNESP, 2007.
- BAKER, Lynne Rudder. *The ontology of artifacts*. In: *Philosophical Explorations* (7), 99-112, 2004.
- BATESON, Gregory. *Mind and Nature. A necessary unity*. New York. E.P. Dutton. 1979.
- BATESON, Gregory. *Propósito Consciente versus Natureza*.
- BENNETT, M.R. & HACKER, P.M.S. *Fundamentos filosóficos da neurociência*. Instituto Piaget. Tradução de Rui Alberto Pacheco. Lisboa, 2005.
- BENSAUDE-VINCENT, B. & NEWMAN, W.R. *The artificial and the natural: an evolving polarity*. MIT Press, Cambridge – Massachusetts, 2007.
- BERG OLSEN et alli. *A Companion to the Philosophy of Technology*. Blackwell Publishing Ltd. West Sessex, UK, 2009.
- BODEN, Margaret. *The Philosophy of Artificial Intelligence*. Oxford: Oxford University Press, 1990.
- BOOLE, G. *An investigation of laws of thought, on which are founded the mathematical theories of log and probabilities*. London: Walton and Marberly, 1854.
- BOSTROM, N. *Superintelligence, paths, dangers, strategies*. Oxford University Press. Oxford, UK. 2014.

- BRNA, Paul. *Artificial intelligence in educational software: has its time come?* British Journal of Educational Technology, Vol 30, Nº 1, p.79-81, 1999.
- BROOKS, R. *Inteligência sem representação*. Tradução: Pedro Rocha de Oliveira e João de Fernandes Teixeira. Em: www.filosofiadamente.org/images/stories/textos/representacao.doc Acesso em: 28/11/2011.
- CASSIRER, E. *An essay on man. An introduction to a philosophy of human culture*. Yale University Press, 1944.
- CASSIRER, E. *Antropologia Filosófica. Um ensaio sobre o homem: introdução a uma filosofia da cultura humana*. Tradutor Vicente Felix de Queiroz. Editora Mestre Jou, São Paulo. s/d.
- CASTRO, F.S. & LANDEIRA-FEERNANDEZ, J. *Alma, corpo e a antiga civilização grega: as primeiras observações do funcionamento cerebral e das atividades mentais* In: Psicologia: Reflexão e Crítica. Vol. 24, número 4, Porto Alegre, 2011. http://www.scielo.br/scielo.php?pid=S0102-79722011000400021&script=sci_arttext
- COSTA, Newton da. *Aquiles e a Tartaruga*. Jornal de Resenhas. <http://www1.folha.uol.com.br/fsp/resenha/rs1108200103.htm>
- CHEMERO, A. *Radical Embodied Cognitive Science*. Massachusetts: MIT/Bradford, 2009.
- CHEUNG, B. et alii. *SmartTutor: an intelligent tutoring system in web-based adult education*. Journal of Systems and Software, Vol. 68, Nº 1, p. 11-25, 2003.
- CHOU, Chih-Yueh et alii. *Redefining the learning companion: the past, present, and future of educational agents*. Computers & Education, Vol. 40, Nº 3, p. 255-269, 2003.
- CHRISTIAN, B. *O humano mais humano: o que a inteligência artificial nos ensina sobre a vida*. Tradução: Laura Teixeira Motta. 1ª. Edição. São Paulo: Companhia das Letras, 2013.
- CHURCHLAND, Paul. *The engine of reason, the seat of soul*. Massachusetts: MIT Press, 1996. COLE, David. 'Thought and Thought Experiments', Philosophical Studies, Vol. 45, Nº: 3, p. 431-44, 1984.
- CHURCHLAND, Paul. *The ontological problem*. Cambridge, Massachusetts: the MIT Press, 2013.
- CLARK, A. *Being there: putting brain, body, and world together again*. Massachusetts: MIT/Bradford, 1998.
- CLARK, A. *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*. Cambridge: MIT Press, 1991.
- CLARK, A. *Mindware. An introduction to the philosophy of cognitive science*. New York: Oxford University Press, 2001.

- CLARK; A.; CHALMERS, D. *The Extended Mind*. In: Analysis, n° 58, V.1, p.10-23, 1998.
- COLE, D. *The Chinese Room Argument*. In: ZALTA, E. N. (Org.) The Stanford Encyclopedia of Philosophy. 2012. Disponível em: <http://plato.stanford.edu/archives/win2012/entries/chinese-room/> Acesso em: 15/12/2012
- CUMMING, G. *Artificial intelligence in education: an exploration*. Journal of Computer Assisted Learning, N° 14, p. 251-259, 1998.
- DARWIN, C. *A origem das espécies: esboço de 1842*. Rio de Janeiro: N. Compton, 1996.
- DAVIDSON, Donald. *Essays on actions and events*. Oxford: Oxford University Press, 1989.
- DENNETT, D. - *Brainstorms*. S.I., Bradford Books, 1978.
- DENNETT, Daniel. *The Intentional Stance*. Cambridge: MIT Press, 1987.
- DENNETT, Daniel. *Tipos de Mente*. Rio de Janeiro: Rocco, 1997.
- DESCARTES, R. *Meditações metafísicas*. São Paulo: Nova Cultural, 1999.
- DOTY, R. W. (2007). *Alkmaion's discovery that brain creates mind: A revolution in human knowledge comparable to that of Copernicus and of Darwin*. *Neuroscience*, 147, 561-568. [[Links](#)]
- DREYFUS, H. L. *O que os computadores não podem fazer: crítica da razão artificial*. Rio de Janeiro: A casa do Livro Eldorado, 1975.
- DREYFUS, H. L. e DREYFUS, S.E. *Mind over machines*. New York: Free Press, 1986.
- DUPUY, J.P. *Nas origens das Ciências Cognitivas*. Tradução: Roberto Leal Ferreira. São Paulo: Editora Unesp, 1996.
- ECO, Umberto. *A busca da língua perfeita na cultura europeia*. Tradução Antonio Angonese. Bauru, SP: EDUSC, 2001.
- FODOR, J. - *Psychological explanation*. New York, Randon House, 1968.
- FODOR, J. - *The language of thought*. Sussex, The Harvester Press, 1976.
- FODOR, Jerry. *The modularity of mind*. Cambridge, Massachussetts, MIT Press, 1983.
- FOLEY, R. *Os humanos antes da humanidade: uma perspectiva evolucionista*. Tradução Patrícia Zimbres. São Paulo: Editora Unesp, 2003.
- FREGE, G. *Lógica e Filosofia da linguagem*: seleção, introdução, tradução e notas de Paulo Alcoforado. 2ª. Edição ampliada e revisada. São Paulo: EDUSP, 2009.
- FURTADO, André. *O espelho mágico de M.C. Escher*. <https://www.lume.ufrgs.br/bitstream/handle/10183/135249/000736567.pdf?sequenc e=1> s/d.

- GEHLEN, Arnold. *A alma na era da técnica*. Problemas de psicologia social na sociedade industrializada. Tradução de Manuela Pinto dos Santos. Livros do Brasil – Lisboa, s/d.
- GONZALEZ, M. E. Q.; BROENS, M. C. *Darwin e a virada naturalista na filosofia*. In: MORAES, J. Q. (Org.). *Darwin e a origem do homem*. Campinas: UNICAMP, 2010.
- GUNDERSON, K. - *M mentality and machines*. New York, Anchor Books, 1971.
- HAACK, Susan. *Filosofia das Lógicas*; tradução de Cezar Augusto Mortari, Luiz Henrique de Araújo Dutra. São Paulo: Editora UNESP. 2002.
- HADOT, Pierre. *The veil of Isis. An Essay on the History of the Idea of Nature*. Translated by Michael Chase. Harvard University Press, Cambridge, Massachusetts, 2006.
- HASELAGER, W. F. G. & GONZALEZ, M. E. Q. *Conhecimento comum e auto-organização*. In E. RABOSSI (Ed.), *La mente y sus problemas: Temas actuales de filosofía de la psicología*. Buenos Aires, pp. 95-105, 2004. Disponível em: <http://www.nici.ru.nl/~haselag/publications/ConhecimentoComumAutoorg.pdf> Acesso em: 05/05/2012.
- HASELAGER, W.F.G. *Auto-organização e comportamento comum: opções e problemas*. In: GONZALEZ, M.E.Q; D'OTTAVIANO, I.M.L; SOUZA, G.M. (Orgs). *Autoorganização: estudos interdisciplinares*. Campinas, Vol. 38, pp. 213-235, 2004.
- HAWKINS, J. & BLAKESLEE, S. *Intelligence*. Traduction: Bernard Jolivalt. Pearson Education France. 2005.
- HEISENBERG, Werner. *Physics and beyond*. Harper and REow, Publishers. London, 1971.
- HOFSTADTER, Douglas. *Gödel, Escher e Bach. Um entrelaçamento de gênios brilhantes*. Tradução de José Viegas Filho. Brasília; Universidade de Brasília: São Paulo: Imprensa Oficial do Estado, 2001.
- HORGAN, John. *A mente Desconhecida: por que a ciência não consegue replicar, medicar e explicar o cérebro humano*. Tradução Laura Teixeira Motta. São Paulo; Compainha das Letras, 2002.
- http://www.cbpm.com.br/paginas/oque_geologia.php retirada da página em 03 de março de 2016 as 09:57
- <https://pt.wikipedia.org/wiki/Alquimia>
- IHDE, Don. *Techinics and práxis*. Dordrecht, Reidel. 1979.
- IHDE, Don. *Technology and the life world. From Garden to Earth*. Bloomington:Indiana University Pres, 1990.
- IHDE, Don. *Heidegger's Technologies: postphenomenological Perspectives*. New York: Fordham University Press. 2001.

- JEFFERSON, Geoffrey. *The Mind of Mechanical Man*. British Medical Journal de 25 de Junho de 1949.
- JUNG, C. (1980). *Psychology and Alchemy*. Princeton: Princeton University Press.
- KANT, I. *Lógica*. Rio de Janeiro, Tempo Brasileiro, 1992.
- KANT. *Crítica da Razão Pura*. Tradução de Manuela Pinto dos Santos e Alexandre Fradique Morujão. (5ª. Edição) Fundação Calouste Gulbenkian. Lisboa. 2001.
- KOSELLECK, Reinhart. *Estratos do tempo: estudos sobre história*. Tradução Markus Hediger. 1. Ed. - Rio de Janeiro: Contraponto: PUC-Rio, 2014.
- KRAUSS, L. & Starkman, G.D. *Universal limits on computation, Astrophysics* (2004) (<http://arxiv.org/abs/astro-ph/0404510>)
- KURZWEIL, Ray. *A era das máquinas espirituais*. Tradução de Fábio Fernandes. São Paulo: Aleph, 2007.
- LEIBNIZ, G.W. *Dissertacion acerca del arte combinatorio*. Tradução de Manoel Correia M. Ediciones Universidad Católica de Chile. 1992.
- LENZI, Eduardo Barbosa et VICENTINI, Max Rogério. *Vico e a História como Ciência*. In: *Acta Scientiarum*. v. 24, n. 1, p. 201-210, Maringá. 2002.
- LEVERE, T.H. *Transforming Matter. A history of chemistry from alchemy to the buckyball*. The Johns Hopkins University Press. Baltimore-Maryland, 2001.
- LEWIS, David. *An argument for the identity theory, with addenda*. In: ROSENTHAL, David (ed.). *Materialism and the mind-body problem*. Indianapolis: Hackett, 2000. P: 162-171.
- LLOYD, S. *Ultimate physical limits to computation*, Nature, 406, 1047-1054 (2000)
- LOVEJOY, Arthur Oncken. *The great chain of being. A study of the history of an idea*. Harvard University Press. Cambridge, Massachusetts. 2001.
- LOVEJOY, Arthur Oncken. *Reflections on human nature*. The Johns Hopkins Press. Baltimore, 1961.
- LOVEJOY, Arthur Oncken. *The revolt against dualism. Na inquiry concerning the existence of ideas*. The Open Court Publishing Company e W.W. Norton & Company, Inc. 1930.
- LÖHR, Robert. *A máquina de Xadrez*. Tradução de: André del Monte e Christina Michahelles. Rio de Janeiro: Best-bolso, 2012.
- MATURANA, H. R. *The mind is not in the head*. Journal of Social and Biological Structures 8(4): 308-311, 1985 Available at <http://cepa.info/575>
- MCCARTHY, J.; HAYES, P.J. *Some Philosophical Problems from the Standpoint of Artificial Intelligence*. In: MICHIE, D. (Org.). *Machine Intelligence*, V. 4, New York, pp. 463-502, 1969.
- MIGUENS, S. *Alguns problemas de filosofia da Inteligência Artificial*. In: *Intelectu* (revista electrónica de divulgação filosófica), Porto, n° 3, pp. 111 – 111, 2000.

- MITCHAM, Carl. *Philosophy and Technology II: information technology and computers in theory and practice*. D. Reidel Publishing Company. Dordrecht, Holland. 1986.
- MITCHAM, Carl. *Que es la filosofía de la tecnología?* Traducción del original inglés: César Cuello Nieto y Roberto Méndez Sting. Editorial Anthropos, Barcelona, 1989.
- MITCHAM, Carl. *Thinking Through Technology. The path between engineering and philosophy*. The University of Chicago Press. 1994.
- MORAVEC, Hans. *Mind Children: the future of robot and human intelligence*. Cambridge-Massachussetts, Harvard University Press, 1988.
- MORAVEC, Hans. *Robot: Mere machine to transcendent mind*. New York: Oxford University Press, 1999.
- MULLER, G.A. & PRIBRAM, J. *Plans and the structure of behavior*. New York, Holt, 1960.
- MUMFORD, Lewis. *Technics and Civilization*. Routledge & Kegan Paul Ltda. London, 1934.
- MUMFORD, Lewis. *El mito de la máquina. Técnica y evolución humana*. Traducción: Arcadio Rigodón. Pepitas de Calabaza Ed., 2010.
- MURARO, Rose Marie. *A automação e o futuro do homem*. Petrópolis, Rio de Janeiro. Editora Vozes, 1969.
- NAGEL, Thomas. "What is it like to be a bat?" *The Philosophical Review*, Vol. LXXXIII, Nº 4, p. 435-50, 1974.
- NEUMANN, J. von. *The computer and the brain*, Yale University Press, New Haven, USA (1958)
- NEWELL, A & SIMON, H. *The simulation of human thought*. Current Trends in Psychological Theory, n? I: 152-179, 1961.
- NEWELL, A. & SIMON, H. *GPS a program that simulates human thought*. In: FEINGEMBAUM & FELDMAN, eds. - *Computers and thought*. New York, McGraw-Hill, 1968. p. 279-293.
- NEWELL, A. & SIMON, H. - *Human problem solving*. New York, Prentice Hall, 1972.
- NEWELL, A. *The Logic Theory Machine: a complex information processing system*. California: The RAND Corp., 1956.
- NEWELL, A.; SHAW, J.D.; SIMON, H.A. Report on a General Problem-Solving Program for a Computer. In: UNESCO. *Information Processing: Proc. International Conference on Information Processing*, Paris, p. 256-264, 1958. Disponível em: http://bitsavers.trailingedge.com/pdf/rand/ipl/P-1584_Report_On_A_General_ProblemSolving_Program_Feb59.pdf Acesso em: 15/08/2012.
- NEWELL, A.; SIMON, H.A. *Human Problem Solving*. Englewood Cliffs: Prentice Hall, 1972.

- NEWTON, I. & LEIBNIZ, G. *La polémica sobre la invención del cálculo infinitesimal – Escritos y Documentos*. Tradução de Antonio J. Durán & José Luis Arántegui. Editorial Críticas, S.L. Barcelona, 2006.
- NICOLELIS, Miguel. *Muito além do nosso eu*. São Paulo: Compainha das Letras, 2011.
- NILSSON, N. - Problem-solving methods in Artificial Intelligence. New York, McGraw Hill, 1971.
- ORTEGA Y GASSET, J. *Obras Completas*. Tomo V. Revista de Occidente. Madrid, 1964.
- ORTEGA Y GASSET, J. *Obras Completas*. Tomo IX. Revista de Occidente. Madrid, 1965.
- PATEL, A. *The future of Computation*, Quantum Physics, (2005) (<http://arxiv.org/abs/quant-ph/0503068>)
- PATTISON MUIR, Matthew Moncrieff. *The story of alchemy and the beginnings of chemistry*. The Project Gutenberg eBook. <http://pgdp.net>, 2004.
- PENROSE, Roger. *Lo grande, lo pequeño y la mente humana*. Traducción: Javier García Sanz. Cambridge University Press, Madrid, 1999.
- PENROSE, Roger. *Shadows of the mind: a search for the missing science of consciousness*. Oxford University Press. United Kingdom, 1996.
- PENROSE, Roger. *The emperor's new mind*. Oxford University Press. United Kingdom, 1999.
- PEREIRA, J.A. *A percepção em Thomas Reid*. Revista Kínesis, Vol. I, n° 02, p. 133 – 143, 2009.
- PICH, R.H. *Thomas Reid, o Método de Filosofar e a rejeição do ceticismo*. In: Revista Dissertatio, Porto Alegre, n° 32, p. 343-275, 2010.
- PINTO, S.M.C.R. *A Natureza Histórica da Cognição: Debates filosóficos na Teoria dos Sistemas Dinâmicos na Ciência Cognitiva*. Dissertação de Mestrado. Universidade do Porto. Portugal, 2007.
- PIRES DA SILVEIRA, J. A. *Searle e Dennett: duas perspectivas de estudo da mente*. In: *Problemata*: R. Intern. Fil. Vol. 04. No. 02. (2013), p. 238-258 ISSN 15169219. DOI: <http://dx.doi.org/10.7443/problemata.v4i2.1675>
- PITT, J.C. & DURDIN, P.T. *Broad and Narrow interpretations of philosophy of technology*. Kluwer Academic Publishers. Dordrecht, The Netherlands, 1990
- PLACE, U.T.; SMART, J. *Is Consciousness a Brain Process?* In: *British Journal of Psychology*, n° 47, pp. 44–50, 1956. Disponível em: <http://home.sandiego.edu/~baber/analytic/Place1949.html> Acesso em: 15/09/2012.
- PLATÃO. *Teeteto*. Desterro: Ed. Barba Ruiva, 2008. Disponível em: www.barbaruiva.libertar.org Acesso em: 10/09/2011.

- PORCHAT, O. *Rumo ao Ceticismo: A Filosofia e a Visão Comum de Mundo*. São Paulo: Editora UNESP, 2007. PORT, R.;
- RAMACHANDRAN, V.S. & HIRSTEIN, W. *The Science of Art: a neurological theory of aesthetic experience*. Journal of Consciousness studies, 6 (6-7), 15-51, 1999.
- REGAI, J. (1992). *The Philosopher's Stone: Alchemy and Chemistry*, Alif: Journal of Comparative Poetics , No. 12, pp. 58-77.
- ROOB, A. (2006). *The Hermetic Museum: Alchemy & Mysticism*. Koln:Taschen.
- RORTY, Richard. *A filosofia e o espelho da natureza*. Tradução de Antônio Trânsito; revisão técnica de César Ribeiro de Almeida. Rio de Janeiro: Relume-Dumará, 1994.
- ROSENBERGER, R. & VERBEEK, P. *Postphenomenological investigations*. Lexington Books – London, 2015.
- ROSENTHAL, David (ed.). *Materialism and the mind-body problem*. Indianapolis: Hackett, 2000.
- RUSSEL, S. & NORVIG, P. *Inteligencia artificial: um enfoque moderno*. Segunda edición. Pearson Educación, Madrid. 2004.
- RYLE, G *Critical Essays*. New York: Routledge, 2009.
- RYLE, G. *Collected Essays*. New York: Routledge, 2009.
- RYLE, G. *The concept of Mind*. London: Penguin, 1949.
- SCHANK, R.; ABELSON, R. *Scripts, Plans, Goals, and Understanding*, Hillsdale, NJ: Lawrence Erlbaum, 1977.
- SCHELER, M. *A situação do homem no cosmos*. Tradução de Artur Morão. Edição texto & grafia, Lisboa. 2008.
- SEARLE, J. *Intencionalidade*. Tradução: Julio Fischer e Tomás Rosa Bueno. São Paulo: Martins Fontes, 1995.
- SEARLE, J. *Intentionality*. Cambridge: Cambridge University Press, 1983.
- SEARLE, J. *Mente, Linguagem e Sociedade: Filosofia no mundo real*. Tradução: F.Rangel. Rio de Janeiro: Rocco, 2000.
- SEARLE, J. *Mind, Language and society. Philosophy in the real world*. New York: Basic Books, 1998.
- SEARLE, J. *Minds, brains, and programs*. Behavioral and Brain Sciences, n° 3, V.3, pp. 417- 457. 1980. Disponível em: <http://www.bbsonline.org/Preprints/OldArchive/bbs.searle2.html> Acesso em 04 de outubro de 2011.
- SEARLE, John. *A redescoberta da mente*. Tradução Eduardo Pereira e Ferreira. São Paulo: Martins Fontes, 1997.
- SEARLE, John. *Mente, cérebro e ciência*. Lisboa: Edições 70, 1997a.
- SHANNON, C.E. *A mathematical theory of communication*, Bell SystemTech. J. 27, 379-423; 623-656 (1948)

- SHAPIRO, L. *Embodied Cognition. (New problems of philosophy)*. New York: Routledge, 2011.
- SIBILIA, Paula. *El hombre postorgánico. Cuerpo, subjectividad y tecnologías digitales*. 1^a ed. Buenos Aires: Fondo de Cultura Económica, 2005.
- SIEMER, Julika; ANGELIDES, Marlos. *A comprehensive method for the evaluation of complete intelligent tutoring systems*. DecisionSupport Systems, Nº 22, p. 85-102, 1974.
- SIMON, H. *Models of discovery*. Cambridge, Mass. MIT Press, 1974.
- SIMON, H. *As ciências do artificial*. Tradução de Luís Moniz Pereira. Armênio Amado - Editor, sucessor. Coimbra, 1981.
- SIMON, H. *Models of man: mathematical essays on rational human behavior in a social setting*. New York: Ed. John Wiley and Sons, 1957.
- SLOMAN, A. - *The computer revolution in philosophy*. Sussex, Harvester Press, 1978.
- TABACHNICK, D. 7 KOIVUKOSKI, T. *Globalization, technology and philosophy*. State University of New York Press, 2004.
- TEIXEIRA, 1. de F. & QUILICI GONZALES, M.E. *Inteligência Artificial e teoria de resolução de problemas*. Trans/Form/Ação, São Paulo, 6: 45-52, 1983.
- TEIXEIRA, J.F. *Cérebros, Máquinas e Consciência*. São Carlos: Editora da UFSCar, 1996.
- TEIXEIRA, J.F. *Cérebros, máquinas e consciência: uma introdução à filosofia da mente*. São Carlos; EDUFSCar, 1996.
- TREFIL, James. *Somos diferentes? Um cientista explora a inteligência ímpar da mente humana*. Tradução Alexandre Tort; Revisão Técnica Filadelfo Cardoso Santos. Rio de Janeiro: Rocco, 1999.
- TURING, A. *Computing machinery and intelligence*. In: FEINGEMBAUM & FELDMAN, eds. - *Computers and thought*. New York, McGraw-Hill, 1968. p. 11-35.
- TURING, Alan. *Computing machinery and intelligence*. In: BODEN, Margaret (ed.). *The Philosophy of Artificial Intelligence*. Oxford: Oxford University Press, 1990. P. 40-66.
- VARELA, F. *Conhecer as Ciências Cognitivas: tendências e perspectivas*. Trad. Maria Teresa Guerreiro. Lisboa: Instituto Piaget, 1994.
- VARELA, F.; THOMPSON, E.; ROSCH, E. *A mente corpórea: Ciência Cognitiva e experiência humana*. Trad. Joaquim Nogueira Gil e Jorge de Sousa. Lisboa: Instituto Piaget, 2001.
- VICO, Giambattista. *Principios de uma ciéncia nueva: em torno a La naturaleza común de las naciones*. Tradução e Prólogo de José Carner. Mexico-DF: Fondo de Cultura Económica, 2006.

- VIGNAUX, G. *As Ciências Cognitivas: uma introdução*. Coleção Epistemologia e Sociedade. Tradução: Maria Manuela Guimarães. Lisboa: Instituto Piaget, 1991.
- YABLO, Stephen. *A paradox without self-reference*, Analysis, 53,4, 1993.
- WASSON, Barbara. *Advanced educational technologies: the learning environment, Computers in Human Behavior*, Vol. 13, Nº 4, p. 571-594, 1997.
- WITTGENSTEIN, Ludwig. *Investigações Filosóficas*. Lisboa: Fundação Calouste Gulbenkian, 1987.
- WHITBY, Blay. *Inteligência Artificial: um guia para iniciantes*. Tradução Cláudio Blanc. São Paulo. Madras, 2004.
- WUNDT, W.M. *Principles of physiological psychology*. Translate by Edward B. Titchener (1904). <http://www.blackmask.com>
- ZARANDONA, F.D. del C. *Blaise Pascal: la certeza y la duda*. Editorial Vision Net. Madrid. 2008